ENCYCLOPÆDIA BRITANNICA

Fourteenth Edition, Revised

1966

v8: EDW-EXT

ENCYCLOPÆDIA BRITANNICA

THE UNIVERSITY OF CHICAGO

The Encyclopædia Britannica
is published with the editorial advice of the faculties
of The University of Chicago and of a
committee of members of the faculties of Oxford, Cambridge
and London universities and of a committee
at The University of Toronto

*

"LET KNOWLEDGE GROW FROM MORE TO MORE
AND THUS BE HUMAN LIFE ENRICHED."

A New Survey of Universal Knowledge

ENCYCLOPÆDIA BRITANNICA

Volume 8

EDWARD *TO* EXTRACT

ENCYCLOPÆDIA BRITANNICA, INC.

WILLIAM BENTON, PUBLISHER

CHICAGO · LONDON · TORONTO · GENEVA · SYDNEY · TOKYO

©
1966
BY ENCYCLOPÆDIA BRITANNICA, INC.
COPYRIGHT UNDER INTERNATIONAL COPYRIGHT UNION
ALL RIGHTS RESERVED UNDER PAN AMERICAN AND UNIVERSAL COPYRIGHT
CONVENTIONS BY ENCYCLOPÆDIA BRITANNICA, INC.
PRINTED IN THE U. S. A.
Library of Congress Catalog Card Number: 66-10173

FOUNDED A.D. 1768

ENCYCLOPÆDIA BRITANNICA

Volume 8

EDWARD *TO* EXTRACT

EDWARD (EADWEARD), **SAINT,** THE CONFESSOR (*c.* 1003–1066), king of the English from 1042 to 1066, son of Aethelred II the Unready and of Emma, daughter of Richard I, duke of Normandy, was born at Islip, Oxfordshire, between 1002 and 1005. He was sent to Normandy (1013) when Sweyn I replaced Aethelred as king of England and he brought (1014) Aethelred's reply to the invitation asking him to return. Edward lived in Normandy, except for a brief visit to his mother in 1036 (the year of the murder of his brother Alfred), from 1016 to 1041, when he came to the court of his half brother Hardicanute, whom he succeeded in 1042, being crowned on April 3. One of his first acts was to seize his mother's goods (1043); she had favoured her son Hardicanute, and one source accuses her of supporting the claim of Magnus of Norway, which was based on an agreement with Hardicanute that if either died without heir the survivor should succeed. Expecting Magnus to invade, Edward collected his fleet at Sandwich (1045), but Magnus was prevented by his war with Sweyn Estrithson, king of Denmark, and his death (1047) removed the threat.

Edward was overshadowed by the power of Godwin (*q.v.*), earl of the West Saxons, whose daughter Edith he married in 1045. A breach occurred in 1051. Godwin, claiming to be acting against the foreigners brought in by the king, defied him, and Edward, with the support of the earls Leofric of Mercia and Siward of Northumbria, outlawed the Godwin family and dismissed his queen. However, his encouragement of foreigners and, according to some authorities, his promise of the succession to William, duke of Normandy, lost him sympathy, so that in 1052 Godwin and his sons could gather a large force and compel the king to reinstate them. The Norman prelates Robert, archbishop of Canterbury, and Ulf, bishop of Dorchester, fled with other Normans. Stigand replaced Robert as archbishop, an offense against canon law which enabled William to secure papal support for his invasion in 1066. When Godwin died (1053) his son Harold, afterward Harold II (*q.v.*), became the chief power in the land. By 1057 his three brothers were provided with earldoms, but in 1065 the Northumbrians successfully petitioned for the expulsion of Tostig, though he was a favourite of the king.

There were other unruly elements in this reign. A Dane, Osgod Clapa, outlawed in 1046, raided Essex in 1049. The Welsh made a raid in 1052, and when in 1055 and again in 1058 Earl Aelfgar of Mercia was outlawed, he was twice reinstated with the help of Welsh forces and in 1058, of a large fleet from Norway also. After Aelfgar's death Harold and Tostig reduced Wales (1062). The English interfered in Scottish politics (1054) when Siward invaded and routed Macbeth. Tostig became a friend of King Malcolm, but this did not prevent Malcolm from raiding Northumbria in Tostig's absence (1061).

Early in the reign England was drawn into continental politics. Edward maintained a fleet at Sandwich in 1048 to hamper the movements of Count Baldwin V of Flanders, who was at war with the emperor Henry III. Baldwin retaliated later by supporting English exiles. An embassy, led by Bishop Ealdred of Worcester, to the emperor in 1054 concerned the succession to the English throne. It resulted in the arrival (1057) of Edward, son of Edmund Ironside, and his family, but Edward died the same year. After the return to power of Godwin and his sons, a peaceful succession of William became unlikely, and on his deathbed Edward named Harold to succeed him.

Edward, later called "Confessor," has been blamed for behaving more like a monk than a king. Yet his reputation for piety may have helped to preserve some dignity for the crown in an age of overdominant magnates. His introduction of foreigners, natural in a king educated abroad, lost him the sympathy of his subjects. There is little evidence that the foreigners made much contribution to learning and the arts, in which, except for architecture, the English were in advance of the Normans. Edward's new church at Westminster, dedicated on Dec. 28, 1065, was the first to be built in the new continental style. Edward died on Jan. 5, 1066, and was buried the next day at Westminster. He was canonized in 1161. His feast days are Jan. 5 and Oct. 13 (translation).

BIBLIOGRAPHY.—*Anglo-Saxon Chronicle;* H. R. Luard (ed.), *Lives of Edward the Confessor* (1858), of which the first is written with Godwinist sympathies; *Florentii Wigorniensis monachi chronicon ex chronicis,* ed. by B. Thorpe (1848–49); Guillaume de Jumièges, *Gesta Normannorum ducum,* ed. by J. Marx (1914); F. M. Stenton, *Anglo-Saxon England,* 2nd ed. (1947); *English Historical Documents,* vol. ii, ed. by D. C. Douglas, pp. 14–17, 50 f., 110–141, 428–430 (1953); R. R. Darlington, "The Last Phase of Anglo-Saxon History," *History,* new series, vol. xxii (1937–38). (D. WK.)

EDWARD (EADWEARD), **SAINT,** THE MARTYR (*c.* 963–978), king of the English from 975 to 978, was the son of King Edgar by his first wife, Aethelflaed. On Edgar's death one party wished his younger son, Aethelred, to succeed him, but Edward was elected and crowned at Kingston-upon-Thames before the end of the year. His reign saw a great reaction from the promonastic policy of Edgar, led by Aelfhere, ealderman of Mercia, but reaching further afield, including Kent and Northumbria, where the earl, Oslac, was banished in 975 for some unstated cause. Monks were expelled and estates withdrawn from several houses, but the monastic cause was upheld by Aethelwine, ealderman of East Anglia, and his brother Aelfwold, and by Brihtnoth, ealderman of Essex, and eventually the reaction was halted. An estate belonging to Ramsey abbey was declared by the shire moot to be a royal estate illegally granted, but Edward did not press his claim to it. Little else is known of his short reign. Assemblies were held at Kirtlington, Oxfordshire and Calne, Wiltshire, both in 977, and a few charters survive. Edward was assassinated on March 18, 978, when visiting his brother at Corfe, Dorset, and was buried unceremoniously at Wareham, but a year later Ealderman Aelfhere translated his remains to Shaftesbury where it was claimed that they worked miracles. There is no contemporary evidence for the tale told a century later attributing his murder to his stepmother Aelfthryth. A code (1008) of Aethelred II enjoined that his festival was to be observed over all England.

BIBLIOGRAPHY.—*Anglo-Saxon Chronicle; Vita Oswaldi,* ed. by J. Raine, *Historians of the Church of York and Its Archbishops,* vol. i, pp. 443–446, 448–452 (1858); F. M. Stenton, *Anglo-Saxon England,* 2nd ed. (1947); *English Historical Documents,* vol. i, ed. by D. Whitelock, pp. 48, 208–211, 407, 522 f., 839–843, 857 (1955). (D. WK.)

EDWARD (EADWEARD) THE ELDER (d. 924), king of the West Saxons from 899 to 924, was a great military commander. His fame rests securely on his campaigns against the Danish armies which during the reign (871–899) of his father, Alfred, had occupied the eastern half of England. When Alfred died (Oct. 26, 899) Edward had to face trouble from an older cousin, Aethelwald, who may have thought he had a better claim to the throne. Aethelwald failed to attract sufficient support and took refuge among the Northumbrian Danes. Edward was crowned on June 8, 900. In 902 Aethelwald persuaded the Danes of East Anglia to attack Edward's territories but he was killed in battle. After an interval, Edward and the Danes made peace (906) at Yttingaford (now Tiddenfoot in Linslade, Buckinghamshire). The struggle broke out again in 909 when Edward sent a combined force of Mercians and West Saxons against the Northumbrian Danes. They retaliated in 910, raiding extensively in English Mercia, but they were decisively defeated (Aug. 5) at Tettenhall, Staffordshire, after which no more is heard of an effective Danish army in Northumbria.

Danish armies still controlled the eastern midlands and East Anglia, and the main theme of the years 910–916 was the building of a series of fortified enclosures (boroughs) to protect English territory from raids and to serve as bases for advance. Edward built two fortresses at Hertford (911, 912), two at Buckingham (914), one at Witham (912), one at Bedford (915) and one at Maldon (916). His sister Aethelflaed (*q.v.*) built a complementary series in the northwest midlands. The complete integration of West Saxon and Mercian policies is illustrated also by Edward's assumption (911) of direct control of the London-Oxford area, a traditionally Mercian district which Alfred had given (886) into the keeping of Aethelred, ealderman of the Mercians. Edward's fortress building was interrupted by Danish raids from Northampton and Leicester (913) and by the arrival of Vikings in the Severn (914), but by the end of 916 much territory had been recovered from the Danes and in 917 a great offensive was launched. In the first half of April Edward's forces occupied Towcester and after the middle of May he built a fortress at Wigingamere (apparently a few miles southeast of Cambridge). The Danes, alarmed by these advances, made three unsuccessful attacks on English positions. The struggle was at its height in the summer of 917 when Aethelflaed captured Derby, one of the great Danish strongholds in the midlands. By the end of the year Northampton, Huntingdon and Colchester were in Edward's hands and the whole Danish army of East Anglia, including that of Cambridge, had submitted. A final campaign was planned for 918. Edward advanced to Stamford, Aethelflaed occupied Leicester and the end was in sight when Aethelflaed died at Tamworth (June 12). Edward broke off his campaign to assert his authority over the Mercians. The princes and peoples of Wales also "sought him for their lord," and by the end of the year the last independent Danish armies in the midlands had submitted. Edward was lord of all the peoples south of the Humber.

His later fortresses were at Thelwall and Manchester (919), at Nottingham and Bakewell (920) and at Cledemutha (921), probably at the mouth of the river Clwyd in Wales. His main concern was now to protect his territories against Scandinavians resident outside his authority, especially against Irish-Norwegian adventurers who were rapidly gaining control of Northumbria. He seems to have succeeded Aethelflaed as the leader of an anti-Norse coalition of Northumbrian Danes, English, Picts, Scots and Britons. The entry (under the year 923 or 924) in the Anglo-Saxon Chronicle to the effect that all these peoples, and also the Irish-Norwegians, chose Edward "for father and for lord" probably reflects a general pacification of the north.

Other aspects of the reign are not unimportant. There were advances in the organization of the English church, in the minting of coins and in law, local government and administration. But Edward was, above all else, a great military leader. He died at Farndon on Dee on July 17, 924, a few days after quelling an obscure disturbance in Chester, perhaps an attempt by Mercians and Welshmen to escape from West Saxon domination.

BIBLIOGRAPHY.—*Anglo-Saxon Chronicle; Chronicon Aethelweardi* in *Monumenta historica Britannica,* ed. by H. Petrie (1848); *Florentii Wigorniensis chronicon ex chronicis,* ed. by B. Thorpe (1848–49); William of Malmesbury, *Gesta Regum,* ed. by W. Stubbs (1887–89); W. de G. Birch, *Cartularium Saxonicum* (1885–93); F. M. Stenton, *Anglo-Saxon England,* 2nd ed. (1947); *English Historical Documents,* vol. i, ed. by D. Whitelock (1955); F. T. Wainwright, "The Submission to Edward the Elder," *History,* new series, vol. xxxvii (1952). (F. T. W.)

EDWARD I (1239–1307), king of England from 1272 to 1307, was born at Westminster on June 17, 1239, the eldest son of Henry III and Eleanor of Provence. In 1254 he was given the duchy of Gascony, Oléron, the Channel Islands, Ireland, Henry's lands in Wales, the earldom of Chester and the castles of Bristol, Stamford, Grantham, Tickhill and the Peak. Henry negotiated Edward's marriage with Eleanor, half sister of Alfonso X of Castile. Edward reached Gascony in June 1254 and Castile in October, when Alfonso knighted him at Burgos. He married Eleanor at Las Huelgas (*c.* Oct. 31), returning to Bordeaux to organize his scattered apanage. He now had his own household and officials, chancery and seal, with an exchequer at Bristol castle; though nominally governing all his lands, in Gascony and Ireland he merely enjoyed the revenues. He returned to England in Nov. 1255 and attacked Llewelyn ap Gruffydd, prince of Gwynnedd, to whom his Welsh subjects had appealed when his seneschal, Geoffrey Langley, applied the English shire and hundred system to Edward's Welsh lands. Edward, receiving no help from either Henry or the marcher lords, was defeated ignominiously. His arrogant lawlessness and his close association with his Poitevin uncles increased his unpopularity, especially when he supported the Poitevin resistance to the Provisions of Oxford (1258). But after the Poitevins were expelled, Edward fell under the influence of Simon de Montfort, his uncle by marriage, with whom he made a formal pact.

As spokesman for the "community of the bachelors of England," Edward intervened dramatically to support the radical Provisions of Westminster (Oct. 1259), and in the dangerous crisis early in 1260 he supported Montfort and the extremists, though finally he deserted Montfort and was forgiven by Henry (May 1260). (*See* ENGLISH HISTORY: *Henry III* [*1216–72*].) He was sent to Gascony in Oct. 1260 but returned early in 1263; his violence in robbing the Temple and his quarrel with the Londoners harmed Henry's cause. At Lewes (May 14, 1264) his vengeful pursuit of the Londoners early in the battle contributed to Henry's defeat; he surrendered and became a hostage in Montfort's hands. He escaped at Hereford in May 1265 and took charge of the royal-

ist marchers' forces, penned Montfort behind the river Severn and by lightning strategy destroyed a large relieving army at Kenilworth (Aug. 1), trapped and slew Montfort at Evesham (Aug. 4) and rescued Henry. Shattered and enfeebled, Henry allowed Edward effective control, and his extreme policy of vengeance, especially against the Londoners, revived and prolonged rebel resistance. Finally the papal legate Ottobuono, Edward's uncle Richard, earl of Cornwall, and other moderates persuaded Henry to the milder policy of the *Dictum de Kenilworth* (Oct. 31, 1266), and after some delay the rebels surrendered. Edward took the cross (1268), intending to join Louis IX's crusade against Tunis, but was delayed by lack of money until Aug. 1270. Louis died before Edward's arrival, and Edward, after wintering in Sicily, went to Acre, where he stayed from May 1271 to Sept. 1272, winning fame by his energy and courage and narrowly escaping death by assassination, but achieving no useful results. On his way home he learned in Sicily of Henry III's death on Nov. 16, 1272.

Accession and Character.—Edward had nominated Walter Giffard, archbishop of York, Philip Bassett, Roger Mortimer and his trusted clerk Robert Burnell to safeguard his interests during his absence. After Henry's funeral the English barons all swore fealty to Edward (Nov. 20, 1272). His succession by hereditary right and the will of his magnates was proclaimed and England welcomed the new reign peacefully, Burnell taking charge of the administration with his colleagues' support. This quiet succession demonstrated England's unity only five years after a bitter civil war. Edward could journey homeward slowly, halting in Paris to do homage to his cousin Philip III for his French lands (July 26, 1273), staying several months in Gascony and reaching Dover on Aug. 2, 1274, for his coronation at Westminster on Aug. 19. Now 35 years old, Edward had redeemed a bad start. He had been arrogant, lawless, violent, treacherous, revengeful and cruel; his Angevin rages matched those of Henry II. Loving his own way and intolerant of opposition, he had still proved susceptible to influence by strong-minded associates. He had shown intense family affection, loyalty to friends, courage, brilliant military capacity and a gift for leadership; handsome, tall, powerful and tough, he had the qualities men admired. He loved efficient, strong government, enjoyed power and had learned to admire justice, though in his own affairs it was often the letter, not the spirit of the law that he observed. Having mastered his anger, he had shown himself capable of patient negotiation, generosity and even idealism, and he preferred the society and advice of strong counselors with good minds. As long as Burnell and Queen Eleanor lived, the better side of Edward triumphed, and the years until about 1294 were years of great achievement. Thereafter his character deteriorated for lack of independent advice, and his rule degenerated into autocratic militarism and megalomania.

Parliament and Statutes.—Shrewdly realistic, Edward understood the value of the "parliaments" which since 1254 had distinguished English government and which Montfort had deliberately employed to publicize government policy and to enlist widespread, active support by summoning representatives of shires and boroughs to the council to decide important matters. Edward developed this practice swiftly, not to share royal power with his subjects but to strengthen royal authority with the support of rising national consciousness. From 1275 to 1307 he summoned knights and burgesses to his parliaments in varying manners. The parliament of 1295, which included representatives of shires, boroughs and the lesser clergy, is usually styled the Model parliament, but the pattern varied from assembly to assembly as Edward decided. By 1307 parliament, thus broadly constituted, had become the distinctive feature of English politics, though its powers were still undefined and its organization embryonic.

Edward used these parliaments and other councils to enact measures of consolidation and reform in legal, procedural and administrative matters of many kinds. The great statutes promulgated between 1275 and 1290 are the glory of his reign. Conservative and definitory rather than original, they owed much to Burnell, Edward's chancellor. With the vast developments and reorganization of the administrative machine which Burnell coordinated, they created a new era in English government. The *quo warranto* inquiry, begun in 1275, the Statutes of Gloucester (1278) and of *Quo Warranto* (1290) sought with much success to bring existing franchises under control and to prevent the unauthorized assumption of new ones. By distraint of knighthood and other supporting measures, Edward strove, unsuccessfully, to restore the feudal army and strengthen local government institutions by compelling minor landowners to assume the duties of knighthood. His land legislation, especially the clause *De donis conditionalibus* in the miscellaneous Statute of Westminster II (1285) and the statute *Quia Emptores* (1290), eventually helped to undermine feudalism, quite contrary to Edward's purpose. The Statute of Mortmain (1279), by requiring royal consent, controlled the acquisition of land by ecclesiastical bodies. The Statute of Winchester (1285) codified and strengthened the police system for preserving public order. The Statute of Acton Burnell (1283), the Statute of Rhuddlan (1284) and the Statute of Merchants (1285) showed practical concern for trade and merchants. These are but the most famous of many statutes aimed at efficiency and sound administration.

Wars.—*Wales.*—Meanwhile, Edward destroyed the autonomous principality of Wales which, under Llewelyn ap Gruffydd, had expanded to include all Welsh lordships and much territory recovered from the marcher lords. Domestic difficulties had compelled Henry III to recognize Llewelyn's gains by the treaty of Shrewsbury (1267), but Edward was determined to reduce Llewelyn and used Llewelyn's persistent evasion of his duty to perform homage as a pretext for attack. He invaded Wales by three co-ordinated advances with naval support (1277), blockaded Llewelyn in Snowdonia, starved him into submission and stripped him of all his conquests since 1247. He then erected a tremendous ring of powerful castles encircling Gwynedd and reorganized the conquered districts as shires and hundreds. When English rule provoked rebellion, he methodically reconquered the principality, killing both Llewelyn and his brother David (1282–83). By the Statute of Rhuddlan he completed the reorganization of the principality on English lines, leaving the Welsh marches unaffected. A further Welsh rising in 1294–95 was ruthlessly crushed and Wales remained supine for over 100 years.

France.—After 1294 matters deteriorated. Queen Eleanor died in 1290, Burnell in 1292, and Edward never thereafter found such good advisers. The conquest and fortification of Wales had badly strained his finances; now endless wars with Scotland and France bankrupted him. He quarreled bitterly with both clergy and barons, behaving as a rash and obstinate autocrat refusing to recognize his limitations. Philip III and Philip IV of France had both cheated him of the contingent benefits promised by the treaty of Paris (1259). By constant intervention on pretext of suzerainty they had nibbled at his Gascon borders and undermined the authority of his administration there. After doing homage to Philip IV in 1286, Edward visited Gascony to reorganize the administration and restore authority. On returning to England in 1289 he had to dismiss many judges and officials for corruption and oppression during his absence. In 1290, having systematically stripped the Jews of their remaining wealth, he expelled them from England. French intervention in Gascony was now intensified; affrays between English and French sailors inflamed feelings; and in 1293 Philip IV tricked Edward's brother Edmund, earl of Lancaster, who was conducting negotiations, into ordering a supposedly formal and temporary surrender of the duchy, which Philip then refused to restore. The Welsh rising and Scottish troubles prevented Edward from taking action and when at last, in 1297, he sailed to attack France from Flanders, his barons refused to invade Gascony and Wallace's rising forced him to return. He made peace with Philip (1299), by Boniface VIII's persuasion, married Philip's sister Margaret and eventually recovered an attenuated Gascon duchy.

Scotland.—For over 100 years relations between England and Scotland had been amicable and the border had been remarkably peaceful. Edward inaugurated 250 years of bitter hatred, savage warfare and bloody border forays. The deaths of Alexander III of Scotland (1286) and his granddaughter Margaret, the Maid of Norway (1290), whom Edward planned to marry to his heir,

Edward of Caernarvon (afterward Edward II), ended the line of succession. Many dubious claimants arose, and the Scottish magnates requested Edward's arbitration. Edward compelled the nobles and the claimants to recognize his suzerainty, and only then adjudged John de Balliol king (1292). Balliol did homage and was crowned, but Edward's insistence on effective jurisdiction, as suzerain, in Scottish cases eventually provoked the Scottish nobles to force Balliol to repudiate Edward's claims and to ally with France (1295). Edward invaded and conquered Scotland (1296), removing to Westminster the coronation Stone of Scone. William Wallace led a revolt in 1297 and Edward, though brilliantly victorious at Falkirk (July 22, 1298), could not subdue the rebellion despite prolonged campaigning (1298–1303).

Last Years.—The strain of these wars provoked heavy collisions between Edward and his magnates. He had quarreled violently with his archbishops of Canterbury, John Pecham (1279–92) and Robert Winchelsea (1293–1313), over ecclesiastical liberties and jurisdiction. In 1297 Winchelsea, obeying Boniface VIII's bull *Clericis laicos* (1296), rejected Edward's demands for taxes from the clergy, whereupon Edward outlawed the clergy. His barons now defied his orders to invade Gascony and, when Edward went to Flanders, compelled the regents to confirm the charters of liberties, with important additions forbidding arbitrary taxation (1297), thereby forcing Edward to abandon the campaign and eventually to make peace with France. Although Pope Clement V, more pliant than Boniface, allowed Edward to exile Winchelsea and intimidate the clergy (1306), the barons had exacted further concessions (1301) before reconciliation. Edward renewed the conquest of Scotland in 1303, captured Stirling in 1304 and executed Wallace as a traitor in 1305, but when Scotland seemed finally subjected, Robert the Bruce revived rebellion and was crowned in 1306. On his way to reconquer Scotland, Edward died at Burgh by Sands, near Carlisle, on July 7, 1307.

See also references under "Edward I" in the Index volume.

BIBLIOGRAPHY.—F. M. Powicke, *King Henry III and the Lord Edward* (1947), *The Thirteenth Century* (1953); T. F. Tout, *Chapters in the Administrative History of Medieval England,* vol. ii (1920), *Edward I* (1903); D. Pasquet, *Essay on the Origins of the House of Commons* (1925); T. F. T. Plucknett, *Legislation of Edward I* (1949); J. E. Morris, *The Welsh Wars of Edward I* (1901); E. M. Barron, *The Scottish Wars of Independence,* 2nd ed. (1934). (R. F. T.)

EDWARD II (1284–1327), known as Edward of Caernarvon, king of England from 1307 to 1327, the fourth son of Edward I and his first wife, Eleanor of Castile, was born at Caernarvon on April 25, 1284, and became heir to the throne on the death of his brother Alfonso (Aug. 1284). He was granted the royal lands in Wales and the earldom of Chester (1301), but there is no evidence that he placed particular emphasis on this Welsh connection. More important for the future were his friendship with Piers Gaveston, the Gascon knight whom Edward I banished in Feb. 1307, and his marriage in Jan. 1308 with Isabella, daughter of Philip IV the Fair of France.

Edward became king on July 7, 1307. He immediately reinstated some of his father's more prominent opponents and recalled Gaveston from exile, making him earl of Cornwall (Aug. 1307). At his coronation (Feb. 1308) Edward took a fourfold oath, which added to the traditional three promises an undertaking to maintain the laws and rightful customs chosen by the people of the realm. In this the king bound himself to little that was new, but the oath was subsequently used against him.

During the two decades of his reign Edward was engaged in an almost continuous struggle with the barons, which culminated in his deposition and death. In this conflict parliament became increasingly important as a political assembly. Edward's personality explains much of his failure. Though physically strong and a good horseman, he was not intelligent. A source written in 1315 describes his interest in such rustic pursuits as making ditches, and his unconventional tastes separated him in outlook from the majority of the feudal magnates. He thus turned to favourites such as Gaveston for friendship and advice, while his faults of character lost him any more widespread support.

The Ordinances.—The conflict between Edward II and the barons which arose from the baronial attempt to restrict the king's power, particularly in the choice of advisers and control of the household, developed immediately over Gaveston. The barons compelled the king to banish him (June 1308), but the favourite was soon recalled (July 1309). His unpopularity and the favour shown to him by Edward resulted in the formation of a baronial committee of 21, the lords ordainers, who drafted the document known as the ordinances (1311). Their terms included demands for the renewal of the sentence of banishment on Gaveston and the limitation of the royal prerogative in such fields as finance, declaration of war and appointments in the household. Parliaments were to be held at least annually. Although the ordinances still envisaged parliament as a baronial assembly, they may mark an advance in the commons' participation, for after 1311 knights and burgesses were summoned more frequently.

Edward was obliged to accept the ordinances, but he endeavoured to build up a party to overthrow them and quickly annulled the sentence on Gaveston. Such actions constituted a direct challenge to the barons. A force under Aymer de Valence, earl of Pembroke, and John de Warenne, earl of Surrey, captured Gaveston at Scarborough castle and, despite a promise of personal immunity, he was executed on Blacklow hill (June 1312). Although a formal reconciliation was effected in 1313, Gaveston's death bred lasting enmity between Edward and certain of the leading barons.

Edward had to wait 11 years to annul the ordinances and avenge Gaveston. These years were occupied by Scottish affairs and by the ambitious actions of the king's cousin Thomas, earl of Lancaster. While Edward quarreled with the ordainers, Robert I the Bruce was consolidating his power in Scotland. A treaty in 1312 confirmed the cession of the Western Isles from Norway to Scotland, and in 1313 Scottish rule was restored in the Isle of Man. Stirling castle, the only important stronghold left in the hands of the English, was besieged in 1313. An English army, led north by Edward in an effort to save the garrison, was decisively defeated at Bannockburn on June 24, 1314. In the long run the battle proved a major step toward the practical and legal recognition of Scotland's independence. Immediately, it left Bruce free to devastate the northern counties of England and forced Edward more under the control of the leading barons. Thomas of Lancaster forced changes in the royal household at the York parliament of 1314, and in 1315 England was virtually under his control. Like Edward, however, Lancaster proved to have faults of character. He was suspected of intriguing with Bruce and he allowed personal enmity toward Edward to dislocate the work of government. His failure of statesmanship led to a new baronial grouping, headed by the earl of Pembroke, which historians have called the "middle party." Their policy was to maintain the ordinances and extend administrative reform. By the treaty of Leake in Aug. 1318, they effected a formal reconciliation between Lancaster and the king on the basis of the ordinances, and established a standing council whose consent was necessary to acts of sovereignty. One member of the council was nominated by Lancaster.

The treaty of Leake revealed the weakness of Edward's position. The Household Ordinance of York (Oct. 1318) reviewed household appointments, and Edward's reputation was further diminished by the Scottish victory over the English forces at Myton-upon-Swale (Sept. 1319). At this juncture Edward found new friends in the Despensers (*see* DESPENSER). Hugh le Despenser the younger had inherited, through his wife, Eleanor, part of the estates of Gilbert de Clare, earl of Gloucester. With Edward's help he now attempted to acquire the lordship of Gower, as a preliminary to seizing the whole of the Clare property in Wales. But he aroused the hostility of the marcher lords and was defeated in Wales. Lancaster called various assemblies in the north, seeking to rouse opinion there, and a parliament held in July 1321 finally banished the Despensers. Edward took up arms on their behalf against what proved to be a divided opposition. Sir Roger Mortimer of Chirk and his nephew, Roger Mortimer of Wigmore, surrendered in the west, and on March 16, 1322, Lancaster was defeated at the battle of Boroughbridge and executed near his own castle of Pontefract (March 22).

At last free of baronial control, Edward revoked the ordinances

by the Statute of York (May 1322), which also stated that traditional methods of legislation by the king in parliament should be observed. The view that the statute intended to confer new rights on the commons is extremely doubtful. The period after Boroughbridge saw reforms which make the reign a turning point in administrative history. The work of the chamber was expanded under Hugh le Despenser the younger and that of the wardrobe restricted more closely to the affairs of the household. At the exchequer, Walter Stapledon, bishop of Exeter, was responsible for the reorganization of the exchequer records.

Isabella's Rebellion and the Deposition.—The Despensers were the main architects of administrative reform, but their growing wealth created resentment. In particular they aroused the antagonism of Queen Isabella, whose estates had been seized (1324) on the excuse that a French invasion was planned. Isabella sailed for France (March 1325) in an attempt to settle the perennial dispute with the French about Gascony, and in September was joined there by her son, the future Edward III. With various baronial exiles, including Roger Mortimer of Wigmore who had been in France since 1323 and whose mistress she became, she crossed to Essex in Sept. 1326, and declared her intention of removing the Despensers. Edward found himself without support. He fled to the west with the Despensers, but they were captured and executed, and Edward was imprisoned at Kenilworth (Nov. 1326).

Isabella and Mortimer determined to remove Edward from the throne. A revolutionary assembly which termed itself a parliament, and in which the commons were fully represented, met on Jan. 7, 1327. In this assembly Edward III was chosen king, and by the so-called Articles of Deposition Edward II was declared incompetent to govern. A deputation representative of the different estates of freemen in the realm went to Kenilworth and compelled Edward to renounce the throne; his son's reign was held to begin on Jan. 25. Despite these attempts to legalize revolution, it was considered unsafe for Edward to remain alive. In April 1327 he was removed from Kenilworth to Berkeley castle, where, after two attempted rescues, he was almost certainly murdered (Sept. 1327). He was buried in Gloucester, in St. Peter's abbey (now the cathedral), where miracles were later said to have been performed at his tomb.

BIBLIOGRAPHY.—H. Johnstone, *Edward of Carnarvon* (1946); M. Mckisack, *The Fourteenth Century* (1959); J. C. Davies, *The Baronial Opposition to Edward II* (1918); T. F. Tout, *The Place of the Reign of Edward II in English History*, 2nd ed. (1936), *Chapters in the Administrative History of Medieval England*, vol. ii (1920); W. Stubbs, *Constitutional History*, vol. ii, 4th ed. reprinted (1929); B. Wilkinson, *Constitutional History of Medieval England*, vol. ii and iii (1948–58); M. V. Clarke, *Medieval Representation and Consent* (1936).

(J. TA.)

EDWARD III (1312–1377), king of England from 1327 to 1377, the eldest son of Edward II and Isabella of France, was born at Windsor on Nov. 13, 1312. He was summoned to parliament as earl of Chester (1320) and was made duke of Aquitaine (1325), but he never received the title of prince of Wales. He was sent to France to do homage (Sept. 1325) to his uncle Charles IV for Guienne, Gascony and Ponthieu and remained abroad until he accompanied his mother and Roger Mortimer of Wigmore in their expedition to England (Sept. 1326). To raise funds for this enterprise he was betrothed to Philippa, daughter of William II, count of Hainaut. The prince was proclaimed keeper of the realm (Oct. 26, 1326), and after Edward II had been forced to resign, he was accepted as king, his reign being held to begin on Jan. 25, 1327. He was crowned on Jan. 29, 1327.

During the next four years Isabella and Mortimer governed in his name, though nominally his guardian was Henry, earl of Lancaster. In the summer of 1327 he took part in an abortive campaign against the Scots. He married Philippa at York on Jan. 24, 1328, and his eldest child, Edward, later called the Black Prince, was born on June 15, 1330. Soon afterward, Edward made a successful effort to throw off his degrading dependence on his mother and Mortimer. While a council was being held at Nottingham, he entered the castle by night, through a subterranean passage, took Mortimer prisoner (Oct. 1330) and procured his execution (Nov. 1330). Edward discreetly ignored his mother's liaison with Mortimer and treated her with every respect, but her political influence was at an end.

Edward III now began to rule as well as to reign. Young, ardent and active, he sought to restore England to the position it had acquired under Edward I. He resented the concession of independence made to Scotland by the treaty of Northampton (1328), and the death of Robert I the Bruce in 1329 gave him a chance of retrieving his position. The new king of Scots, his brother-in-law, David II, was a mere boy, and the Scottish barons who had been exiled by Bruce for their support of the English took advantage of the weakness of his rule to invade Scotland in 1332. At their head was Edward Balliol whose victory at the battle of Dupplin Moor (Aug. 1332) established him for a brief time on the Scottish throne. Balliol was defeated by a Scottish coalition (Dec. 1332), whereupon Edward III for the first time openly took up his cause and in person won a victory at the battle of Halidon Hill (July 1333). David II fled to France but the Scots always despised Balliol as a puppet of the English king and David was able to return in 1341.

The French Wars.—During the 1330s England gradually drifted into a state of hostility with France, for which the most obvious reason was the dispute over English rule in Gascony. Contributory causes were Philip VI's support of the Scots, Edward's alliance with the Flemish cities, then on bad terms with their overlord, and the revival (1337) of Edward's claim, first made in 1328, to the French crown. Philip declared Gascony forfeit (1337) and Edward visited Coblenz (Sept. 1338), where he confirmed an alliance with the emperor Louis IV (V) the Bavarian. Edward twice endeavoured to invade France from the north (1339, 1340) with the help of his German and Flemish allies, but the only result of his campaigns was to reduce him to bankruptcy. Edward assumed the title of king of France in Jan. 1340. At first he may have done this to gratify the Flemings, whose scruples in fighting their overlord, the French king, disappeared when they persuaded themselves that Edward was the rightful king of France. But his pretensions to the French crown gradually became more important and the persistence with which he and his successors urged them made stable peace impossible for more than a century. This was the struggle famous in history as the Hundred Years' War (*q.v.*). Until 1801 every English king also called himself king of France.

Crécy and Calais.—Edward was present in person at the great naval battle off Sluis in June 1340, in which he temporarily destroyed the French navy. Despite this victory he was so exhausted by his land campaign that he was forced to make a truce (Sept. 1340) and return to England. He unfairly blamed his chief minister, John de Stratford, archbishop of Canterbury, for his financial distress and vindictively attacked him. Before the truce expired a disputed succession to the duchy of Brittany gave Edward an excuse for renewing hostilities with France. He went to Brittany (1342) and fought an indecisive campaign. During the following years he spent much time and money in rebuilding Windsor castle and instituting the Order of the Garter, in fulfillment of a vow (Jan. 1344) to restore the Round Table of Arthur. A new phase of the French war began when Edward landed in Normandy (July 1346), accompanied by Prince Edward. At first he showed some want of strategic purpose. In an expedition that was little more than a large-scale plundering raid he marched via Caen almost to the gates of Paris. The campaign was made memorable by his decisive victory over the French at Crécy in Ponthieu (Aug. 26), where he scattered the army with which Philip VI sought to cut off his retreat to the northeast. Edward laid siege to Calais in Sept. 1346 and received its surrender in Aug. 1347. Other victories in Gascony and Brittany and the defeat and capture of David II at Neville's Cross near Durham (Oct. 1346) further emphasized his power, but this was the most solid and lasting of his conquests. He ejected most of the French inhabitants of Calais, colonizing the town with Englishmen and establishing there a base from which to conduct further invasions of France. Nevertheless, in the midst of his successes, want of money forced him to make a new truce (Sept. 1347).

Poitiers and the Treaty of Calais.—Edward returned to England in Oct. 1347. He celebrated his triumph by a series of splendid

tournaments and completed his scheme for the establishment of the Order of the Garter (*c.* 1349). In 1348 he rejected an offer of the imperial throne. In the same year the Black Death first appeared in England and raged until the end of 1349. Its horrors hardly checked the magnificent revels of Edward's court, and neither the plague nor the truce stayed the slow course of the French war, though the fighting was indecisive and on a small scale. Edward's martial exploits during the next years were those of a gallant knight rather than of a responsible general. Conspicuous among them were his famous combat with Eustace de Ribaumont in a melee near Calais (1349) and the hard-fought naval victory over the Spaniards off Winchelsea (1350). Although the English commons were now weary of the war, efforts to make peace, initiated by Pope Innocent VI, came to nothing and large-scale operations began again in 1355 when Edward led an unsuccessful raid out of Calais. He harried the Lothians in the expedition famous as the Burned Candlemas (Jan. and Feb. 1356), and in the same year he received a formal surrender of the kingdom of Scotland from Balliol. His exploits were, however, eclipsed by those of his son, whose victory at Poitiers (Sept. 19, 1356), resulting in the captivity of the French king John II, forced the French to accept a new truce. Edward entertained his captive magnificently, but forced him by the treaty of London (1359) to surrender so much territory that the French repudiated the agreement. In an effort to compel acceptance, Edward landed at Calais (Oct. 28) and besieged Reims, where he planned to be crowned king of France. The strenuous resistance of the citizens frustrated this scheme and Edward marched into Burgundy, eventually returning toward Paris. After this unsuccessful campaign he was glad to conclude preliminaries of peace at Brétigny, near Chartres (May 8, 1360). This treaty, less onerous to France than that of London, took its final form in the treaty of Calais, ratified by both kings (Oct. 1360). By it Edward renounced his claim to France in return for the whole of Aquitaine.

The Years of Decline: 1360–77.—The treaty of Calais did not bring rest or prosperity to either England or France. Fresh visitations of the Black Death (1361, 1369) intensified social and economic disturbances and desperate but not very successful efforts were made to enforce the Statute of Labourers (1351), which was intended to maintain prices and wages as they had been before the pestilence. Other famous laws enacted during the 1350s had been the statutes of Provisors (1351) and Praemunire (1353), which reflected popular hostility against foreign clergy. These measures were frequently re-enacted and Edward formally repudiated (1366) the feudal supremacy over England still claimed by the papacy by reason of King John's submission in 1213.

When the French king Charles V repudiated the treaty of Calais, Edward resumed the title of king of France, but he showed little of his former vigour in meeting this new trouble and left most of the fighting and the administration of his foreign territories to his sons Edward and John of Gaunt, duke of Lancaster. While they were struggling with little success against the rising tide of French national feeling, Edward's want of money made him a willing participant in the attack on the wealth and privileges of the church. The chancellor, William of Wykeham, bishop of Winchester, and the treasurer, Thomas Brantingham, bishop of Exeter, were driven from office (1371) and replaced by laymen, who proved, however, less effective administrators. Meanwhile Aquitaine was gradually lost, and the defeat of the earl of Pembroke off La Rochelle (1372) reflected English weakness at sea. The Black Prince returned to England in broken health (1371) and John of Gaunt's march through France from Calais to Bordeaux (1373) achieved nothing. Edward's final attempt himself to lead an army abroad (1372) was frustrated when contrary winds prevented his even landing his troops in France. In 1375 he was glad to make a truce, which lasted until his death. By it the only important possessions remaining in English hands were Calais, Bordeaux, Bayonne and Brest.

Edward was now sinking into his dotage. After the death of Queen Philippa in 1369 he fell entirely under the influence of a greedy mistress called Alice Perrers, while the Black Prince and John of Gaunt became the leaders of sharply divided parties in the royal court and council. John of Gaunt returned to England in April 1374 and by the help of Alice Perrers obtained the chief influence with his father, but his administration was neither honourable nor successful. His chief enemies, headed by William of Wykeham, were the higher ecclesiastics, whom he further irritated by his support of the Lollard leader John Wycliffe. The secular opposition to Gaunt was led by the Black Prince and Edmund Mortimer, earl of March, the husband of Edward's granddaughter Philippa of Clarence. At the famous Good parliament of 1376 popular indignation against the ruling party came at last to a head. Alice Perrers was removed and some of Gaunt's followers were impeached. Before the parliament had concluded its business, however, the death of the Black Prince (June 8, 1376) robbed the commons of their strongest support. John of Gaunt regained power and the acts of the Good parliament had been reversed when Edward III died, on June 21, 1377, at Sheen (now Richmond).

Character and Family.—Edward III possessed extraordinary vigour and energy of temperament; he was an admirable tactician and a consummate knight. His court, described at length in Jean Froissart's famous chronicle, was the most brilliant in contemporary Europe, and he was himself well fitted to be the head of the magnificent chivalry that obtained fame in the French wars. Though his main ambition was military glory, he was not a bad ruler of England, being liberal, kindly, good tempered and easy of access. His need to obtain supplies for carrying on the French wars made him favourable to his subjects' petitions and contributed to the growing strength of parliament. His weak points were his wanton breaches of good faith, his extravagance, his frivolity and his self-indulgence. Like that of Edward I, his ambition transcended his resources, and before he died even his subjects sensed his failure.

Edward had seven sons and five daughters. Five of his sons played some part in the history of their time, these being Edward, the Black Prince; Lionel of Antwerp, duke of Clarence; John of Gaunt, duke of Lancaster; Edmund of Langley, afterward duke of York; and Thomas of Woodstock, afterward duke of Gloucester. John and Edmund are also important as the founders of the rival houses of Lancaster and York.

See also references under "Edward III" in the Index volume.

BIBLIOGRAPHY.—J. Barnes, *The History of Edward III* (1688); W. Longman, *The Life and Times of Edward the Third* (1869); J. MacKinnon, *The History of Edward the Third* (1900); W. Stubbs, *Constitutional History of England*, vol. ii (1866 and many later editions); M. McKisack, *The Fourteenth Century* (1959); F. A. Walters, *Last Silver Coinage of Edward III* (1902); G. Unwin, *Finance and Trade Under Edward III* (1918); E. Perroy, *The Hundred Years War*, Eng. trans. by W. B. Wells (1951). (T. F. T.; J. R. L. H.)

EDWARD IV (1442–1483), king of England from 1461 to 1483 except for the brief period (Oct. 1470–April 1471) when the deposed Henry VI was restored, was the eldest surviving son of Richard, duke of York, by Cicely, daughter of Ralph Neville, earl of Westmorland, and was born at Rouen on April 28, 1442. His father was the lineal representative both of Edward III's third son, Lionel, duke of Clarence, and of his fifth son, Edmund, duke of York, whereas the rival house of Lancaster could trace descent only from the fourth son of Edward III, John of Gaunt. When the government of Henry VI became discredited by weakness and unchecked faction at home and by disastrous failure in France, it was therefore possible for Richard of York to claim the throne on the ground of hereditary right (in spite of Lancastrian objections to a descent from Clarence that depended on two women, Philippa and Anne Mortimer). His claim was the more formidable because his marriage had brought him the support of the powerful Neville connection, above all of the earls of Salisbury and Warwick. In 1459 York, Salisbury and Warwick joined forces at Ludlow, headquarters of York's great estates on the Welsh marches, where Edward, who had from boyhood been styled earl of March, normally resided. However, the Yorkist forces melted away on the approach of the royal army; the duke of York had to flee to Ireland and Edward, with Salisbury and Warwick, his uncle and cousin, sped to Calais. As governor of Calais, Warwick commanded a fleet which protected them from attack and enabled them to invade Kent. Having occupied London, they marched north to meet the royal

army at Northampton, where they heavily defeated the Lancastrians (July 10, 1460) and captured the king. When Lancastrian forces slew Richard of York at Wakefield (Dec. 30, 1460), Edward was in the Welsh marches. Hastily gathering an army there, he defeated the earls of Pembroke and Wiltshire at Mortimer's Cross (Feb. 2, 1461) and then marched on London where he was acclaimed king in Westminster hall (March 4). Nine days later he left for the north with Warwick and Norfolk and their forces, and at Towton (March 29) he won a complete victory over the Lancastrians. Henry VI and Queen Margaret fled to Scotland, while Edward returned to London for his coronation (June 28, 1461).

The Struggle With Warwick.—Edward was at this time a strikingly handsome young man, over six feet tall, with golden hair and charming manners. Time was to show him a born general and a strong ruler, but at the moment he appeared to be merely a dissolute youth, caring only for fighting, drinking, women and pageantry. He owed his throne largely to his cousin Richard Neville, now earl of Salisbury (1461) as well as of Warwick, who was in the first years of Edward's reign the most powerful man in England. It was Warwick who crushed Lancastrian resistance in the far north of England between 1462 and 1464 and conducted England's diplomacy. Edward, however, was winning many friends (especially in London) by his comeliness and charm and was determined to assert his independence. In Sept. 1464, when Warwick was in the midst of negotiations for a peace with France, to be sealed by a royal match with a sister-in-law of Louis XI, he was astounded and humiliated by the king's announcement that he had secretly married (May 1, 1464) a beautiful young widow, Elizabeth Woodville. Not only did Edward offend Warwick and other Yorkist nobles by this hasty marriage to a woman of relatively unimportant family—a woman who was, moreover, the widow and daughter of Lancastrians—but, by showering favours on Elizabeth's two sons by her first husband, on her five brothers and her seven sisters, he began to build up a group of magnates who would be a counterpoise to the Nevilles. Gradually Warwick lost all influence at court, and when he was once more negotiating for an alliance with France, Edward increased his humiliation and resentment by revealing that he had already concluded an alliance (1467) with France's enemy Burgundy. Edward's sister Margaret was married in July 1468 with great pomp to Duke Charles the Bold of Burgundy and the brothers-in-law planned a joint invasion of France.

As a countermove the king of France, Louis XI, encouraged Warwick to stir up risings in the north of England against Edward IV (1469 and 1470). Warwick found a tool in Edward's weak brother George, duke of Clarence, to whom Warwick married (1469) his elder daughter and co-heiress, Isabel, in spite of Edward's orders. The king, always overtrustful of others in his earlier years, was suddenly taken prisoner (July 1469) by Warwick's brother George Neville, the archbishop of York, and for over two months remained a captive of Warwick. But Edward had by now too many supporters (especially in London) for him to be kept under tutelage for long. He regained his freedom in Oct. 1469, and when it was proved that the Lincolnshire rebellion of March 1470 was instigated by Warwick and Clarence, he summoned them to account. They fled, however, to France (April 1470), made terms with their former bitter foes, Queen Margaret and the Lancastrian exiles, and with the help of Louis XI invaded England (Sept. 1470).

Edward prepared to resist, but was suddenly surprised by Warwick's brother John Neville, marquess of Montagu, in whom Edward had characteristically trusted too much. With only hours to spare, Edward fled with a few faithful supporters to Holland (Oct. 1470). At first Charles of Burgundy refused to help him for fear of provoking a French invasion, but when it became clear that for Louis XI the prime purpose of the revolution in England was to make possible a joint Anglo-French invasion of Burgundy, Charles retaliated by supplying men and money to help Edward regain his throne. Edward landed with his faithful brother Richard, duke of Gloucester, and a small force at Ravenspur on the Humber (March 1471). As a general he was always bold and swift, and soon he was well on the way to London, outmaneuvering Warwick and joined by the vacillating Clarence, who had by now realized that his ambitions were thwarted by Warwick's alliance with Queen Margaret. Having gained the great prize of London and gathered recruits there, Edward came out to fight Warwick at Barnet, on Easter day, April 14, 1471. The battle ended in a complete victory for Edward, and Warwick and Montagu were slain on the field. On that very day Queen Margaret belatedly landed in Dorset with her only son, Edward, prince of Wales. Her advisers hoped to gain Lancastrian support in Wales, and it became a race for time between Edward IV's forces and hers as to whether she could get there before he overtook her. At Tewkesbury, after some remarkable forced marches (one of over 40 mi. at a stretch), he caught up with her army on May 4. There he won another crushing victory. Nearly all the remaining Lancastrian leaders were killed on the field or executed afterward and, after murdering Henry VI (May 21–22) and repelling an attack on London, Edward was secure for the remainder of his life.

The Second Half of the Reign.—He was now able to revive the project of an invasion of France in concert with the duke of Burgundy. He made great preparations in 1474 and obtained a large grant from parliament. In 1475 he invaded France with the largest army, it was said, that had ever left England, but he found the duke of Burgundy very ill-prepared and the French formidable and willing to buy him out. Hence the treaty of Picquigny was made by which Edward agreed to withdraw from France in return for 75,000 gold crowns down and a pension of 50,000 gold crowns a year while both kings survived. These sums helped to free Edward from dependence on parliamentary grants. As he grew older, he became more avaricious and showed considerable ingenuity in raising money by reviving obsolescent rights and using doubtfully legal devices. Commercial treaties with France (1475), Burgundy (1468) and the Hanseatic league (1474) combined with external peace and growing internal order to revive trade strikingly after 1475, and this benefited the customs duties and other revenues. Edward became a trader himself, transporting goods in his own ships (for he encouraged shipbuilding) and those of foreign merchants. He began a reorganization of the revenues from the crown estates, experimenting with methods of improving yields and promoting more efficient auditing under officials accountable to the obedient and flexible royal household treasury instead of to the unadaptable exchequer. All these and other measures enabled him to leave behind a fortune; he was the first king to die solvent after Henry II. Indeed, some of his improved financial administration was continued and developed by Richard III and Henry VII.

As the last decade of Edward's reign saw an improvement in the royal finances, so it did in law enforcement. There is evidence that the king took energetic steps to repress unruliness, and although progress here was necessarily slow, it impressed contemporaries. One especially disturbed area was Wales and the Welsh marches; Edward used the royal estates there (much strengthened by the addition of the great Yorkist estates round Ludlow) as a foundation on which to base a council that acted in the name of his infant heir, the prince of Wales, and employed the royal prerogative to make a start in repressing disorder. It was the forerunner of the council of Wales and the marches which at last brought peace to that turbulent region.

Modern research has emphasized these administrative achievements of Edward IV, and contemporary and Tudor historians viewed his later years as a time of prosperity and success. Yet some dark shadows remained. The king's complex character included traits of energy and idleness, ability and dissipation. It is true that some of his leisure hours were devoted to artistic pursuits, such as the rebuilding of St. George's chapel, Windsor, and to forming a library, especially of the illuminated Flemish manuscripts which he had learned to admire during his exile in Bruges in 1470–71. He was also a friend and patron of the printer William Caxton, and his book collection became the foundation of the Old Royal library, later one of the glories of the British museum. But much of his leisure was spent in less profitable ways of self-indulgence, and the frequency of his amours enabled Richard of

Gloucester, after his death, to question the validity of his marriage and so to ruin his sons. As a young man Edward had been trustful and openhanded, but his experiences made him increasingly suspicious and avaricious. The former trait led him to execute (1478) his own brother Clarence, who in former years had been repeatedly pardoned for his follies but had now by acts of defiance given fresh grounds for offense. The latter characteristic made the king especially angry when, in 1482, Louis XI, in order to come to terms with the rulers of Burgundy, tacitly repudiated the treaty of Picquigny and the annual tribute that it provided. Edward contemplated a fresh invasion of France, the popularity of which is attested by the alacrity with which the usually parsimonious commons voted a subsidy for this purpose. Before it could be levied, however, Edward fell ill, and died at Westminster on April 9, 1483, after an illness of ten days, at the age of only 40. Contemporaries differed as to whether his death was due to chagrin, debauchery or a quartan fever. By Elizabeth Woodville he had seven children who survived him: two sons, Edward (afterward Edward V) and Richard, duke of York, who were probably murdered in the Tower of London in Aug. 1483, and five daughters, of whom the eldest, Elizabeth, married Henry VII.

See also references under "Edward IV" in the Index volume.

Bibliography.—J. H. Ramsay, *Lancaster and York*, vol. ii (1892); C. L. Scofield, *The Life and Reign of Edward IV*, 2 vol. (1923); J. R. Lander, "Edward IV: the Modern Legend: and a Revision," *History*, vol. xli (1956), "Council, Administration and Councillors, 1461 to 1485," *Bulletin of the Institute of Historical Research*, vol. xxxii (1959); B. P. Wolffe, "The Management of the English Royal Estates Under the Yorkist Kings," *English Historical Review*, vol. lxxi (1956); A. R. Myers, *The Household of Edward IV* (1959). *See* also bibliography for C. H. Williams, "The Yorkist Kings 1461–1485," *Cambridge Medieval History*, vol. viii (1936). (A. R. M.)

EDWARD V (1470–1483?), king of England from April to June 1483, was the eldest surviving son of Edward IV and Queen Elizabeth (Woodville). He was born between Nov. 1 and 3, 1470, in the sanctuary of Westminster abbey where his mother had taken refuge in Oct. 1470 when the king had fled to Holland. On June 26, 1471, after Edward IV had returned and crushed all his foes, he created his son prince of Wales, and the child received formal grants of the principality of Wales, the counties palatine of Chester and Flint and the duchy of Cornwall (July 17). A business council was appointed for the affairs of the principality (Feb. 20, 1473), and Edward was sent with his mother to Ludlow, to be titular ruler of the lands granted to him in Wales and the marches. The queen's brother Anthony Woodville, Earl Rivers, was made his governor, and Sir Richard Grey, her younger son by her first marriage, was chosen as one of his councilors. Except for brief intervals the prince seems to have stayed at Ludlow for the remainder of his father's reign, as representative of the royal authority in the Welsh marches.

By the time Edward IV died (April 9, 1483), the ambition of the queen's kindred and their influence with the young prince alarmed many of the older nobility and the king's councilors. By his last will Edward IV had left the care of his heir and of the kingdom to his brother Richard, duke of Gloucester, who forestalled the designs of the Woodville party by arresting Earl Rivers and Sir Richard Grey as they brought the young king from Ludlow to London for his coronation. Elizabeth Woodville took shelter again in the sanctuary of Westminster with her younger surviving son, Richard, and her five daughters. The king was lodged in the Tower (then still a royal residence as well as a prison). By a mixture of threats and cajolery the queen was induced to give up her younger son on Monday, June 16, to Gloucester, who took the boy to join his brother in the Tower. On June 22 a supporter of Gloucester, Ralph Shaw, preached a sermon at St. Paul's cross which challenged the validity of Edward IV's marriage, claimed that his children were all bastards and asserted that Gloucester was the rightful king. An assembly of lords and commons met at Westminster on June 25, Richard's claim was laid before it and a deputation from it urged Richard to assume the royal dignity. This he did and on June 26 the brief reign of Edward V came to an end.

For a short while afterward the two princes were seen playing together in the Tower garden, and then more rarely behind bars and windows; finally they disappeared forever. What happened to them will probably never be certainly known; much the most likely explanation is that they were murdered, probably in Aug. 1483, at the instigation of their uncle, although the responsibility for their death has been attributed to Henry Stafford, duke of Buckingham. It has also been suggested that they survived a further two years and were then murdered by the order of Henry VII. In 1674 workmen demolishing a staircase in the Tower discovered beneath the foundations a wooden chest containing the bones of two children. As the presumed remains of the two princes they were placed in an urn in Westminster abbey. The urn was opened in 1933 and the bones were examined by an eminent anatomist. His findings, and those of other anatomists and orthodontists who have inspected his numerous photographs, show that the age range of the material is consonant with the ages of the two princes in Aug. 1483, when they were widely rumoured to have been murdered.

See P. M. Kendall, *Richard the Third* (1955), for an appraisal and full bibliography. (A. R. M.)

EDWARD VI (1537–1553), king of England and Ireland from 1547 to 1553, was born at Hampton Court on Oct. 12, 1537, the only child of Henry VIII by his third wife, Jane Seymour. His mother died 12 days after his birth and he himself was a frail child who was never expected to have a long life. This did not prevent a strenuous education. Sir John Cheke, Sir Anthony Cooke and Roger Ascham all helped to teach him Latin, Greek and French, and by the age of 13 he had read Aristotle's *Ethics* in the original and was translating Cicero's *De philosophia* into Greek.

Edward was only nine when he succeeded his father as king (Jan. 28, 1547). At first the government was conducted by the duke of Somerset as protector, but factions soon developed around the king, each striving to control his person, not because of his personality but because ability to act in the king's name was indispensable to the wielding of effective authority. The protector's brother tried to bribe him with pocket money; John Dudley, earl of Warwick (soon to be duke of Northumberland), established a complete dominion over his mind by more subtle means, overthrew Somerset and then put Edward forward at the age of 14 as entitled to all the power of Henry VIII. But the king was only Northumberland's mask; of his personal influence on the course of history during his reign there is hardly a trace, though his youth enabled conservative men like Stephen Gardiner to maintain that the royal supremacy established over the church by Henry VIII was, or should be, in abeyance during a royal minority.

Edward's health began to fail by 1552 and the first signs of a rapid consumption were apparent in Jan. 1553. In May 1553 it was known that he was dying. He drew up a will which in its first draft excluded his half sisters Mary and Elizabeth, and indeed all females, from the throne and devised the crown to "the Lady Jane Grey's heirs male," Lady Jane being married to Northumberland's son Guildford Dudley. This draft was manipulated so as to read "the Lady Jane and her heirs male." For all this Northumberland was undoubtedly responsible, but Edward, with his own high ideas of the divine right of kings and the divine truth of Protestantism, seems to have thought himself both entitled and bound to override the succession as by law established and to exclude a Catholic heir. His last recorded words were vehement injunctions to Archbishop Thomas Cranmer to sign the will. He died at Greenwich on July 6, 1553, and was buried in Westminster abbey on Aug. 8.

The early age at which Edward VI died makes it impossible to form a confident estimate of his character and abilities. In some respects he manifested exceptional and precocious talent, but he was perhaps too much of a recluse to have become a successful ruler. His writings show awareness of the evils of the times but leave it uncertain whether he had the practical sagacity and the energy to remedy them. Moreover, he showed signs of all the Tudor obstinacy and was a zealot into the bargain, as no other Tudor was except Mary. Had he lived, the combination might well have involved England in disasters far greater than any that ensued from his premature death.

See also references under "Edward VI" in the Index volume.

Bibliography.—J. G. Nichols (ed.), *Literary Remains of King Edward VI*, 2 vol. (1857); A. F. Pollard, *England Under Protector Somer-*

set (1900), *Political History of England, 1547–1603* (1910); J. D. Mackie, *The Earlier Tudors, 1485–1558* (1952). (R. B. Wm.)

EDWARD VII (1841–1910), king of Great Britain and Ireland from 1901 to 1910, was born in London, at Buckingham palace, on Nov. 9, 1841, the second child and eldest son of Queen Victoria and Prince Albert. At birth he became duke of Cornwall and he was only a month old when his mother created him prince of Wales. He was christened Albert Edward. His childhood was happy and uneventful, with the austere surroundings of Windsor castle and Buckingham palace relieved as he grew older by the paradise of Osborne—the description is Queen Victoria's—and by highland life at Balmoral. He inherited his father's tastes for sports. Both parents were disappointed to notice that even in childhood the boy's interest in people and his love of chit-chat were far more strongly developed than strictly intellectual interests.

Education.—A rigorous educational program was devised. Till he was seven Edward remained in the hands of an accomplished governess, Sarah, Lady Lyttelton. Then, tutors were appointed—Henry Birch in 1849, F. W. Gibbs in 1851—in whose hands the prince was almost isolated from boys of his own age. In Nov. 1858 Edward entered his 18th year, at the end of which he would, for purposes of sovereignty, come of age. At the same time Gibbs's appointment was ended and Col. Robert Bruce, a somewhat pedantic Scottish disciplinarian, took over as the prince's governor. After some foreign travel with Bruce in Germany and the Mediterranean countries, Edward attended lectures at Edinburgh university for a short time prior to his entry into Oxford as an undergraduate (Oct. 1859) where he was admitted to Christ Church, although his parents declined to allow him to have rooms in college. He showed no greater aptitude for learning, but he enjoyed Oxford. In 1860 he undertook an important tour of Canada and the United States, staying with Pres. James Buchanan at the White House.

Edward returned to Oxford in Oct. 1860 and then transferred (Jan. 1861) to Trinity college, Cambridge, living outside the university at Madingley hall. During the long vacation (June–Sept. 1861) he served with the Grenadier guards in Ireland. He was summoned from Cambridge to Windsor castle to be present at the deathbed of his father on Dec. 14, 1861. The death of the prince consort was a dire misfortune for the prince of Wales, because his mother regarded the disclosure of his somewhat wild behaviour, when he was serving with the Grenadier guards, as having clouded the closing weeks of her husband's life. She apparently could not help contrasting Edward's character with Albert's. In consequence the queen never relied on her son, and allowed him no say either in the business of the nation or in the affairs of her court or family. The prince deserves credit for the long-suffering patience with which he bore his exclusion. Even when he was past 50 he had to endure snubs and criticism from the queen on official matters.

Marriage and Social Life.—Edward was married at Windsor on March 10, 1863, to Alexandra (*q.v.*), eldest daughter of Prince Christian (afterward King Christian IX) of Denmark. The bride had been selected for him after much family consultation, and although the tastes and interests of the prince and princess were by no means identical, the marriage was happy. The prince consort had arranged that the prince, when he came of age, should live at Marlborough house in London and at Sandringham in Norfolk. After a week's honeymoon at Osborne, the prince started that unvarying routine, which was not to be broken till he became king nearly 40 years later: London for the season, Norfolk for shooting, Christmas and holidays, Abergeldie (near Balmoral) for a part of the autumn, and (as he grew older) an increasing number of visits to relations and friends abroad. The prince and princess of Wales had five children who reached maturity—Albert Victor, duke of Clarence (d. 1892), George (afterward King George V), Louise (afterward princess royal and duchess of Fife), Victoria and Maud (afterward queen of Norway).

The princess and her children formed a singularly close knit family circle; their life was centred on Sandringham which they loved dearly, but the prince had many friends and interests beyond that flat, bleak countryside of west Norfolk. He took a leading part in the social life of London and equally enjoyed the less formal country house visits where were transacted some of the politics and where was exerted much of the influence of late Victorian life. Nor did the escapades of social life by-pass the prince. He gave evidence in the divorce case brought by Sir Charles Mordaunt against Lady Mordaunt in Feb. 1870, and it was well known that the two men cited as corespondents were his close friends. Naturally his guilt was widely believed, though he was not directly implicated, and he wrote to his mother of his indignation at the "gross imputations" made against him. He had to face some booing both on the race course and in one of the London theatres. Twenty-one years later he again appeared in the witness box to give evidence in the celebrated baccarat or Tranby Croft case where one of his companions in that game of chance was detected cheating. If these were the hazards of the prince's easy social life there was another side to it. Few men had a surer knowledge of human nature and a shrewder grasp of what was happening. The prince hardly ever read a book; his letters are perfunctory trifles, but it was Sir Charles Wentworth Dilke, with advanced opinions on home and imperial affairs, who noticed that "it is worth talking seriously to the prince."

Through his family connections Edward was welcome at most of the courts and capitals of Europe and he became in fact something of a free-lance diplomatist. He was the first heir to the throne to visit India, touring the country in the winter of 1875–76. He was outspokenly critical of the treatment of Indians by a certain type of British official. He was a careful student of housing conditions in London and himself served on the royal commission on housing appointed by the Liberal government in 1884. Although the prince was a familiar figure on the race course (he won both the Derby and St. Leger with Persimmon in 1896, repeating these successes with Diamond Jubilee in 1900) and at the Royal Yacht squadron at Cowes or at one of the huge shooting parties which delighted the hearts of Victorian sportsmen, these were recreations which he intended should not diminish his stature as a public personality. His mother had excluded him from any recognized place in the machinery of government, yet he was by no means ill-equipped for the sovereignty of the British empire which he assumed at her death on Jan. 22, 1901.

Accession.—The king felt that his first task was to restore to the crown some of the traditional splendour (and authority) which had lapsed during the 40 years of his mother's widowhood. Buckingham palace, which had scarcely been used since 1860, was refurbished; parliament, in spite of the general mourning for the queen, was opened in conspicuous state by the king in Feb. 1901. Plans went ahead for the elaborate ceremony of the coronation which was fixed for June 26, 1902. Two days beforehand, when London was filled with distinguished foreign and imperial visitors, the king had to undergo an immediate and dangerous operation. He recovered and the coronation (though shorn of some of its glory) was held on Aug. 9. The coronation honours list published on the day originally fixed for the ceremony included a new order of chivalry—the Order of Merit—to mark distinguished personages in all walks of life. This was the king's idea, and the nomination to the Order was kept in his own hands.

King Edward's interest in affairs at home was certainly less than his concern with foreign affairs, and this may explain why some critics thought that he handled state business somewhat perfunctorily. Though he was broadly Liberal in his sympathies—unlike his mother he was always a warm admirer of Gladstone—and he greatly preferred his Liberal prime minister, Sir Henry Campbell-Bannerman, to the Conservative Arthur Balfour, the king was apprehensive of violent change and for this reason deprecated the somewhat extravagant speeches of David Lloyd George and Winston Churchill when they were ministers in the Liberal cabinet at the end of his reign. He was never unmindful of the possibility of a European war—a possibility which made him critical of what he regarded as too costly schemes of social reform. He wholeheartedly endorsed the great schemes of army reform undertaken at the war office by R. B. (afterward Lord) Haldane. "He supported me strenuously" were Haldane's own words.

Interest in Foreign Affairs.—Edward's reign will rightly be

remembered for the effort he made to strengthen the position of Great Britain in Europe. Feeling against Britain was strong, especially in France, Germany and the Netherlands, as a result of the South African War, but peace with the Boers was made in 1902, and in 1903 Edward traveled to Lisbon and Rome, returning through Paris, where the dislike for all things British was at its most vocal. In a public speech the king referred to his continuing devotion to Paris which dated back to the visit he paid to that city with his parents in 1855 when the emperor Napoleon III was at the height of his power. A superbly delivered speech in French—the king spoke French and German perfectly and had a tolerable command of Italian—at a state banquet at the Elysée palace, opened a new era in Anglo-French relations. King Edward certainly did not create the entente cordiale, but his visit to Paris revealed to the world new feelings of friendship which were developing between the two peoples. His most controversial visit—for which he was criticized by the left wing in parliament—was to Russia in 1908. The king's relations with his nephew, the German emperor, were personally not good but there was never any question of excluding the emperor from the friendship of Great Britain, and in fact Edward paid more visits to his nephew in Berlin than to any other European sovereign. Apart from any natural aptitude for diplomacy, Edward VII was helped by being nearly related to all the ruling houses of Europe except those of Austria, Italy and Spain. The latter was brought into the English circle in 1906 by the marriage of the king's niece to Alfonso XIII. How much England, under Edward, had become the centre of the European network of monarchy is suggested by the fact that 8 reigning sovereigns or their consorts and 16 other royal personages had luncheon at Windsor castle on Nov. 17, 1907.

Character and Death.—The familiar impression of the king recorded in countless photographs revealed a man of fashion with his well-trimmed beard, neat boots and the top hat with its suggestion of a curling brim. On the race course and at countless public functions they seemed almost the insignia of royalty and identified the king. Equally familiar are the photographs of him at the centre of a country house party of enormous proportions surrounded by the leading members of the aristocracy, or after a shoot with a truly royal bag laid out before him. But these gay pictures did not reveal the whole man. He was deeply worried by the international situation. These anxieties depressed him and explained why toward the close of his reign he began to speak of abdication. Nor did the public, seeing the debonair outward man, realize that his health was not robust. From 1906 his doctors were concerned about him, and it was noticed that he fell asleep in public during the state visit which he paid to Berlin in the spring of 1909. The public was stunned by the news of his serious illness on May 5, 1910, at Buckingham palace; he died the following evening. Edward VII was an immensely popular sovereign—frowned on perhaps by the middle classes but admired in the world of fashion and beloved by the ordinary English people. The secret of his personal sway lay in his personality—neatly described by his admirer Haldane as "so strong and direct."

BIBLIOGRAPHY.—Sir Sidney Lee, *King Edward VII*, 2 vol. (1925–27); Virginia Cowles, *Edward VII and His Circle* (1956); Lord Redesdale, *King Edward VII. A Memory* (1915). *See* also F. J. C. Hearnshaw (ed.), *Edwardian England* (1933); *Journals and Letters of Reginald, Viscount Esher*, 4 vol., ed. by M. V. Brett and Oliver, Viscount Esher (1934–38); Sir F. Ponsonby, *Recollections of Three Reigns* (1951).

(R. T. B. F.)

EDWARD VIII (1894–), king of Great Britain and Northern Ireland from Jan. 20 to Dec. 11, 1936, the eldest child of George, duke of York, afterward King George V, and of Princess Mary of Teck, afterward Queen Mary, great-grandson of Queen Victoria, was born at White lodge in Richmond park on June 23, 1894. He was christened Edward Albert Christian George Andrew Patrick David by Edward Benson, archbishop of Canterbury, in the drawing room at White lodge on July 16, 1894.

His extreme youth was largely passed in Norfolk, in a small house on the Sandringham estate known as York cottage. Neither of his parents believed in an easygoing environment for their children, and he was strictly brought up, although a gayer atmosphere was introduced when his grandparents, King Edward VII and Queen Alexandra, came to Sandringham. The prince passed successfully into the Royal Naval college, Osborne, at the beginning of 1907, and he moved on to Dartmouth college in 1909. The death of his grandfather on May 6, 1910, meant that he became at once heir to the throne and duke of Cornwall, and that he would have to forgo the North American training cruise which was the conclusion of the Dartmouth course. He was admitted to the Order of the Garter on June 10, 1911, and was able to wear the robes at his father's coronation on June 22; he was formally invested as prince of Wales at Caernarvon castle in July 1911.

Prince of Wales.—After these ceremonies the prince was allowed to serve on the battleship "Hindustan" for three months and then he was sent abroad (1912) to learn something of the languages and politics of Europe by staying in France and Germany. He went up to Oxford university in Oct. 1912, and was allowed to lead a comparatively normal undergraduate life at Magdalen college. Although the president of his college once observed that he would never be "bookish," he absorbed some learning, especially from his tuition in constitutional law by Sir William Anson, the warden of All Souls' college, and he made friends and enjoyed himself. He was occasionally called from Oxford to help his parents entertain distinguished visitors, and after one such function he confided to his diary "what rot and a waste of time, money and energy all these state visits are." Immediately after the outbreak of World War I he was commissioned in the Grenadier guards (Aug. 6, 1914), and on Nov. 16 went to France as aide-de-camp to the commander in chief, Sir John French. In Sept. 1915 he was appointed to the staff of Lord Cavan, who was in command of the then recently formed guards division. He felt acutely his exclusion from the front line and the fighting but Earl Kitchener, the secretary of state for war, could not risk his being taken prisoner. He was sent to Suez in 1916, returning to the western front later in that year. He went with Lord Cavan to the Italian front immediately after the disaster of Caporetto (Oct. 1917) but was back in France for the closing weeks of the war.

After the war the prince began the round of royal duties which was familiar to his father and to the older generation of courtiers; he received the freedom of cities, opened new buildings and made conventional speeches at public banquets. However, these tasks were varied by a remarkable series of imperial tours, to Canada (1919), to New Zealand and Australia (1920), to the Mediterranean, India and Ceylon and the far east (1921–22). He traveled again to Canada and the United States (1924) and to South Africa (1925). King George V was a little apprehensive of the informal behaviour of the prince on these imperial visits, but he readily admitted their remarkable success.

In the ten years between the close of his imperial tours and the death of his father, the prince took his share in the official life and public ceremonies of the country, also making visits throughout the British Isles which were intended to encourage industry and local enterprise. During the 1920s the prince's most frequent recreations were hunting and riding. He enjoyed point-to-point racing and steeplechasing but after one or two falls he stopped in deference to the opinions of the king and the prime minister. He then took to golf with enjoyment, and by persistence became a reasonably handicapped player. He learned to fly his own aircraft (1929), and later purchased a Gypsy Moth aircraft for his personal use. In London the prince lived in York house, a rather small private house in St. James's palace. In 1930 King George gave him Fort Belvedere, an 18th-century house belonging to the crown near Sunningdale. The Fort, as he always called it, gave him privacy and the sense of making a home that was entirely his own. He worked arduously in the garden and woodlands, becoming in the 1930s something of an authority on horticulture, especially on the growing of roses. He soon began to regard the Fort as a refuge from the official world which he increasingly disliked. There he entertained a private circle of friends, not drawn from the conventional aristocracy in which his father and grandfather had moved. During these years his popularity rivaled, if it did not exceed, that of his grandfather King Edward VII when he was prince of Wales. He seemed to attach to himself, especially at the great Toc H ceremonies in the Albert hall on Armistice day,

the ideals and anguish of the generation of World War I.

King Edward VIII.—His father died after a short illness on Jan. 20, 1936, and with traditional splendour the prince was proclaimed king on the following day in St. James's palace, as King Edward VIII. The contrast between the reign which had ended and the new one which was beginning was made clear by the new king in his first broadcast on March 1. After a long and keenly felt tribute to King George he interpolated into the speech a personal passage referring to the opportunities he had had to get to know the people of almost every country in the world; he emphasized that although he spoke as king he was still the same man who had had that experience. Outwardly the reign moved smoothly, its formal ceremonies reduced by mourning. The summer was marked by an alarming incident when a loaded revolver was pointed or thrown at the king as he was returning from the presentation of new colours in Hyde park, and by the great ceremony in France when he unveiled the Vimy Ridge memorial. There was some criticism of the king for supposed economies in the royal residences, for informality and for curtailing the long established ritual of presentation parties at court. He moved into Buckingham palace during the autumn.

The Abdication.—The king had intended to spend a holiday in the south of France but because of the unsettled state of that country he went instead for a cruise in the Adriatic. He was accompanied by a private circle of friends which included Mrs. Wallis Warfield Simpson, concerning whose association with the king there was much lurid comment in the continental and U.S. press.

This comment caused concern throughout the summer and early autumn both to the royal household and to the cabinet. This was increased when it was known that Mrs. Simpson was about to divorce her husband at the Ipswich assizes in October. The divorce was heard and the decree nisi granted on Oct. 27. Seven days previously Stanley Baldwin, the prime minister, asked for an audience with the king. He explained the embarrassment and difficulties which were being caused by the king's private friendship but the two men parted under a degree of misapprehension—the prime minister thinking that he had made some impression, the king feeling that he had convinced the prime minister that his friendship was a personal matter unrelated to the state. Thus matters drifted until Nov. 13, when the king received a letter from his private secretary, Maj. A. H. L. (afterward Lord) Hardinge, warning him of the dangers into which he was running and urging that Mrs. Simpson should forthwith go abroad. On Nov. 16 the king again saw the prime minister, told him—and this had not been clearly stated at the interview in October—that he hoped to marry Mrs. Simpson and added that if he could not marry her and remain king, he was "prepared to go." During the next ten days the prime minister was engaged in preparing the cabinet, the press, parliament and the dominions for the abdication of the king which he had rightly told the king was "grievous news." However, the situation was complicated during the closing days of November by a proposal of Lord Rothermere, strongly supported by other influential men outside the prime minister's established political circle, that the king should marry Mrs. Simpson, withholding from her the dignity of queen consort. It is possible that if opinion could have been fairly polled on this proposal, which had respectable constitutional origins, considerable support for it might have been forthcoming. It was, however, doomed by being somewhat hurriedly and forcibly put to the dominions and by the explosion of the whole matter in press and parliament on Dec. 3. On the following day the word "abdication" appeared in the newspapers for the first time, but the king hesitated while Winston Churchill and other members of parliament urged that he should not be hurried. On the reassembly of the house of commons, after a week end for reflection, the support for the prime minister was overwhelming and it became clear to the king that he could only carry out his wishes by reigning over a divided nation. Mrs. Simpson had left England on Dec. 3, and on Dec. 8 she published a statement that she was prepared to withdraw from a position which had become untenable. Lord Beaverbrook had been trying to urge this course. He felt that, if the decision could be postponed and public excitement allowed to subside, it might later become possible for the king to achieve his marriage without having to abandon his throne. But this came too late to affect events and on the following day the prime minister announced that on Dec. 10 he would make a formal statement. This proved to be a message from the king announcing that he had that morning signed an instrument of abdication. This was passed by parliament on the morning of Dec. 11, unanimously except for a republican gesture by the Independent Labour party. It is right to recall an obiter dictum of Baldwin's to the effect that throughout those critical days "the king could not have behaved better than he did." On the same evening the former king broadcast to the nation explaining the reasons which had compelled him to abdicate—"I now quit altogether public affairs, and lay down my burden." This broadcast is believed to have had a larger audience than any other ever heard in Great Britain. That night on board the destroyer "Fury" he left Portsmouth for France.

Duke of Windsor.—The new king, George VI, created Prince Edward duke of Windsor at his accession council on Dec. 12. The duke spent the first months of his exile in Austria and was married to Mrs. Simpson by a clergyman of the Church of England at the Château de Candé, France, on June 3. A few days previously the new king, on the advice of the cabinet, had accorded the duke the right to the title of royal highness for himself alone; the duchess was thereby excluded from enjoying her husband's rank—a decision which was deeply wounding to the duke and the legality of which has been questioned. For the next two years the duke and duchess lived mainly in France, visiting various other European countries including Germany (Oct. 1937), where the duke had an interview with Adolf Hitler. The outbreak of World War II failed to close the breach between the duke and his family and, after visiting London, he accepted a position as liaison officer with the French. On the fall of France he traveled to Madrid where he was subjected to a fanciful plan of the Nazis to use him against the established government in England. When he reached Lisbon he was offered by Winston Churchill the governorship of the Bahamas, a British colony in the West Indies, and he remained there for the duration of the war. After 1945 he lived in France, making occasional visits to the U.S. and London. He attended the funerals of his brother George VI in 1952 and of his mother Queen Mary in 1953. His memoirs, *A King's Story*, were published in 1951, and those of the duchess of Windsor, *The Heart Has Its Reasons*, in 1956. (R. T. B. F.)

EDWARD (DUARTE) (1391–1438), king of Portugal from 1433 to 1438, was the son of John I of Portugal. His mother was Philippa of Lancaster, daughter of John of Gaunt, duke of Lancaster. A writer and philosopher, Edward promulgated some notable reforms including the so-called *lei mental* (April 8, 1434) which was intended to facilitate recovery of the many grants made by the crown under John I by establishing that grants of royal property were inalienable and indivisible and must revert to the crown in default of a legitimate male heir. An attempted expedition against Tangier in 1437 was a military disaster and the king's brother Fernando was taken prisoner and was to die in captivity in 1443. Edward died at Tomar on Sept. 9, 1438, and was succeeded by his son Afonso V. (V. R. R.)

EDWARD, THE BLACK PRINCE (1330–1376), also known as Edward of Woodstock, the eldest son of Edward III and Philippa of Hainaut, was one of the most famous commanders in the Hundred Years' War. His sobriquet, the "Black Prince," said to have come from his wearing black armour, has no contemporary justification and is found first in Richard Grafton's *Chronicle of England* (1569).

Born at Woodstock on June 15, 1330, Edward was created earl of Chester (March 1333), duke of Cornwall (Feb. 1337)—the first appearance of this rank in England—and prince of Wales (May 1343); he was prince of Aquitaine from 1362 to 1372. His first campaign was served under his father in northern France (1346–47) and at the battle of Crécy (Aug. 26, 1346) he won both his spurs and the famous ostrich plumes and with them the mottoes used by himself and subsequent princes of Wales, *homout; ich dene* ("Courage; I serve"; the words are here spelled as Edward himself wrote them; later variants include *houmout* and *ich dien*

or *ich diene*). One of the original knights of the Garter, he was sent to France with independent command (1355), winning his most famous victory over the French at Poitiers (Sept. 19, 1356). The French king John II, brought captive to England, was treated by the prince with a celebrated courtesy, but he was obliged to pay a ransom of 3,000,000 gold crowns and to negotiate the treaties of Bretigny and Calais (1360) by which Aquitaine was ceded to the English.

Edward married his cousin Joan, the divorced and widowed countess of Kent (Oct. 1361). He was created prince of Aquitaine in July 1362 and left England in 1363 to take up his duties. His powers and his opportunities were great but his rule was a failure, and he was personally largely to blame. His court at Bordeaux, that of a foreign conqueror, was extravagant; the 13 *sénéchaussées* into which the principality was divided administratively followed their earlier French pattern and allowed local French loyalties to subsist; his relations with the many bishops were unfriendly, while the greater nobles, Arnaud Amanieu, sire d'Albret, Gaston II, comte de Foix, and John I, comte d'Armagnac, were hostile. He summoned several estates or parliaments, but always to levy taxes. In 1367 he undertook to restore Pedro the Cruel of Castile to his throne. This was Edward's most unstatesmanlike act, for though he won his classic victory at Nájera (April 3, 1367), the campaign ruined his health, his finances and any prospect of sound rule in Aquitaine, where in 1368 the nobles and prelates appealed against him to Charles V of France as suzerain. Edward's reply to the French king's citation to answer the appellants before the *parlement* of Paris in May 1369 is well known—he would appear with 60,000 men at his back. He had, however, alienated the towns and peasantry as well as the nobles, and by March 1369 over 900 towns, castles and strong places had declared against him. Relying on mercenaries whom he could not afford to pay, he was powerless to quell the revolt and the terrible sack of Limoges (Oct. 1370) merely redounded to his discredit. He returned to England a sick and broken man (Jan. 1371) and formally surrendered his principality to his father in Oct. 1372, alleging that the revenues of the country were insufficient to defray his expenses. He had no successor as prince of Aquitaine.

Edward's position in England, where throughout his life he was heir apparent, was that of a typical 14th-century magnate. The registers of his household from 1346 to 1348 and from 1351 to 1365 have survived and add to what we know of him from the chroniclers and from his biographer, the herald of Sir John Chandos. In one important respect all these sources paint the same picture, that of a man constantly living beyond his means. However, his generosity extended to his tenants as well as to his knightly companions, and faithful service was rewarded, as in 1356 when the ferry of Saltash was granted to William Lenche who had lost an eye at Poitiers.

The prince paid two visits to Chester (1353, 1358). Cheshire furnished many of his archers who wore a rudimentary uniform of a short coat and hat of green and white cloth with the green on the right. Wales he never visited, but he was in Cornwall in 1354, and spent Christmas 1362–63 at Restormel castle. Near London, when not at his castle at Berkhampsted, he stayed at his manors of Kennington or (until 1362) Vauxhall. His wardrobe was near Ironmonger lane and his exchequer at Westminster. At his death his lands yielded roughly £8,600 a year—his Welsh principality about £4,700; Chester £1,300; Cornwall about £2,300 and the scattered English possessions about £300. His council and household officials were men of ability, particularly his receiver-general, Peter Lacy.

Edward appears to have shared the interests of his class, jousting, falconry, hunting, gaming. He was literate and conventionally pious, substantially endowing a religious house at Ashridge (1376). He had the customary fine presence of the Plantagenets and shared their love of jewels. The Black Prince's ruby in the present imperial state crown may or may not have been given to him by King Pedro after the battle of Nájera, but he would certainly have prized it as a connoisseur. Similar artistic interest is shown in his seals adorned with their ostrich feathers, and in the elegant gold coins which he issued as prince of Aquitaine.

The last five years of the prince's life are obscure. Some contemporaries suggest that he supported the commons when political discontent culminated in the Good parliament (April 1376), but he knew he was dying and he was probably seeking the best means to ensure the succession of his second, but only surviving son, Richard of Bordeaux (afterward Richard II). He died at Westminster on June 8, 1376, and was buried at Canterbury, where his tomb with his accoutrements still stands.

BIBLIOGRAPHY.—M. K. Pope and E. C. Lodge (eds.), *Life of the Black Prince by the Herald of Sir John Chandos* (1910); *Register of Edward the Black Prince*, 4 vol. (1930–33); T. F. Tout, *Chapters in Administrative History*, vol. v (1930); *R. Delachenal, Histoire de Charles V*, vol. iv (1928); Sir Israel Gollancz, *Ich dene* (1921).

(I. P. S.)

EDWARD, LAKE, situated in one of the tectonic troughs which border the east of the Republic of the Congo, lies a few miles south of the equator at an altitude of 2,992 ft. It is oval and about 47 mi. in length, with a maximum breadth of about 32 mi. On the northeast it is connected by the Kazinga channel, about 20 mi. long, with the smaller Lake George (or Dweru) which is crossed by the equator. The combined area of the two lakes is 970 sq.mi. Two-thirds of Lake Edward are in the Congo and one-third is in Uganda. The Congolese part of the lake is incorporated in the Albert National park, the rest in the Queen Elizabeth National park.

In the south the Rutshuru river forms the principal affluent of Lake Edward, and gathers the waters which descend from the Virunga volcanoes. This and the other affluents have already silted up much of the lake; they flow through a wide plain, marshy in places. Lake George receives the waters of the Ruwenzori river and drains into Lake Edward. Lake Edward empties northward into Lake Albert through a single outlet, the Semliki river, which flows through a plain situated at the bottom of a tectonic trough and is a headstream of the White Nile (*see* NILE). The banks are in many places fault formations and are therefore generally abrupt except in the south and north (the Rutshuru and Semliki plains) and along part of the east coast where low terraces of dry steppe scattered with treelike euphorbia rise between the lake and the watershed of the trough. On the west the edge of the trough exceeds 10,000 ft. in altitude. About 20 mi. N. the Ruwenzori (*q.v.*) massif rises to nearly 16,800 ft. The waters of Lake Edward are clear, light green in colour and brackish. Fish, waterfowl, crocodiles and (in the southern swamps) hippopotamuses abound. In the dry season the lake is often overhung by a thick haze and in the rainy season is subject to violent storms.

Lake George was discovered in 1875 by H. M. Stanley, then traveling westward from Uganda, and by him was named Beatrice gulf in the belief that it was a part of Lake Albert. In 1888–89 Stanley, approaching the Nile region from the west, traced the Semliki to its source in Lake Edward, which he discovered, naming it after Albert Edward, prince of Wales, afterward Edward VII. Stanley also discovered the Kazinga channel. The accurate mapping of the lake (which, however, shrank considerably in the first half of the 20th century) was mainly the work of British officials and travelers such as Sir F. J. D. Lugard and Sir Harry Johnston (*qq.v.*), while Emin Pasha (*q.v.*) and the German zoologist Franz Stuhlmann first explored its southern shores. Lake Edward was found, by the Anglo-German boundary commission of 1902–04, to lie within the sphere of influence of the Congo Free State as defined in the agreement of May 12, 1894, between that state and Great Britain. *See* also AFRICA: *Physical Geography: Geology;* GREAT RIFT VALLEY.

(HE. NI.)

EDWARDS, ALFRED GEORGE (1848–1937), first archbishop of Wales, was born at Llanymawddwy, Merionethshire, on Nov. 2, 1848, the younger brother of Henry Thomas Edwards (*q.v.*), and graduated from Jesus college, Oxford, in 1874. After a successful headmastership of Llandovery college, he became vicar of Carmarthen in 1885, where he began the great work of his life, the defense and strengthening of the Welsh national church. An ardent believer in establishment, he fought his church's many enemies on every ground of fact and principle by sermons, politics and publications, of which the most important were *The Truth About the Church in Wales* (1889) and *Land-*

marks in the History of the Welsh Church (1912). In 1889 he was appointed bishop of St. Asaph, and from the disestablishment in 1920 until his retirement in 1934 he was archbishop of Wales. He died at St. Asaph, July 22, 1937.

See G. Lerry, *Alfred George Edwards, Archbishop of Wales* (1940). (G. F. A. B.)

EDWARDS, HENRY THOMAS (1837–1884), clergyman of the Church of Wales who led a largely successful movement for an increase in the Welsh character of his church, born at Llanymawddy, Merionethshire, on Sept. 6, 1837, was successively vicar of Aberdare (1866), vicar of Caernarvon (1869) and dean of Bangor (1876). His open letter to Gladstone "On the Church of the Cymry" (1870) attracted much notice; in it he remarked that the spread of nonconformity in Wales was largely due to "the withering effect of an alien episcopate." Deeply convinced of the advantages of religious education and national churches, he promoted his cause at Bangor through an excellent clerical education society, aiming to supply clergymen able to speak Welsh. Worn out by incessant labours, his restless mind began to give way in 1883 and he committed suicide in his brother's vicarage at Ruabon, May 24, 1884. (G. F. A. B.)

EDWARDS, JONATHAN (1703–1758), American theologian, was born Oct. 5, 1703, at East Windsor, Conn. His father Timothy was pastor of the church there; his mother Esther was a daughter of Solomon Stoddard, pastor of the church at Northampton, Mass. Jonathan was the fifth child and only son among 11 children; he grew up in an atmosphere of Puritan piety, affection and learning. After a rigorous schooling at home, he entered Yale at 13. He was graduated in 1720 but remained at New Haven for two years studying divinity. After a brief New York pastorate (1722–23), he received his M.A. in 1723; during most of 1724–26 he was a tutor at Yale. In 1727 he became his grandfather's colleague at Northampton. The same year, he married Sarah Pierrepont, who combined a deep, often ecstatic, piety with personal winsomeness and practical good sense. To them were born 11 children.

As a precocious child of 9 or 10 years, he composed a brief paper on the soul. His essay on the flying spider, probably written shortly before he went to college, exhibits his remarkable powers of observation and analysis. Another on the rainbow, written about the same time, shows that he had already mastered Newton's optical theories. He habitually studied with pen in hand, recording his thoughts in numerous hand-sewn notebooks; one of these, a "Catalogue" of books, evidences the wide variety of his reading.

Under the influence of Newton, Locke (whose *Essay* he had read in his sophomore year), the Cambridge Platonists and Reformed divines, the young theological student sketched in his manuscripts the outlines of his philosophical theology, stating the doctrines of Calvinism in terms of contemporary philosophy. Absolute Nothing is inconceivable, therefore something exists: Being which is eternally everywhere. But since existence is existence only for consciousness, the universe depends for its being every moment on the intelligence and will of God, and "spirits only are properly substance." Atoms themselves are only units of resistance maintained by God's power and formed by his idea. Reality is therefore a spiritual universe contained in the mind of God and realized as human experience by the direct action of the divine will. This is true for the knowledge of God given by grace, as well as for physical perception. Being all-sufficient fullness, God must communicate himself by exercising all his attributes. The happiness of his creatures is indeed God's end in creation, but that happiness consists in contemplating and rejoicing in God's glory as manifested in creation and redemption. Intelligent beings are highest in the great chain of being, but their vision of God is the completing link. Only those beings which "consent" with the divine Being (by "love to Being in general") can be said to have positive being; those which dissent from that Being sin infinitely and deserve eternal punishment. Yet God reveals his glory even in punishing sinners, and evil contributes to the grand design. Cause and effect hold throughout the universe, but the connection between them is a "constituted" one, continued at the sovereign pleasure of God; this applies also to the successive moments of a person's existence. Though grace is independent of human means, means as well as ends are objects of the decrees, and God has so constituted things that means (*e.g.*, sermons, sacraments, even the fear of hell) are the "matter" upon which grace acts; grace is therefore experienced in the cause–effect realm as men devote themselves to the worship and service of God. Thus the predestinarian preacher could appeal to the emotions and wills of men.

The peculiar dynamic of Edwards' theology came from his own experience of God. He progressed in turn from preoccupation with his own salvation to an intellectual "conviction" of divine sovereignty, and thence (in a moment of Scripture meditation) to a "new sense" of and "delight" in God's glory revealed in Scripture and nature. This became the centre of Edwards' piety: a direct, intuitive apprehension of God in all his glory, a sight and taste of Christ's majesty and beauty far beyond all "notional" understanding, immediately imparted to the soul (as the 1734 sermon title puts it) by "a divine and supernatural light." This alone confers worth on man, and in this consists his salvation. What such a God does must be right; hence Edwards' cosmic optimism. The acceptance and affirmation of God as he is and does, and the love of God simply because he is God, became the positive elements in all Edwards' preaching.

At Stoddard's death in 1729, Edwards became sole occupant of the Northampton pulpit, the most important in Massachusetts outside of Boston. In his first published sermon, preached in 1731 to the Boston clergy and significantly entitled *God Glorified in the Work of Redemption, by the Greatness of Man's Dependence upon Him, in the Whole of It*, Edwards blamed New England's moral ills on its assumption of religious and moral self-sufficiency. Since God is the saints' whole good, faith, which abases man and exalts God, must be insisted on as the only means of salvation. His sermons on "Justification by Faith Alone," delivered in Nov. 1734, were also directed against that Arminianism which was becoming so congenial to the colonists' enterprising spirit (*see* ARMINIANISM). These were immediately followed by a great revival in Northampton, in the winter and spring of 1734–35, during which more than 300 made professions of faith. Edwards' *Faithful Narrative of the Surprizing Work of God* (1737) which he had witnessed made a profound impression in America and Europe, particularly through his description of the types and stages of conversion experience and (less fortunately) his accounts of some of the more "surprizing" cases.

In 1740–41 came the "Great Awakening" throughout the colonies, led by men like George Whitefield and Gilbert Tennent, whose "pathetical" sermons to huge crowds resulted in violent emotional reactions and mass conversions. Edwards himself, though he held his own congregation relatively stable, employed the "preaching of terror" on several occasions, as in the Enfield sermon, *Sinners in the Hands of an Angry God* (1741). The Awakening produced not only conversions and changed lives; there were excesses, disorders, ecclesiastical and civil disruptions. While increasingly critical of the "experiences" and practices associated with the revival, to the extent of personally rebuking Whitefield, Edwards maintained that it was in essence a genuine work of God, to be furthered and purified. In defense and criticism of the revival Edwards wrote *The Distinguishing Marks of a Work of the Spirit of God* (1741), *Thoughts on the Revival* (1742) and *A Treatise Concerning Religious Affections* (1746), the last a classical delineation of the Christian life as it revolves around the love of God. His last explicit defense of revival religion was the "Reflections" in his edition of the memoirs of David Brainerd (1749), a young New Light (revivalist) missionary to the Indians, who died in 1747. (*See* also REVIVALISM.)

Meanwhile, Edwards' relations with his own congregation had become strained for a number of reasons, among which was Edwards' changed views on the requirements for admission to the Lord's Supper. To the "Half-Way Covenant," whereby baptized but unconverted children of believers might have their own children baptized by "owning the covenant," Stoddard had instituted the soon widespread practice of admitting to the Eucharist all who were thus "in the covenant," even though they knew themselves to be unconverted. Edwards had come to believe that a

profession of the essentials of a Christian experience, not merely doctrinal knowledge and a moral life, was necessary for admission to full communion. The public announcement of his position in 1749 precipitated a violent controversy which was climaxed by his dismissal. On July 1, 1750, Edwards preached his dignified and restrained *Farewell Sermon*. In this controversy Edwards wrote two books, the *Qualifications for Communion* (1749) and a *Reply to Solomon Williams* (1752), pastor at Lebanon, Conn., and a relative of Edwards. Though Edwards himself was defeated, his position finally triumphed in New England and facilitated the separation of church and state after the Revolution.

In 1751, Edwards became pastor of the frontier church at Stockbridge, Mass., and missionary to the Indians there. Hampered by language difficulties, illness, Indian wars and conflicts with powerful personal enemies, he nevertheless discharged his pastoral duties and found time to write his famous work on the *Freedom of the Will* (1754), in which he attacked the notion of a self-determining will. According to Edwards, the Arminians considered the will a separate faculty, capable of determining its own volitions in the face of the strongest contrary motives. After dismissing the self-determining will as a logical absurdity, Edwards defined volition as merely the realization in act of the soul's "prevailing inclination" and hence determined by "the greatest apparent good." Man is free to do as he pleases, yet his choices occur in causal series and are foreseen and foreordained by God. Man is not free to do otherwise than he pleases, yet he is still morally accountable; for it is the nature rather than the cause of volitions which renders them objects of moral judgment.

By 1757 Edwards had finished his *Original Sin* (1758); this was mainly a reply to the English divine, John Taylor, of Norwich, whose works attacking Calvinism had "made a mighty noise in America." Edwards defended the doctrine on empirical and Biblical grounds, and offered his theory of "constituted identity" to account for the unity of mankind with Adam, a unity, however, not in Adam's sinful act, but in his prior sinful disposition. Edwards perceived the threat in Taylor's notion of man's innate goodness and autonomy; the whole Christian conception of supernatural redemption was at stake. He therefore planned further treatises, of which he completed the posthumously published dissertations on *The End for Which God Created the World* and *The Nature of True Virtue*. He also projected books on other subjects, notably a "History of the Work of Redemption," a complete theology combining Biblical, historical and systematic materials "in an entire new method."

Late in 1757, Edwards accepted the presidency of the college at Princeton, N.J. (later Princeton university), and arrived there in January. He had hardly assumed his duties when he contracted smallpox from an inoculation and died at Princeton on March 22, 1758.

Edwards was the greatest theologian of American Puritanism, perhaps the greatest America has produced. For sheer speculative power and magnificence of conception he has not been surpassed by any American philosopher. Though he did not live to publish his great systematic work, enough remains to show that he had achieved a remarkable synthesis of Reformed theology with the Newtonian world-view, Lockean empiricism with Augustinian illuminism, the Christian "plan of salvation" and concept of history with Platonic idealism and Neoplatonic emanationism. His influence on American religion was tremendous. His disciples not only created the "New England theology" but helped recover for Christianity the intellectual leadership of postrevolutionary America. By his writings and example, he gave impetus to the infant evangelical missionary movement. As chronicler, preacher, apologist and critic of the evangelical awakening, he influenced its patterns and practices, helped correct its extravagances, and wedded it to heart worship of God and intense moral earnestness.

See also CONGREGATIONALISM; *United States*.

BIBLIOGRAPHY.—Clarence H. Faust and Thomas H. Johnson (eds.), *Jonathan Edwards, Representative Selections*, best handbook, extensive bibliographies (1935); Ola Elizabeth Winslow, *Jonathan Edwards*, best biography (1941); Robert E. Spiller, *et al.* (eds.), *Literary History of the United States*, an excellent annotated bibliography of editions and studies, vol. 3, pp. 481–485 (1948); Perry Miller, *Jonathan Edwards*, a brilliant interpretive study (1949); Perry Miller (ed.), *The Works of Jonathan Edwards*, a complete critical edition (1957–). (T. A. SR.)

EDWARDS, LEWIS (1809–1887), minister of the Calvinistic Methodist Church of Wales and educationist, was born Oct. 27, 1809, at Pen-Llwyn, Cardiganshire. With slender resources he sought a more extensive education than was then available within Wales. His church accepted him as a preacher in 1829 and he went to Edinburgh university in 1833 where he was befriended by John Wilson ("Christopher North") and permitted to graduate in three years. He was ordained in 1837 and married the granddaughter of Thomas Charles (*q.v.*) of Bala. With his brother-in-law David Charles he opened a school at Bala, to prepare men for the ministry, which became in 1867 the theological college for his church in north Wales. Through his influence his denomination adopted a more Presbyterian form of church government on the Scottish model. He was twice moderator of the general assembly which united the Calvinistic Methodist associations of north and south Wales and owed its institution to him. He died on July 19, 1887. Edwards initiated excellent periodical literature in Welsh: *Yr Esboniwr* ("The Expositor"), 1844 ff.; *Y Traethodydd* ("The Essayist"), 1845 ff. Through his literary and theological essays he greatly influenced Welsh culture. *See* also EDWARDS, THOMAS CHARLES.

See T. C. Edwards, *Bywyd a Llythyrau y Parch. Lewis Edwards, D.D.* (1887). (B. H.)

EDWARDS, RICHARD (1522?–1566), English playwright and musician, author and composer of well-known madrigals. He was born in Somersetshire about 1522, became a scholar of Corpus Christi college, Oxford, in 1540 and took his M.A. in 1547. He was appointed in 1561 a gentleman of the chapel royal and master of the children. In 1564 he entered Lincoln's Inn where at Christmas he produced his play *Damon and Pithias*. Tragic in subject, but with scenes of vulgar farce, it was written in roughly constructed rhymed lines, varying in length and neglecting the caesura. Edwards' song "Awake ye woeful wights" belongs to this play in which music, influenced by the Italian madrigal, plays a considerable part.

On Sept. 2, 1566, his lost play *Palamon and Arcite* was performed before Queen Elizabeth I in the hall of Christ Church, Oxford. A number of shorter pieces are preserved in *Paradyse of Daynty Devises*, a collection which he edited but which was published (1576) after his death. Shakespeare quotes Edwards' song "Where griping griefs" in *Romeo and Juliet*.

The *Historie of Damocles and Dionise* is assigned to Edwards in the 1578 edition of the *Paradyse*, where the words of his madrigal "In going to my naked bed" are published. The music appears in the Mulliner manuscript (*c.* 1560), in the British museum, London, which has another composition of his, "O the silly man," to words by Francis Kinwelmarsh, as well as the words and music of "Where griping griefs."

Edwards died in London on Oct. 31, 1566.

BIBLIOGRAPHY.—L. Bradner, *Life and Poems of Richard Edwards* (1927); E. H. Fellowes, *English Madrigal Composers* (1948); *The Mulliner Book*, ed. by D. Stevens (1951); D. Stevens, *The Mulliner Book, a Commentary* (1952).

EDWARDS, THOMAS CHARLES (1837–1900), Welsh Nonconformist minister and educationist, was born at Bala, Merioneth, on Sept. 22, 1837, the son of Lewis Edwards (*q.v.*). After graduating in London, he matriculated at St. Alban Hall, Oxford, in 1862, obtained a scholarship at Lincoln college in 1864, and took a first class in the school of Literae Humaniores in 1866. In 1867 he became minister at Windsor Street, Liverpool, but left it to become first principal of the University College of Wales at Aberystwyth in 1872. When the college was destroyed by fire in 1885 he collected £25,000 to rebuild it, the remainder of the necessary £40,000 being given by the government and by the people of Aberystwyth. In 1891 Edwards became principal of the theological college at Bala. He died on March 22, 1900.

His chief works were a *Commentary on I Corinthians* (1885), the *Epistle to the Hebrews* (Expositor's Bible, 1888) and *The God-Man* (Davies lecture, 1895).

EDWIN (d. 632), king of Northumbria from 616 to 632, was

the son of Aelli of Deira. During Aethelfrith's reign (593–616) he had been in exile, first in north Wales and then, after the defeat of the Welsh by Aethelfrith near Chester, with Raedwald, king of East Anglia. His cause was espoused by Raedwald, whose army defeated and killed Aethelfrith near the river Idle in 616. Edwin thereupon succeeded to Northumbria and drove out the sons of Aethelfrith.

He conquered Elmet, a Welsh kingdom broadly corresponding in area with the West Riding of modern Yorkshire, and expelled its king, Certic. He invaded north Wales, took possession of Anglesey and besieged Cadwallon, king of Gwynedd, in Priestholm, or Puffin Island, off the eastern point of Anglesey. Edwin also took possession of the Isle of Man, but excepting Elmet these conquests did not endure. He was recognized as overlord by the other Anglo-Saxon rulers of his day, save Eadbald of Kent to whom he was closely related by marriage. Bede comments on the peaceful state of his kingdom and adds that on his journeys he was preceded by standard-bearers.

Edwin was the first Christian king of Northumbria and in this lies the real importance of his reign. Bede has left a deeply moving account of his conversion. A condition of his marriage to Aethelberg, daughter of Aethelberht, Christian king of Kent, was that she should be allowed to practise her religion in Northumbria and when she went north in 625 she was accompanied by Paulinus, who had been consecrated bishop. Pope Boniface V sought to persuade Edwin to be baptized but he hesitated many months. On the day of the birth of his daughter Eanflaed an emissary from Cwichelm, king of Wessex, sought to assassinate him and he was saved from death only by the sacrifice of his thegn Lilla who threw himself in the way of the knife. He vowed to receive baptism if he returned successfully from an expedition of vengeance against Cwichelm, but on his return he still hesitated until Paulinus reminded him of a vision he had seen while an exile at the court of Raedwald. Finally he submitted the matter to his council who were persuaded to turn to Christianity. Coifi, the heathen priest, rode from the meeting to desecrate the temple which he had previously served at Goodmanham. Edwin was baptized at York on Easter Eve, 627, in a wooden church built for the occasion. Paulinus was established as bishop in York and traveled over Northumbria baptizing many people in the river Glen near Yeavering and the river Swale near Catterick.

In 632 Cadwallon, Christian king of Gwynedd, allied himself with Penda, a heathen nobleman of Mercia, and the two invaded Northumbria. They were engaged by Edwin in Hatfield Chase (Oct. 632). Edwin's army was destroyed and he and his son were killed. The succeeding kings in Bernicia and Deira (*qq.v.*) reverted to heathenism and Paulinus fled to Kent with all his fellow workers save James the Deacon who remained preaching near Catterick.

See Bede, *Historia ecclesiastica,* bk. ii, ch. 5–20 (1896); F. M. Stenton, *Anglo-Saxon England,* 2nd ed. (1947). (P. H. Bn.)

EDWY (Eadwig) (d. 959), king of the English from 955 to 957 and king of Wessex and Kent from 957 to 959, the eldest son of Edmund I and Aelfgifu, succeeded his uncle Edred at the age of (at most) 15. According to an early, but biased, source, he quarreled with Dunstan who made him return to his coronation feast which he had left in order to visit a lady Aethelgifu and her daughter Aelfgifu, whom later he married. An 11th-century addition to the Anglo-Saxon Chronicle says Archbishop Oda separated them because they were too near of kin. Dunstan was forced to leave the country. Monastic writers, therefore, are biased against Edwy, and depict him as an enemy to the monastic revival. Though he appears not to have shared the interest of his uncle and brother in it, many charters show him making grants to monasteries. In 957 the Mercians and Northumbrians, tiring of him and his youthful advisers, chose his brother Edgar as king, Edwy retaining England south of the Thames. He died on Oct. 1, 959. The chronicler Aethelweard says people called him "All-fair" because of his beauty.

Bibliography.—*Anglo-Saxon Chronicle;* W. Stubbs (ed.), *Memorials of St. Dunstan* (1874); F. M. Stenton, *Anglo-Saxon England,* 2nd ed. (1947); *English Historical Documents,* vol. i, ed. by D. Whitelock, pp. 45 f., 205, 512–514, 829 f. (1955). (D. Wk.)

EECKHOUT, GERBRAND VAN DEN (1621–1674), Dutch biblical, genre and portrait painter, one of the most gifted pupils of Rembrandt, was born in Amsterdam on Aug. 19, 1621. His father, the goldsmith Jan Pietersz van den Eeckhout, sent him to study painting under Rembrandt, with whom he stayed from 1635 to 1640. He was a favourite pupil and remained a close friend of Rembrandt. His usual style is based so closely on that of his master that many of his pictures, especially the small-scale biblical subjects of his earlier years, have passed as works of Rembrandt himself. Eeckhout was one of the most successful of this school in adopting the broader and bolder technique of Rembrandt's mature style, though he seldom approaches the master in humanity or depth of feeling. From the middle years of the century, Eeckhout's narrative pictures tend to become larger and more ambitious, often with life-size figures, and more brilliant in colour. Though in general they remain true to the Lastman-Rembrandt tradition of biblical painting, they represent its more theatrical and artificial side; the deep chiaroscuro of his early manner gives way to a harder and cruder use of light and shade.

Eeckhout was a competent and successful painter of portraits, most of which belong to the earlier part of his career. In surprising contrast to his normal Rembrandtesque style are a number of highly finished genre subjects—guardroom scenes, backgammon players and so on—in the manner of G. Terborch. This adaptability to popular taste goes far to prove that Eeckhout's considerable talent was of an imitative rather than a creative nature. An example of the early style, once thought to be by Rembrandt, is the "Raising of Jairus's Daughter," in Berlin. A good example of genre in the Terborch manner is the "Music Lesson" of 1655, in Copenhagen. The National gallery in London has a fine group portrait of the "Officers of the Wine Guild of Amsterdam," dated 1657. Eeckhout's later biblical pictures, including some dated examples which closely follow the early Rembrandtesque style, are well represented in European and American collections.

Eeckhout died at Amsterdam on Sept. 22, 1674. (R. E. W. J.)

EEDEN, FREDERIK WILLEM VAN (1860–1932), Dutch writer and doctor of medicine, was born at Haarlem, April 3, 1860. He studied medicine at Amsterdam and with Willem Kloos and Albert Verwey founded (1885) *De Nieuwe Gids,* a literary periodical devoted to the work of modern authors and the expression of new social ideas. Later he became a doctor at Bussum, near Hilversum, where he started a clinic for physical therapy. In 1898 he founded Walden, an agricultural colony based on the ideas of Thoreau. Van Eeden's was a many-sided personality, essentially ethical in outlook, having affinities with Tolstoy. After much doubt, and disappointment caused by his own tendency to overestimate his gifts and therefore to feel unappreciated, in 1922 he found peace in the Roman Catholic Church. He died at Bussum, June 16, 1932.

Although in his early days van Eeden was chiefly known outside his own country for his idealistic social theories, his fame is based on his literary work. He first attracted attention with *De kleine Johannes* (1885; Eng. trans. 1895), a symbolic fairytale. *Het lied van schijn en wezen,* the first part of which appeared in 1895, is a long philosophical poem. His shorter poems were collected in *Van de passieloze lelie* (1901). His psychiatric experience provided material for a novel about mental illness, *Van de Koele meren des doods* (1900; Eng. trans. *The Deeps of Deliverance,* 1902).

Van Eeden's criticism and social treatises were collected in *Studies,* six volumes (1890–1917). He also wrote many plays and translated Rabindranath Tagore's work into Dutch.

See G. Kalff, *Frederik van Eeden* (1927); A. Verwey, *Frederik van Eeden* (1939). (Gd. W. Hs.)

EEL, the name that is properly applied to any fish of the order Anguilliformes (Apodes). The true eels are all elongate in form and lack all trace of the pelvic fin (hence the name Apodes, meaning without feet, or, in this case, pelvic fins). The smooth, muscular, snakelike body, usually scaleless (but with embedded scales in a few species) and covered with a copious supply of mucus, renders eels very difficult to grasp. In most species the very

long dorsal fin and the anal fin are joined around the end of the tail to form, with the greatly reduced caudal fin, a single soft-rayed vertical fin. The gill openings are small; the gas bladder has a duct to the esophagus. Compared with other fishes, the eel's vertebrae are greatly increased in number; the face bones, including the jawbones, are reduced in number. All eels pass through a protracted larval stage, known as the leptocephalus, in which the much compressed and generally elongate body is glassy-clear; these larvae commonly reach a length of more than six inches; one five-foot-long larva has been thought to represent the young of a gigantic eel that may have been the basis for reports of sea serpents.

Fresh-Water Eels.—Best known and of greatest commercial importance are the fresh-water eels, family Anguillidae, of Europe (*Anguilla anguilla*), eastern North America (*Anguilla rostrata*) (fig. 1), Japan (*Anguilla japonica*) and various countries bordering the western Pacific, New Zealand and the Indian ocean (other species of *Anguilla*).

The fresh-water eels utilize streams, ponds and estuaries for their growth and development, but like all other eels, they breed and pass through their larval life in the sea. Some if not all migrate long distances from the mouth of streams to the breeding waters. Thus the European eel, after forsaking the stream in which it developed, migrates, using as yet unknown clues, to an oceanic area southeast of Bermuda, where it is assumed that it spawns and dies. The spawning area, long a mystery, was determined by a great Danish scientist, Johannes Schmidt, through a prolonged and famous study of the distribution of the larvae of different size-classes. He further discovered that the American eel breeds in an overlapping area, and that the leptocephalus larva of the European eel, which drifts slowly to Europe in the Gulf Stream, requires three years to reach the stage of transformation known as the elver, at which stage the young eel runs into streams to start its period of growth and development in fresh water. (In the middle ages, long before the remarkable life history of the European eel had been deciphered, it was thought that this fish was spontaneously generated from mud.) The American eel, which has a shorter journey in the open sea, passes through its larval life in a single year.

Fresh-water eels are highly esteemed for food, especially in Europe and Japan. Large numbers are caught, and are consumed either fresh or smoked. The main fishing in Europe is for "silver eels," the stage in which the eels, rich in fats, are starting their journey downstream toward the sea. In Japan, and to a lesser extent in Europe, eels are reared for the market in ponds.

Marine Eels.—In the oceans, particularly in warm seas, several hundred known marine eels represent many types and live in diverse habitats. The most conspicuous eels of the coral reefs are the morays (family Muraenidae), which are sharp-toothed, large-mouthed and often vicious fishes, much feared by skin divers. Morays lack the pectoral as well as the pelvic fins and have the gill openings reduced to pores. The snake eels (Ophichthidae), in which the finless tail tip projects beyond the ends of the dorsal and anal fins, are also widely distributed in warm coastal waters, along with the little worm eels (generally less than a foot long) and members of other families. Various eels occur on the bottom in deep water—among them *Synaphobranchus*, with the gill slits united below; and *Simenchelys*, a blunt-headed eel, which, like the hagfishes, burrows into other fishes to consume the muscle. Still other and even more bizarre eels inhabit the cold black mid-depths of the ocean. Prominent among these are the snipe eels (Nemichthyidae), in which the threadlike jaws are rough with tiny teeth that entangle the prey and the body cavity is extremely short. The excessively elongate tail region is extended into a long filament, except in the bobtail snipe eel, *Cyema atrum* (fig. 2).

TRENT DAVIDSON

FIG. 1.—AMERICAN EEL (ANGUILLA ROSTRATA), EASTERN NORTH AMERICA

BY COURTESY OF CARL L. HUBBS

FIG. 2.—BOBTAIL SNIPE EEL (CYEMA ATRUM), OF MARINE MID-DEPTHS

The name eel is often applied to many other eel-shaped (anguilliform) fishes. For example, the lamprey (*q.v.*) is often called lamprey eel, and the hagfish (*q.v.*) is also called slime-eel. The so-called spiny eels and their relatives are deep-sea fishes belonging to the order Notacanthiformes (Heteromi). The electric-eel (*q.v.*) is not an eel at all, but is more closely allied with catfishes, minnows, etc., in an entirely different order.

See also FISH; and references under "Eel" in the Index volume.

(C. L. Hs.)

EELGRASS (GRASS WRACK), the name applied to certain aquatic plants (*Zostera* species) growing in shallow coastal salt water in cool temperate regions. The lower part of the stem is creeping, the branches growing upward and dividing. The leaves are long and narrow.

Eelgrass, along with other plants of the pondweed family (Potamogetonaceae), fills an important ecological niche; its seeds and vegetative parts provide food for waterfowl.

EFFIGY: *see* SCULPTURE, SEPULCHRAL; WAX FIGURES.

EFFLORESCENCE, the spontaneous loss of water by a hydrated salt, occurs when the aqueous tension of the hydrate (*q.v.*) is greater than the partial pressure of the water vapour in the air. As the vapour pressures of washing soda ($Na_2CO_3.10H_2O$) and Glauber's salt ($Na_2SO_4.10H_2O$) normally exceed that of the water vapour in the atmosphere, those salts effloresce, *i.e.*, lose all or part of their water of hydration, and their surfaces assume a powdery appearance. Hydrated cupric sulfate or blue vitriol ($CuSO_4.5H_2O$), whose aqueous tension is lower, undergoes efflorescence only if the air in contact with it is relatively dry. *See* DELIQUESCENCE.

(J. B. Ps.)

EFIK (the OLD CALABAR), a Nigerian tribe, offshoot of the Ibibio, with a population of approximately 72,000 in 1961. Originally the Efik were a community of fishermen who moved down the Cross river to exploit the prawn fisheries on the estuary. Most of them settled at Ikot Itunko (Creek Town) and became known to Europeans in the 17th century as Old Calabar, being confused with the Kalabari (New Calabar). At that time the Kalabari were the principal Ijaw (Ijo) trading and fishing community on the Rio Real, later known as the Bonny river (*see* also IJAW). By 1880, the Old Calabar and Bonny rivers had become the chief trading ports of the Bight of Biafra, exporting slaves and later palm oil. Old Calabar expanded into several communities, the most important being Duke Town and Creek Town whose kings received the dues (*comey*) paid by the European shipping and were also the head and deputy head of a single secret society, Ekpe (otherwise known as Egbo). During the 20th century, a large part of the Efik population moved out from these towns and settled in farming villages in the forest hinterland, while Duke Town developed into the cosmopolitan township of Calabar and the administrative and educational centre of the Oil Rivers Protectorate. *See* also AFRICA: *Ethnography (Anthropology)*: *West Africa*.

BIBLIOGRAPHY.—D. Forde and G. I. Jones, "The Ibo and Ibibio Speaking People of South-Eastern Nigeria," *Ethnographic Survey of Africa: Western Africa*, pt. iii (1950); D. Forde (ed.), *Efik Traders of Old Calabar* (1956).

(G. I. J.)

EGAN, PIERCE (1772–1849), English sporting writer, whose works were regarded as indispensable reading for fashion-

able men about town in the early 19th century, was born in London. He made his reputation as a boxing reporter, but became an authority on all sports and developed a flare for sensational literature. *The Mistress of Royalty* (1814), describing the prince regent's intrigue with Miss Robinson, set and printed by Egan himself, was his first book; his best-known was *Life in London: or the Day and Night Scenes of Jerry Hawthorn, Esq., and his elegant friend Corinthian Tom* (1821), which was illustrated by George and Robert Cruikshank, and gave a lively account of contemporary "fast" life. Among his other works are the serial *Boxiana* (1818–24), racy, yet accurate, records of the careers of famous pugilists; *Life of an Actor* (1825); *Book of Sports* (1832); and *Pilgrims of the Thames* (1838). He founded (1824) the popular sporting weekly which in 1827 became *Bell's Life in London*. He died in London, Aug. 3, 1849.

EGAS MONIZ, ANTONIO CAETANO DE ABREU FREIRE (1874–1955), Portuguese neurologist, and, with W. R. Hess (*q.v.*), 1949 Nobel laureate in medicine, was born at Avanca on Nov. 29, 1874. After taking his medical degree at the University of Coimbra in 1899 he taught there for several years and then, having decided to specialize in neurology, undertook further courses of study at Bordeaux and at Paris. In 1911 he was appointed to the newly created chair of neurology at the University of Lisbon, and this office he held until his retirement in 1945. Egas Moniz took an active part in the political life of Portugal. He served several times in the Portuguese chamber of deputies between 1903 and 1917, was Portuguese minister at Madrid (1917–18) and foreign minister and leader of the Portuguese delegation to the Paris peace conference (1918–19). Between 1927 and 1937 he introduced and developed cerebral angiography, a method of visualizing the blood vessels of the brain by the injection of substances opaque to X-rays. This method proved to be of considerable value in the diagnosis of intracranial disease. In 1936 he published the results of his pioneer work on the treatment of certain mental diseases by the operation of prefrontal leucotomy, in which part of the white matter of the prefrontal lobes is destroyed. It met with a considerable degree of success in selected cases and opened up new lines of attack on severe and intractable forms of mental illness. It was for this work that he was awarded the Nobel prize. Egas Moniz died in Lisbon on Dec. 13, 1955. (W. J. Bp.)

EGBERT (Ecgberht) (d. 839), king of the West Saxons from 802 to 839, son of Ealhmund, king in Kent in 784 and 786, was driven into exile by Offa of Mercia and Beorhtric of Wessex and spent three years (*c.* 789–*c.* 792) in the Frankish kingdom. From *c.* 792 until he succeeded Beorhtric in 802 he was probably in some Celtic country. He harried Cornwall in 815 and fought against the Britons of this area in 825, and again in 838 when they joined with the Danes. He gave a 10th of his conquests there to the church. In 825 he defeated the Mercians at *Ellandun* (Wroughton), and an army including his son Aethelwulf drove King Baldred from Kent. The people of Kent, Surrey, Sussex and Essex accepted his rule, with Aethelwulf as sub-king. In 829 he conquered the Mercians, and the Northumbrians submitted without fighting. The chronicler therefore calls him *bretwalda*, but his supremacy was short-lived, since Wiglaf of Mercia recovered his throne in 830. Meanwhile Egbert had made a successful expedition into Wales. He was defeated by the Danes in 836, but won a great victory at Hingston Down (Cornwall) in 838. He died in 839.

Bibliography.—*Anglo-Saxon Chronicle*, under the years 784 (Manuscript "F"), 800–836; F. M. Stenton, *Anglo-Saxon England*, 2nd ed. (1947); *English Historical Documents*, vol. i, ed. by D. Whitelock, pp. 25 f., 165, 169–172, 822 f. (1955). (D. Wk.)

EGBERT (Ecgberht) (d. 766), Anglo-Saxon archbishop of York and founder of its cathedral school, was a pupil of Bede. He was appointed bishop of York (732) by his cousin Ceolwulf, king of Northumbria, and received the archbishop's pallium (735) from Pope Gregory III, thus becoming the first northern archbishop since Paulinus (d. 644). Bede wrote to warn him of the evils caused by excessive endowment of monasteries, and Egbert and his brother Eadberht, king of Northumbria (737–758), seem to have acted on this advice. Egbert gave careful instruction to the clerks of his church and wrote an epitome of ecclesiastical law, the *Dialogus ecclesiasticae institutionis*. He also compiled a penitential manual and a pontifical (*see* Coronation). He was a correspondent of St. Boniface, and Alcuin studied under him at York. He died on Nov. 19, 766.

Bibliography.—*Anglo-Saxon Chronicle; Baedae continuatio*, ed. by C. Plummer in *Venerabilis Baedae opera historica*, vol. i. (1896); *The Pontifical of Egbert, Archbishop of York* (1853); A. W. Haddan and W. Stubbs, *Councils and Ecclesiastical Documents*, vol. iii (1871); *English Historical Documents*, vol. i, ed. by D. Whitelock, pp. 16, 160 f., 166, 735, 757, 764 (1955).

EGEDE, HANS (1686–1758), Norwegian missionary, the first man to preach the Gospel to the Eskimos in Greenland, was born on the island of Hinnerøen, Jan. 31, 1686. He studied at Copenhagen university and while there became interested in the project of a mission to Greenland. He was appointed (1706) curate at Vaagen in the Lofoten Islands and spent 13 years winning support and preparing for his expedition. In July 1721 with 43 others Egede reached the Godthaab fiord on the west coast of Greenland and had succeeded in establishing a colony there by 1728. He converted and baptized many Greenlanders, began to teach natives who would be able to carry on his work, and through the Bergen Greenland company initiated trade with Denmark. He returned to Copenhagen in 1736 and became principal of the seminary founded (1737) to prepare missionaries for Greenland. From 1740 to 1747 he was superintendent of the mission in Copenhagen. He died in Stubbekøbing, Den., Nov. 5, 1758. His book, *A Description of Greenland*, was translated into English in 1745.

His son, Paul Egede (1708–1789), completed (1766) the translation of the New Testament into Greenlandic begun by his father, whom he succeeded as superintendent of the mission.

See L. Bobé, *Hans Egede* (1945; Eng. trans. 1952).

EGER: *see* Cheb.

EGER (Ger. Erlau), a small town in the hill country of northern Hungary, Heves county, 39 km. (24 mi.) S.W. of Miskolc, in the valley of the Eger, a tributary of the Tisa. Pop. (1960) 35,375 (mun.). To the north is the limestone plateau of Bükk, the farthest projecting outcrop of Carpathian hill country, although the hills close to the town are volcanic in origin. Eger is an old city (the bishopric dating from the early 10th century) and was one of the richest in medieval Hungary. The remains of the fortress, from which Stephen Dobo defended Eger from the Turks in 1552, are on the hill to the northeast. The minaret (115 ft.) opposite the church of the Brothers of Mercy is the surviving feature of the mosque built during the Turkish occupation. Eger became an archbishopric in 1814, and the number of ecclesiastical buildings which thereafter arose in the city caused it to be known for a time as the "Hungarian Rome." The rich soils of this hill country have been used for centuries for vineyard farming. The well-known Erlauer wine is produced and marketed at Eger and also in the neighbouring towns of Gyöngyös and Hatvan. Eger was scheduled by the Hungarian government as suitable for further manufacturing activity. (H. G. S.)

EGERIA, an Italian goddess, a water spirit or Lympha (*see* Nymphs). She was worshiped in connection with Diana (*q.v.*) at Aricia (see *Aeneid* vii, 762 ff.) and also among the Camenae (*see* Muses) in their grove outside the Porta Capena at Rome. Like Diana she was a protectress of pregnant women and like the Camenae was considered prophetic. She possibly had a male counterpart, Egerius, at Aricia. By tradition she was wife and adviser of King Numa Pompilius, who established the grove at Rome and there consorted with her (Livy I, xxi, 3). Livy, however, rationalizes that Numa gave out this story only to gain approval of his religious reforms, which he reported that she inspired.

Bibliography.—G. Wissowa, *Religion und Kultus der Römer* (1912); W. Roscher, *Ausführliches Lexikon der griechischen und römischen Mythologie*; A. E. Gordon, *Cults of Aricia* (1934). (R. B. Ld.)

EGFRITH (Ecgfrith) (d. 685), king of the Northumbrians from 670 to 685, succeeded his father Oswiu. Early in his reign he suppressed a rising of the Picts, and between 670 and 674 he

defeated an invasion by King Wulfhere of Mercia and annexed Lindsey. He lost this province to King Aethelred of Mercia after the battle of the Trent in 678. In 684 he sent an army which ravaged in Ireland, and in 685 he led his forces against the Picts, but was enticed into mountain country and slain at Nechtanesmere, near Forfar, on May 20, 685. His first wife, Aethelthryth, daughter of King Anna of East Anglia, was encouraged to become a nun by Bishop Wilfrid, and Egfrith's anger at this, fomented by his second wife, Eormenburh, led to Wilfrid's expulsion. Though Wilfrid returned in 680 with favourable letters from the papal see, he was imprisoned for nine months. After his release he left Northumbria, which was then divided into the dioceses of York, Lindisfarne and Hexham, with a see fixed at Abercorn for the Picts. In 684 Egfrith persuaded Cuthbert to become bishop of Hexham, a see which he later exchanged with Eata for Lindisfarne. Egfrith gave estates to Benedict Biscop for the foundation of the twin monasteries at Wearmouth and Jarrow.

BIBLIOGRAPHY.—B. Colgrave (ed.), *The Life of Bishop Wilfrid by Eddius Stephanus* (1927); Bede, *Historia ecclesiastica,* book iv, ch. 5, 12, 17, 19, 21, 26, *Historia abbatum,* sec. 4, 7, 9; F. M. Stenton, *Anglo-Saxon England,* 2nd ed. (1947). (D. Wk.)

EGG, AUGUSTUS LEOPOLD (1816–1863), English genre painter and actor, was born on May 2, 1816, in London. He was a pupil of Henry Sass, and then studied at the Royal Academy, of which he became a member in 1860. He traveled in Italy with Dickens and Wilkie Collins in 1853. Egg was an excellent actor and played in Dickens' company of amateurs; one of his best parts was as John Want in Collins' *Frozen Deep.* He was famous in his day for small anecdotal pictures, of which the London public galleries possess several examples. He died at Algiers on March 26, 1863.

Among his principal pictures are the "Introduction of Sir Piercie Shafton and Halbert Glendinning" (from Scott's *Monastery*); "Buckingham Rebuffed"; "Queen Elizabeth Discovers She Is No Longer Young"; "Peter the Great Sees Catharine for the First Time"; "Charles I Raising the Standard at Nottingham"; the "Life and Death of Buckingham"; two subjects from Thackeray's *Esmond;* "Past and Present, a Triple Picture of a Faithless Wife"; the "Night Before Naseby"; his last exhibited work, the Dinner Scene from *The Taming of the Shrew.*

EGG, the female germ cell or ovum. In popular usage the term ordinarily refers to animal eggs developing outside the body, including their protective shells or capsules. The word is even used synonymously with shell, as "the embryo developing within the egg." To the biologist, the egg is the living germ cell that becomes the embryo through a series of progressive changes.

The following article is divided into two parts: *The Egg in Biology* and *Edible Eggs.*

THE EGG IN BIOLOGY

The Egg as a Germ Cell.—The egg is a large cell, of round or oval form. The diameter of eggs in different species ranges from a few thousandths of an inch to several inches. In multicellular animals (exclusive of sponges, where amoeboid interstitial cells transform into the germ cells), the egg is formed within the ovary, where it is furnished with growth materials by surrounding tissues. In many species special nurse cells contribute material directly to the egg. With few exceptions in animals above the coelenterate level (above the level of jellyfishes and allies), there are accessory structures such as oviducts, associated glands, storage chambers, etc., that convey the egg to the exterior when it is liberated from the ovary.

In lower organisms, as some protozoans, certain algae, etc., the macrogamete (usually nonmotile and the larger of the two germ cells) is considered to be the equivalent of the egg. In mosses and ferns, and their allies, the egg is the central cell of a protective structure, the archegonium, borne on the female gametophyte. In seed plants both archegonium and gametophyte are greatly reduced, and in flowering plants the equivalent of the egg is a single nucleus in the embryo sac.

As in all cells, the primary features of the egg are a nucleus and surrounding cytoplasm, limited externally by a cell membrane. Animal eggs differ from those of plants in two major features: (1) In the course of maturing, the nucleus of the animal egg undergoes the process of meiosis or halving of the array of chromosomes characteristic of all other cells of the species. This meiosis is completed by two successive cell divisions, each of which separates a large daughter cell, the egg, from an abortive sister cell, the polar body, which will subsequently degenerate. In plants, meiosis occurs in the spore ancestral to all gametophyte cells, hence the egg already contains the halved chromosome set. (2) Before it leaves the ovary, the cytoplasm of the animal egg acquires a highly distinctive structure, varying from one species to the next. The degree to which such cytoplasmic specialization occurs in plant eggs is not known; they are much less variable in visible structure than are animal eggs.

A mature egg ready to be shed from the ovary is possessed of a well-defined polarity, the most obvious sign of which is the displacement of the nucleus toward one side of the cell, indicating the point where the polar bodies will separate; this point is known as the animal pole. Some eggs, *e.g.,* those of insects and cephalopod mollusks (octopuses, squids, etc.), show also a marked bilateral symmetry. Various cytoplasmic bodies occupy positions related to the polar axis, to the bilateral axis if present, or to the surface-interior pattern of the egg cell. The peripheral region of the cytoplasm beneath the cell membrane, the cortex of the egg, has a highly specific structure and plays a special role in fertilization and development. Nutrients enter the egg cell across the membrane during the growth period, and the membrane shows marked structural relations to surrounding cells or fluids in the ovary.

The most variable cytoplasmic component of eggs of different animal species is the yolk; *i.e.,* granules of various shapes and sizes composed of food stores, largely protein and fatty in nature, elaborated in the cytoplasm according to a highly specific pattern. It is the accumulation of yolk in eggs of certain animal groups, *e.g.,* fishes, reptiles and birds, that is largely responsible for the gigantic size reached by these cells. Animal eggs can be classified as: (1) lacking yolk, *e.g.,* the eggs of many flatworms, where the embryo derives nourishment from separate yolk cells; (2) having small yolk granules evenly distributed through the cytoplasm, *e.g.,* eggs of echinoderms (starfishes, etc.) and mammals; or (3) as having yolk accumulated preferentially in certain regions of the cytoplasm. In the last two mentioned cases the yolk may be either central in position (arthropods) or aggregated in the vegetative hemisphere, antipodal to the animal pole. In extreme cases of the latter type, as in the eggs of birds, the cell is essentially a vast mass of yolk surmounted by a tiny lens-shaped protoplasmic disc at the animal pole. Nevertheless, even in such extreme cases the yolk itself has a polar and concentric pattern. Plant eggs lack a substance comparable to yolk, although they may contain stored food in organelles associated with plastids.

Membranes that surround the animal egg when the latter is released from the ovary fall into three classes according to their origin. All eggs possess a primary (vitelline) membrane, which is the cell membrane of the egg itself. Some eggs, *e.g.,* those of insects, are also covered by secondary membranes produced by neighbouring cells in the ovary. Tertiary membranes may be formed around the egg by structures outside the ovary, *e.g.,* in the oviduct, or by associated glands either inside or on the exterior of the parent animal. In the familiar egg of the hen, for example, the tough covering of the ovum or "yolk" is the primary or vitelline membrane. The albumen layers, shell membranes and the shell itself are all tertiary coats laid down successively in the oviduct.

Variability in Eggs.—Among animals there are numerous cases where two sorts of germ cell are produced, either successively or simultaneously, in a single organ. The extreme case is the hermaphrodite gland of certain mollusks, where eggs and sperm may occasionally develop side by side. In the microscopic aquatic rotifers it is usual for two types of egg to be produced (dimorphism of eggs) depending on the season: "summer eggs" with thin capsules, which develop without fertilization into females; and smaller eggs, which, if not fertilized, will develop into males, but which if fertilized will enlarge into "winter eggs." These winter

eggs withstand cold or drought and subsequently develop into females that again produce summer eggs. Special chromosomal mechanisms ensure that the normal chromosome quota is maintained through the generations.

In some animal classes a more subtle dimorphism of eggs occurs, in relation to sex determination. The genetic mechanism of sex determination consists of a chromosomal balance such that one sex or the other produces germ cells of two different types: some (usually about 50%) male determining and the rest female determining (*see* HEREDITY: *Mendelian Inheritance: Chromosome Theory of Heredity*). The opposite sex produces germ cells all alike in respect to sex genes. Most commonly the male is the sex producing two types of germ cell: this is the case in humans. However, among butterflies and moths, in birds and probably in some amphibians there are demonstrated cases where the female produces two types of egg, male and female determining respectively, whereas the sperm cells are all uniform.

Ovulation and Fertilization.—The specific physiology of the egg involves successively three unique processes: the growth and organization mechanisms by which the cell differentiates; ovulation, or liberation of the egg from the ovary; and fertilization whereby the egg is united with a sperm to form a zygote or fertilized egg which then develops into a new individual.

In the ripe ovary in some of the marine invertebrates the mature eggs are found floating free; ovulation consists merely in their expulsion. In some cases release of eggs has been shown to be triggered by a chemical stimulus derived from mature males of the species in the vicinity; on the other hand, males have been shown to shed sperm under a similar chemical influence from mature females. In animals with internal fertilization, physiology of internal secretion is probably always involved in the release of eggs at a moment favourable for fertilization. In ovaries where the egg develops in a follicle or sac, the release is more than a mere budding off of a cell or cells; here the maturing of the egg is accompanied by accumulation of fluid within the follicle, resulting in the distention and ultimately the rupture of the follicle wall, at which time the ripe egg is released.

Fertilization is a highly specific reaction involving accumulation and clumping of sperm around the egg; contact between one sperm and the egg surface—in some cases only a restricted area of the egg surface is accessible—and engulfment of the sperm by the egg cytoplasm. The egg then becomes impenetrable to other sperm; in some cases supernumerary sperm are admitted, but they appear to take no part in development. The visible sign of successful contact between an egg and a sperm is the raising of the egg membrane above the surface of the cell (the fertilization membrane). The sperm nucleus is then transported through the egg cytoplasm to unite with the egg nucleus, and the first cell division ensues. Various substances, produced by the egg cortex or surrounding membranes, or by the sperm, have been identified with different phases of the fertilization process. Most of these substances are of protein nature, highly specific as to action, and restricted to their own species; however, in some cases, in lower organisms and plants, less specific substances appear to play a role, at least in the attraction of the cells.

The fact that in numerous species parthenogenesis (*q.v.*; development of the egg without fertilization) occurs normally, and that experimental parthenogenesis is possible in many other cases, emphasizes that the egg contains all the essentials for development, and that the sperm, aside from its genetic contribution, has the role of activating an already prepared system.

The Egg and the Life History.—A close if not always obvious relation exists between the cellular features of the egg and the life history of the species to which it belongs. The protoplasmic framework of the egg sets the structural framework of the new individual; the quantity and arrangement of the yolk decides when and to what extent the new individual must be able to use outside food; and the egg coverings determine what materials can enter and leave the enclosed embryo.

Development from the egg to the adult in animals may be a direct progression, or it may involve a succession of distinct transformations, usually with radical changes of habit. The eggs of many bottom-dwelling marine animals, for example, are released into the sea and rapidly hatch into small surface-swimming larvae very different from the parent. The larvae, after a period of feeding, sink to the bottom and metamorphose into the adult form. The higher insects and most amphibians pass through similar radical changes. Many species have more than one larval or juvenile stage in sequence. Parasitic species in particular may have long and complicated life histories involving very precise adjustment to a specific host at each larval stage.

The number of eggs produced by an adult animal during its life is in direct proportion to the probable mortality during the life history. For every species there is a narrow or wide range of environment in which an egg can develop successfully. The biological arrangements by which species assure their own continuity by providing that a sufficient number of eggs actually do remain in a suitable environment for a sufficient period are numerous and varied.

Plants above the algae and fungi level have solved the problem of milieu for the new generation by maintaining the developing egg within the parent gametophyte tissues, where no new environmental adjustment is required.

In animals there is a great variety of adaptations ensuring a suitable environment to the developing egg. Animals are sometimes classified as oviparous (where eggs are shed to the exterior) and viviparous (eggs retained within the maternal body and dependent thereon for a considerable period). A somewhat intermediate condition, ovoviviparity, is recognized, where encapsulated eggs, capable, at least by inference, of developing outside the maternal body, are nevertheless retained internally until they hatch. These distinctions are difficult to maintain. It is more useful to take the view that there exists a whole spectrum of conditions on the basis of the degree to which the parent furnishes an intimate and specific environment for the egg. The extreme on the one hand would be the marine species in which incompletely mature eggs are shed into the sea, there to complete maturation, be fertilized, evolve their covering membranes and continue development in complete independence of the parent. The extreme on the other hand would perhaps be found in placental mammals, where the egg is fertilized internally and very early in development becomes as intimately associated with the tissues of the wall of the maternal duct as the plant ovum is with the female gametophyte.

Eggs Developing Within the Parent.—The wide variety of structural relations between parent and egg imply an equally wide variety of physiological relations of which little is known. In coelenterates, where the gonads are simple swellings on the external or the gastric surface of the animal, the egg in some species is fertilized *in situ* and the developing individual remains for a time as an appendage of the parent, later dropping off. In some higher invertebrate species, also, development takes place in the ovary or in the body cavity. Among vertebrates, one may cite the familiar guppy, inhabitant of countless home aquariums, where the egg is fertilized and develops inside the sac that encloses the ovary. In some species of parasitic flatworms, individuals may reproduce while still inside the parent body, so that two or even more generations may be found enclosed within one another like Chinese boxes —a situation comparable to the seeds of higher plants. Retention of the encapsulated egg within the maternal ducts may be combined with devices for internal maintenance of the young after hatching. In the toad *Nectophrynoides* the tadpoles hatch within the oviduct, and for a time their tails, which are provided with a rich blood supply, serve as intermediary organs of respiration against the wall of the maternal duct. In viviparous sharks, an egg case encloses a number of the gigantic yolky eggs as they pass down the oviduct. In the lower part of the duct, the capsule is dissolved away and each developing embryo acquires a surface connection to the maternal wall across which materials may pass between the young and the blood of the mother.

The Human Ovum.—The human ovum furnishes a typical example of the mammalian egg cell, adapted for a very specialized viviparous development. Eggs mature in the follicles of the ovary according to a pattern geared to the female endocrine cycle in which the secretions of the pituitary gland, of the follicle itself

and of the ovarian tissue between the follicles are particularly active. In humans, usually only one egg matures at a time. Ovulation is in general geared to the endocrine cycle also. The egg as released from the ovary has a diameter of about 0.006 in.; its cytoplasm is transparent and the yolk granules small. Surrounding the liberated egg are several covering layers produced by the egg membrane and the follicle cells. The egg is conveyed to the oviduct; if sperm are present, fertilization ensues, and meiosis, which had been arrested midway, is completed. Development is initiated as the egg passes down the oviduct to the uterus where, now a hollow ball composed of about 50 cells, it will become embedded in the lining tissue. Around the egg will eventually develop, partly from its outermost layers and partly from the surrounding maternal tissue, the placenta.

Eggs Developing Externally.—In species (all aquatic) where fertilization is external, the egg coverings must be of a nature to permit sperm entrance and yet to protect the egg afterward. The eggs of bony fish, covered by a tough capsule perforated at the animal pole, furnish one solution. Another solution is given by the egg masses of frogs and toads, covered, when laid, by a thin jelly that does not bar the sperm, but that subsequently swells to substantial proportions by gradually absorbing water.

Many aquatic species and all land-living ones except those that do not reproduce bisexually have arrangements for internal fertilization. In many such species elaborate protective capsules are constructed around the fertilized eggs as they leave the body; these may be thickened primary, secondary or tertiary membranes. Frequently a cluster of encapsulated eggs are cemented together in jelly masses, strings or cocoons as a further protection against mechanical shock and drying. Familiar examples are the rectangular egg cases of rays and skates, with their long tendrils for attachment to solid objects; the long strings of egg cases of some snails; the finger-shaped jelly masses of squids; the irregular jelly masses of some tailed amphibians; the cocoons of earthworms; the varied egg cases and pods of insects.

Brooding Devices.—In animals in which little or no development of the embryo occurs before the egg leaves the maternal reproductive tract, the parent may still continue control of the environment. Many animals have devices for holding the eggs to the parental body surface. The fresh-water crayfish carries a mass of eggs on the abdominal appendages and aerates them frequently. The male midwife toad similarly carries an egg mass on his legs. A large number of animal species have brood pouches formed from the external body wall, or in body cavities. Among hydrozoan coelenterates sessile gonophores function as brood chambers; there is one hydromedusan, *Eleutheria*, that has a brood chamber in the apex of the bell. Brooding devices are rare among the worms, but typical of bryozoans and of many crustaceans. In many mollusks, particularly the bivalves, the mantle chamber serves as a brood pouch; the atrial chamber in many ascidians is similarly used. Among the vertebrates a few cases can be cited: the male of the pipefish and sea horse have brood pouches; male or female in some species of fish (especially cichlids) have the habit of using the mouth cavity as a brood chamber. Among amphibians the male *Rhinoderma* carries the eggs in pouches adapted from the vocal sacs. In other species, notably *Pipa*, the female has a brood chamber on her back, made of skin folds. The emperor penguin has a skin pouch for incubating its eggs. Among mammals the pouch of the echidna should be mentioned, as well as the marsupium of the kangaroo, opossum, etc., although the latter is occupied by the young only after they have passed the embryonic stage.

Nests.—A widespread means of regulating the environment of eggs consists not in structural modification of the parent, but in a series of social or behaviour patterns which furnish an equally effective means of control. Among arthropods, insects in particular display elaborate sequences of unlearned behaviour in ensuring a favourable situation for their eggs and adequate food for the immature young upon hatching. These patterns may involve digging or building or covering nest sites, or relations to other animals and plants; stalking and storing prey as do the solitary wasps; or an elaborate social life centring around the care of eggs, as in hive bees, ants or termites. It is among the vertebrates, however, that the most elaborate and flexible breeding patterns have evolved. Bony fish, reptiles and birds in particular furnish continuing objects of interest for naturalists, physiologists and psychologists in regard to mating patterns, nest building and protection or incubation of eggs. In nature these activities are highly seasonal, triggered by changes in the environment. In nonequatorial climates seasonal differences in light, mediated in the individual through the endocrine system, are probably the most important of these changes. (*See* also NEST.)

Fish, Amphibia, Reptiles.—In bony fish, eggs developing outside the body are either of the demersal type (large, opaque, denser than water) or pelagic (relatively small, transparent, lighter than water, and often equipped with buoyant oil globules). The habit of hollowing or scooping out a crude nest in weeds, mud or sand for depositing eggs of the demersal type, and of guarding the nest, is found in some of the most primitive fish species such as the bowfin, as well as in more advanced form in the stickleback, goby, blenny, etc. The male usually guards the nest, but both parents may participate. Some species show remarkable adaptation to some other organism; *e.g.*, the bitterling: the female deposits eggs in the mantle cavity of a clam, which then serves as a brood chamber. Since the larvae of the clam live for a period parasitically on the gills of fish, the clam may at this time place its young out to board with the bitterling. Some fresh-water species, notably the Siamese fighting fish *Betta*, make bubble nests. The eggs are demersal, but as they are released by the female and fertilized, the male gathers them in his mouth before they sink and places them in a mass of foam that he has previously blown on the surface of the water. He guards this bubble nest meticulously and pugnaciously until the eggs hatch.

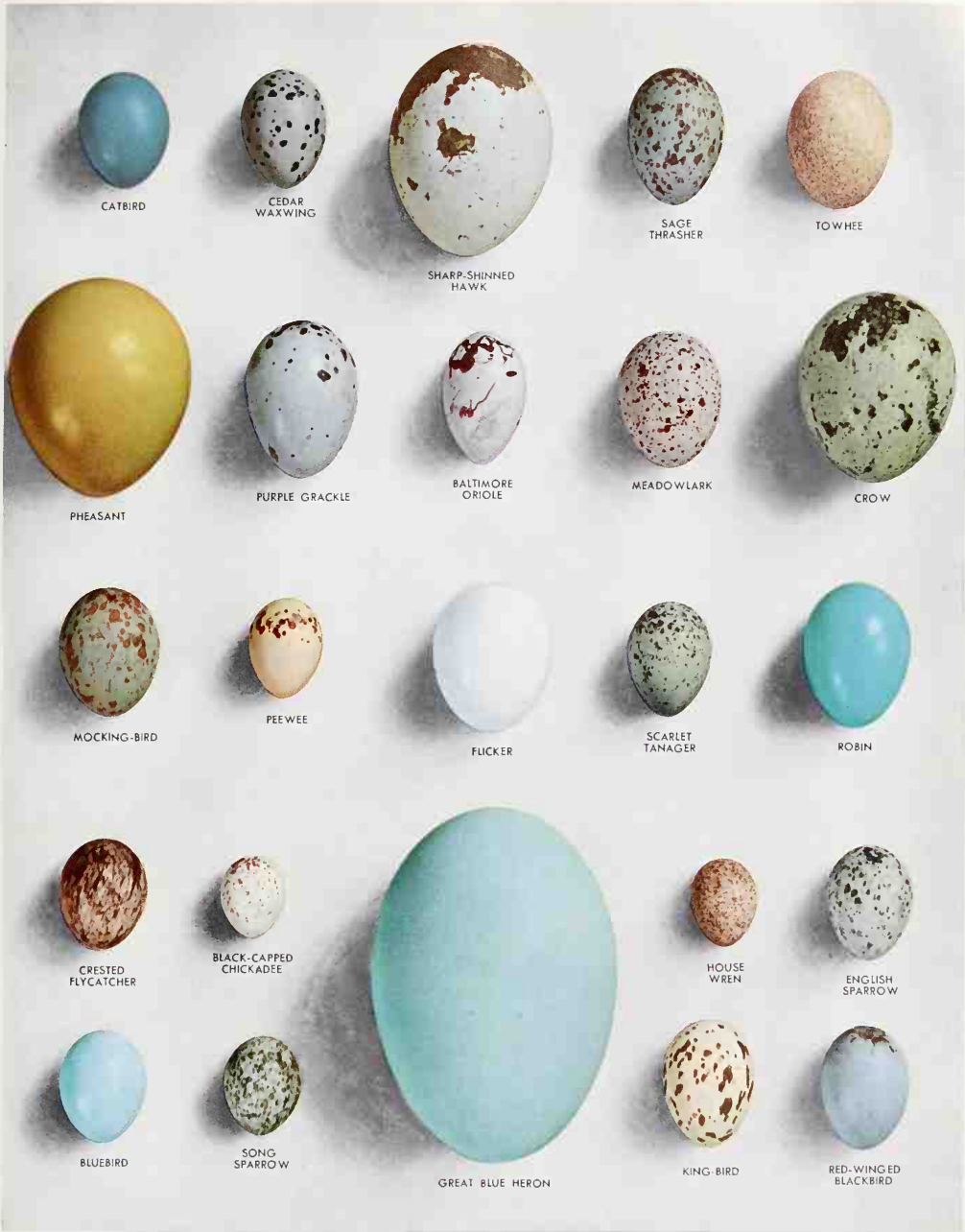
Although some amphibians (Apoda) have the habit of depositing eggs in nests and coiling around them on guard, most of them do not display much parental concern. Reptiles, on the other hand, all protect their eggs by digging or building some variety of nest. Reptilian eggs have a tough or hard shell and are particularly adapted to terrestrial development; marine turtles come on shore to deposit eggs. No reptiles display guarding behaviour.

Birds.—The highly organized breeding behaviour of birds includes a wide variety of adaptations in nesting and egg production. Some species make no nest at all, depositing eggs in irregularities on the ground. The other extreme is found in remarkably elaborate constructions: one might cite particularly the Australian megapodes for their imposing ground nests or mounds; the orioles with their suspended nests; and the grebes, which build raft nests on the water. The choice of site is always peculiar to the species.

The shell of birds' eggs not only affords the embryo maximum physical protection with minimum weight, but in many species is protectively coloured. The colours, due to pigments derived from the breakdown of hemoglobin in the parent body, are added to the shell as it forms in the oviduct.

Perhaps the most interesting birds in respect to eggshell colour and pattern are the cuckoos, comprising several socially parasitic species that have lost their own nesting patterns and instead deposit eggs in the nests of other birds, which then seem tamely to incubate and nourish the intruder at the expense of their own nestlings. One species of cuckoo may thus parasitize several normal species with distinctly different eggs. In such cases the cuckoos seem to be subdivided into specialized races genetically. In each case the cuckoo egg mimics the colour and pattern of the egg of the foster parent, and thus is not rejected by the victim.

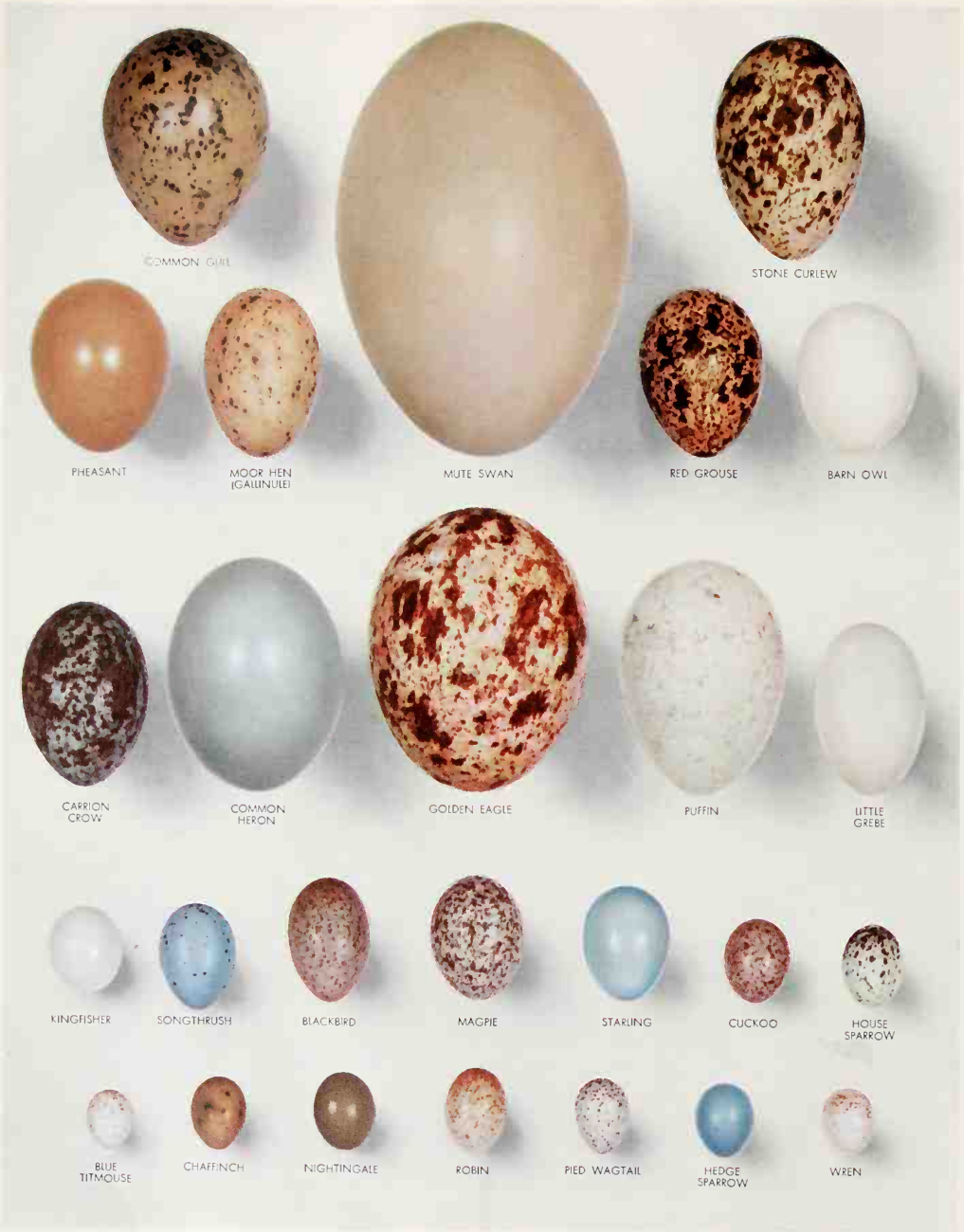
The ovulation pattern in most birds is such that a limited number of eggs are released from the ovary in succession—usually one a day. The number of eggs in such a group or "clutch," within certain limits of variation, is a species character and seems to be nicely adjusted to the survival of the species in spite of enemies without and competition within. Normally, in nature, the laying of a clutch is followed by an incubation period during which the female develops peculiar modifications that promote brooding. Removal of one or more eggs of a clutch will, in some species, stimulate the resumption of laying to compensate the loss. If eggs are successively removed, a dove, for example, will continue laying to the extent of producing over 20 eggs, although the normal clutch



COURTESY CHICAGO NATURAL HISTORY MUSEUM; PHOTOGRAPH, SHIGETA WRIGHT

EGGS OF NORTH AMERICAN BIRDS (LIFE SIZE)

EGG



BY COURTESY OF THE BRITISH MUSEUM (NATURAL HISTORY)

EGGS OF BRITISH AND EUROPEAN BIRDS (THREE-FOURTHS LIFE SIZE)

number is two. It is a point of interest that the cuckoo, as an experienced social parasite, removes at least one of its host's eggs from the nest before depositing its own. Such considerations have led to the notion that a bird can count—to the extent at least of recognizing its own clutch number.

See also BIRD: *Breeding;* EMBRYOLOGY; EMBRYOLOGY, HUMAN; INVERTEBRATE EMBRYOLOGY; VERTEBRATE EMBRYOLOGY; FERTILIZATION; REPRODUCTION.

BIBLIOGRAPHY.—L. B. Arey, *Developmental Anatomy,* 6th ed. (1954); *The Cambridge Natural History* (1895–1909); H. B. Cott, *Adaptive Coloration in Animals* (1940); L. H. Hyman, *Invertebrates* (1940); L. Koch (ed.), *The Encyclopedia of British Birds* (1955); G. K. Noble, *The Biology of the Amphibia* (1931); O. Pettingill, *A Laboratory and Field Manual of Ornithology* (1948); A. L. and A. J. S. Romanoff, *The Avian Egg* (1948); B. H. Willier, P. Weiss and V. Hamburger (eds.), *Analysis of Development* (1955); E. B. Wilson, *The Cell in Development and Heredity* (1928). (D. RK.)

EDIBLE EGGS

Eggs of various animals have long been eaten and enjoyed by man. Although eggs of turtles, crocodiles, sturgeon (caviar) and others are eaten, birds' eggs are consumed in by far the greatest numbers. Whether gathered from the nests of wild fowl or derived as a product of poultry husbandry, eggs among most peoples are highly relished as a delicacy or, if more abundant, as a staple food. Eggs preserved for years in caustic lime have been considered delicacies in China for generations. During aging, the contents shrink to a dark gelatinous mass having none of the obnoxious odours and tastes normally associated with decomposition (*see* FOOD PREPARATION).

By many peoples the avian egg has long been considered a symbol of the earth, life or the soul. It occurred prominently in many early fertility rites, and its use at Easter—originally perhaps a vestige of the early rites but soon adopted as a symbol of the resurrection by early Christians—is universal today in the western world (*see* EASTER).

Chicken Eggs.—The prolific chicken hen has long produced most of the eggs consumed as food. The chicken migrated from its area of origin, assumed by some to be the jungles of India, during prehistoric and historic times, to most parts of the world. Chickens were known in China as early as 1400 B.C. The Romans found them well established in England, in Gaul and among the Germans. Most likely, they did not reach the western hemisphere until the second voyage of Christopher Columbus in 1493.

Chickens are the most widely distributed of domesticated fowl and perhaps the most widely distributed of all animals kept as food producers. The majority of farmsteads and rural villages the world over, especially in the eastern hemisphere, have their flocks of chickens, and the most remote village markets rarely lack eggs.

Modern commercial production of eggs, as distinguished from free range production and occasional subsistence gathering for home use or village markets, is a complex industry. It utilizes much of what modern science has revealed about genetics, animal nutrition and housing and about the preservation and marketing of perishable foods. In many instances automation has been applied to feeding and other operations, and flock size has been increased to many thousands.

Geneticists have developed special breeds of chickens that excel in the production of large numbers of high-quality eggs. The Mediterranean class, particularly several varieties of the Leghorn breed, have been outstanding in pedigree or in-line breeding for high egg production. Some of the dual-purpose or American class birds, however, have also attained importance in pedigree breeding and in developments involving inbreeding, single crossing of inbred lines and multiple (top) crossing of lines. Once basic genetic quality is assured, management of the pullets during their development, and prevention and control of disease are important (*see* POULTRY AND POULTRY FARMING).

Quality standards for eggs are based on the interior changes that occur, particularly in the white (albumen) and yolk. (Shell colour—whether white or brown—is not a standard of quality; colour is merely a breed characteristic.) These changes may be observed by candling the whole shell eggs or by examining the exposed contents with special instruments.

Candling involves holding the shell eggs before a penetrating light in a darkened room, noting internal characteristics of white, yolk and air cell and, on the basis of quality, assigning the appropriate alphabetical grade, as AA (highest quality), A, B and C (lowest quality). Often eggs are classified for marketing according to size, from the largest (jumbo) to the smallest (peewee); these are actually weight standards specified on a per dozen basis.

With instrument measurement, quality is usually given a numerical expression, as for example in Haugh units, which correlates thick white height and egg weight.

Eggs may show all degrees of quality—from the finest (usual) to the inedible (rare)—immediately after they have been laid, and thereafter gradually decrease in quality. On being broken, a fresh grade AA egg spreads over much less area than do eggs of lower grades and displays a round and upstanding yolk, a large portion of firm, thick white and a medium-sized outer margin of thin white.

Annual statistics on world egg production and prices to producers are given in the *Britannica Book of the Year.*

Composition of the Chicken Egg.—The egg is readily separable into three obvious parts: the white, the yolk and the shell. The white, which accounts for about 58% of the weight of the entire egg, is about 87% water and 12% protein. It is composed of four parts: (1) an outer envelope of fluid white; (2) an envelope of firm white; (3) an inner envelope of fluid white; and (4) two small filaments of firm white (chalazas) which hold the yolk in position. The yolk, about 32% of the weight, is approximately 49% water, 32% fat and 17% protein, and consists of alternate concentric layers of light and dark yolk. The shell, about 10% of the weight, is made up of several layers of calcium carbonate and an outer cuticle or skin. Between the inner surface of the shell and the outer envelope of fluid white are the outer and inner shell membranes.

Egg proteins supply all of the amino acids necessary for human growth and maintenance and most of the essential minerals. The vitamin content varies with the season and with the hens' diet. Vitamin A value ranges from 620 international units per 100 g. in winter to 1,460 in summer. Riboflavin varies from 140 micrograms per 100 g. of egg from hens on a basal diet to 440 micrograms from those on a diet rich in alfalfa meal and nonfat milk solids. The yolk contains considerable Vitamin D. (For recommended daily dietary allowances, *see* NUTRITION.)

Processing Eggs.—Following the grading and the rejection of inedible or imperfect shell eggs, the product is sent to nearby retail markets, to more distant wholesale markets, to plants processing egg products or to cold storage for later use. If cold storage periods approach six months, a spray or dip treatment with edible-grade mineral oil is advisable. The oil seals the shell pores and retards quality loss. In countries lacking refrigeration facilities a hot oil (thermo-stabilization) treatment may be used to extend shelf life.

Shell eggs sent to processing plants go into the preparation of liquid blends of whole eggs, of egg yolks, of egg whites or of special mixtures. These products are suitably packaged and frozen at sub-zero temperatures or are dehydrated as egg solids (dried) products. Egg products are used for their binding, emulsifying, leavening and nutritive properties, and for their contribution of colour or flavour to, for example, baked goods, baby foods, candies, ice creams, noodles, prepared cake mixes and salad dressings.

In the production of liquid eggs for freezing or drying the contents of shell eggs are broken out and separated manually or mechanically. All liquid egg products are subjected to thorough mixing and straining to assure homogeneity and freedom from membrane and shell fragments. During the mixing process salt or various carbohydrates may be added for the production of special mixtures. At this point, samples are usually drawn for chemical and bacteriological examination. Until directed to freezing or drying processes, the liquid products are held under refrigeration to retard bacterial spoilage. Liquid eggs may also be pasteurized before freezing or drying to destroy pathogenic bacteria.

Egg solids products are prepared by evaporating the water from

liquid blends in drying units employing heated filtered air as the drying medium. In spray drying, the egg blends are centrifugally or pressure atomized into hot, drying air moving through the system. Dehydration is instantaneous, and the resulting solids are of a powdery appearance and texture. The product is discharged from the dryer by gravity or by mechanical conveyors or sweeps, or a combination thereof, for screening and packaging. In tray drying, the liquid whites are simply measured into aluminum drying trays, which are then passed through cabinet or tunnel drying chambers. Evaporation proceeds at a much slower rate (6–24 hours), and the resulting solids possess a crystalline or granular appearance. An emulsion of beeswax and edible mineral oil, applied to the trays before filling, facilitates removal of the dried product.

Whites are dehydrated by either spray or tray process. Whole egg and yolk products, however, are usually dried by spray process, as the lower temperatures and longer drying intervals encountered in tray drying favours bacterial decomposition.

Liquid whites are desugarized before drying. Desugarization is accomplished by depleting the natural glucose content of the whites with a bacterial or yeast fermentation process or with an enzyme conversion process that converts glucose to gluconic acid. Ingredients called whipping aids may be added to desugarized liquid whites before drying, to enhance whipping when the dried product is reconstituted. Desugarization can be omitted in the production of whole egg and yolk solids products, but if observed, products of greater stability result. When the desugarization step is omitted, the finished egg solids products are susceptible to undesirable chemical reactions during storage and to consequent deterioration of colour, flavour, odour and solubility.

Egg white solids are extremely stable under dry, room-temperature storage conditions. However, whole egg and yolk solids, which have a relatively high fat content, are subject to oxidative changes. Shelf life of these products may be extended by cool, dry storage or by vacuum-gas packaging in nitrogen or nitrogen and carbon dioxide atmospheres.

Duck Eggs.—In some parts of the world the duck is important as an egg producer. Certain types of ducks that have been bred for egg laying have equaled or exceeded the laying records made by chickens. Many have produced 300 to 360 eggs per year. The eggs are larger and heavier than chicken eggs, averaging more than 2½ oz. each and varying in shell colour from white through several shades of green to black. Duck eggs have not been widely produced for food in the United States, but they are popular in other parts of the world, notably western Europe and the orient. The Indian Runner egg-class duck has been introduced widely, but demand for the eggs has not been active except at Easter time.

See also references under "Egg" in the Index volume.

BIBLIOGRAPHY.—*General.*—The egg as a symbol is considered in Sir James G. Frazer's *The Golden Bough*, 12 vol., 3rd ed. (1907–15, reissued 1935–36 and 1955), and in *Funk and Wagnall's Standard Dictionary of Folklore, Mythology and Legend*, ed. by Maria Leach, vol. 1, entry under "egg" (1949). Some account of the use as food of eggs of various animals may be found in A. H. Verrill's *Strange Customs, Manners and Beliefs*, ch. 13 (1946).

Chicken Eggs and Egg Products.—A. William Jasper, "Poultry Farm Practices and Egg Quality," U.S. Department of Agriculture, *Marketing Research Report No. 22* (May 1952); W. E. Hauver, Jr., J. A. Hamann and Lester Kilpatrick, "Egg Grading Manual," U.S. Department of Agriculture, *Agriculture Handbook No. 75* (July 1961); E. M. Mrak and G. F. Stewart, *Advances in Food Research*, vol. 1, 2, 3 and 6; C. O. Chichester, E. M. Mrak and G. F. Stewart, *Advances In Food Research*, vol. 10 (1960); Harry W. von Loesecke, *Drying and Dehydration of Foods*, 2nd ed. (1955); J. W. Koudele and Edwin C. Heinsohn, "The Egg Products Industry of the United States," *Bulletin 423, North Central Regional Publication 108* (May 1960). (J. M. Go.)

EGGLESTON, EDWARD (1837–1902), U.S. author, best known for *The Hoosier Schoolmaster*, was born in Vevay, Ind., on Dec. 10, 1837. Before he was 19 he had become an itinerant preacher, but the hardships of circuit riding broke his health within six months. He then held various pastorates in the Minnesota Methodist conference for several years, and later (1874–79) was minister of the Church of Christian Endeavour in Brooklyn. The remainder of his time was spent in literary work as associate editor of the *Little Corporal* (1866–67), a juvenile paper, and as editor of the *National Sunday School Teacher* (1867–70) in Chicago; for a short period he edited the *Independent* and *Hearth and Home* in New York. He died Sept. 4, 1902.

In all of his work he was true to his early ideal—to write with "photographic exactness" of the real west. The most popular of his books for adults, *The Hoosier Schoolmaster* (1871), is a vivid study of backwoods Indiana. *The End of the World* (1872) deals with the expectation of the end of the world in the mid-19th century; *The Mystery of Metropolisville* (1873) portrays a Minnesota boom town; *The Circuit Rider* (1874) describes frontier religious phenomena; *Roxy* (1878) concerns itself with border politics; *The Graysons* (1888) centres about the young Abraham Lincoln. The later novels and the children's books are of less significance. Toward the end of his life he turned to the writing of history and produced several popular compendiums and two volumes of a projected *History of Life in the United States*—*The Beginners of a Nation* (1896) and *The Transit of Civilization* (1900)—that were pioneering efforts in the field of social history. His novels, while sometimes inclined to exaggeration, are important first-hand pictures of a way of life that has long since passed away.

See William Randel, *Edward Eggleston* (1946), a critical biography with a good bibliography.

EGGPLANT (*Solanum melongena*), a tender perennial plant of the nightshade family (Solanaceae; *q.v.*), closely allied to the potato and known also as eggfruit, aubergine and guinea squash. It is native to southern and eastern Asia, where since remote antiquity it has been cultivated for its fleshy fruit used as a vegetable. For this purpose it is usually grown as an annual. It has an erect, bushy, rather scurfy stem, sometimes armed with a few spines, large ovate, slightly lobed leaves and pendant, violet, usually solitary flowers, about two inches across. The fruit is a large, pendant, egg-shaped, oblong or somewhat pear-shaped berry, varying in length from 2 in. to 12 in., dark purple, white, red or yellowish, sometimes striped and with a shining surface.

J. HORACE MCFARLAND CO.

RIPENING EGGPLANTS (SOLANUM MELONGENA), BLACK BEAUTY VARIETY

The snake eggplant (*S. m. serpentinum*) is a curious variety with very narrow, elongated fruit, sometimes a foot long and curled at the end. The dwarf eggplant (*S. m. depressum*) is a small, nearly smooth and spineless form, with nearly entire leaves and purple fruit about four inches long.

Eggplant requires a warm climate and is grown extensively in eastern and southern Asia. It is grown for market in the United States, chiefly in Florida, New Jersey and Texas. In many states it is grown in kitchen gardens. The plants are started in protected beds and transplanted to the field or garden about 2½ ft. apart in rows about 3½ ft. apart. (V. R. B.; X.)

EGHAM, an urban district of Surrey, Eng., on the Thames, 17 mi. W.S.W. of London by road. Pop. (1961) 30,553. The church of St. John the Baptist, thoroughly restored in 1948–49, contains monuments by John Flaxman. High above the right bank of the river, on Cooper's hill, was the Royal Indian Civil Engineering

college, now a teacher's training college. The hill has been celebrated by the verse of Sir John Denham (1643) and of Alexander Pope, whose "On Cooper's Hill eternal wreaths shall grow" inspired the placing on a spur of the hill in 1953 of a memorial to the air forces of the commonwealth. The wooded slopes of the hill overlook the Magna Carta field of Runnymede (*q.v.*) which is in the parish. Light engineering and the extraction of excellent gravel are the chief industries. The Royal Holloway sanitarium (1885) and the Royal Holloway college for women (opened in 1886), a part of London university, were founded by Thomas Holloway, a maker of patent medicines.

The district also includes the picturesque grounds and artificial lake of Virginia water at the south end of Windsor Great park; the Wentworth golf course; Englefield Green; Thorpe with a Norman and Early English church which once belonged to Chertsey abbey; and the Tudor mansion of Great Fosters, now a hotel. Egham races were held yearly on Runnymede until 1886.

EGILL SKALLAGRÍMSSON (*c.* 910–990), Icelandic poet and adventurer and hero of the *Egils Saga*, probably written *c.* 1220 and often ascribed to the famous Icelandic historian Snorri Sturluson. Although born and bred in Iceland, Egill spent many years abroad. He visited King Aethelstan (924–939) in England and, *c.* 948, fell into the hands of his enemy Eric (Eirík) Bloodaxe, then ruling in York. Egill redeemed his head with an elaborate poem, the *Hofuðlausn* ("Head Ransom"), composed in honour of Eirík. Among other important poems ascribed to Egill are *Arinbjarnarkviða* ("the Lay of Arinbjorn") and *Sonatorrek*, a lament on the death of his son. Egill is remembered as a master of the intricate scaldic measures and imagery. The *Egils Saga* and Egill's verses were translated into English by E. R. Eddison (1930). (G. T.-P.)

EGLANTINE, a fragrant erect branching rose, the sweetbrier, *Rosa eglanteria* (*R. rubiginosa*). The sweetbrier has become extensively naturalized in eastern North America, growing abundantly in pastures, thickets and roadsides from Nova Scotia and Ontario southwestward to Tennessee and Kansas.

Eglantine is frequently alluded to in the writings of English poets, from Chaucer onward. Milton, in *L'Allegro*, is thought by the term "twisted eglantine" to denote the woodbine honeysuckle, *Lonicera periclymenum*, which is still known as eglantine in northeast Yorkshire.

EGLINTON, EARLS OF, a title held by the Scottish family of Montgomerie from 1508. Robert de Mundegumbri (d. 1178) is the first of the name found in Scotland, but the certain affiliation ascends only to Sir John Montgomerie of Eaglesham (d. *c.* 1398), who fought at Otterburn in 1388 and captured Sir Henry Percy (Hotspur). He acquired the baronies of Eglinton and Ardrossan by marriage. His grandson Sir Alexander Montgomerie (d. *c.* 1470) became a lord of the Scottish parliament as Baron Montgomerie in 1445. Sir Alexander's grandson HUGH (1460–1545), 3rd lord, was created earl of Eglinton in 1508. He was succeeded by his grandson HUGH (d. 1546), 2nd earl, whose son HUGH (*c.* 1531–85), 3rd earl, was a supporter of Mary Stuart, for whom he fought at Langside in 1568, and of the Roman Catholic Church. His son and successor, HUGH (1563–86), 4th earl, was murdered in 1586 in a feud with the Cunninghams. In 1612, on the death of HUGH (1584–1612), 5th earl, the earldom and chiefship passed under an entail, confirmed by the crown in 1615, to SIR ALEXANDER SETON (1588–1661), 6th earl, the third son of the 1st earl of Winton. Nicknamed "Graysteel," he was a staunch Presbyterian and became a covenanter. He took the name of Montgomerie. His son HUGH (1613–69), 7th earl, was a royalist in the English Civil War. ARCHIBALD (1812–61), 13th earl, held the famous tournament at Eglinton castle described by Benjamin Disraeli in *Endymion*. He was lord lieutenant of Ireland in 1852 in Lord Derby's ministry and again in 1858–59. The title has descended to ARCHIBALD WILLIAM ALEXANDER (1914–), 17th earl.

See Sir W. Fraser, *Memorials of the Montgomeries, Earls of Eglinton*, 2 vol. (1859). (T. L.)

EGMOND (EGMONT), **LAMORAL,** 4TH COUNT OF (1522–1568), the most popular of the Netherlands magnates on the eve of the great revolt and the most illustrious victim of Spanish oppression, was born at La Hamaide, Hainaut, on Nov. 18, 1522, the younger son of John, 2nd count of Egmond, head of a branch of the ancient house of Egmond (in the county of Holland), whose senior line had already produced three dukes of Gelder (Gueldres; *see* GELDERLAND). He succeeded to the countship in 1541. On his behalf the county of Gaveren (Gaure), inherited from his mother Françoise de Luxembourg, was raised to the rank of principality in 1553. In 1544 he married Sabina of Bavaria (1528–78), daughter of John II, count palatine of Simmern.

Appointed knight of the Golden Fleece in 1546, Egmond was in high favour with the Holy Roman emperor Charles V. In 1554 he was sent to England as the representative of Philip of Spain to ask for the hand of Queen Mary. His distinguished military career was crowned when, as commander of the light cavalry in the Netherlands, he played a brilliant part in the victory over the French at St. Quentin (Aug. 10, 1557) and won the battle of Gravelines (July 13, 1558). The idol of his countrymen, he was in 1559 nominated stadtholder of Flanders and Artois, member of the council of state and commander of the Spanish troops in the Low Countries.

Egmond and other leading noblemen resented King Philip II's attempt to convert the Netherlands into a Spanish dependency (*see* NETHERLANDS, THE: *History*). In conjunction with William the Silent, prince of Orange, and the count of Horn (Philip de Montmorency) he succeeded in 1564 in removing Cardinal Granvelle, the virtual head of the government, from the council of state and in driving him out of the country. Discontent, however, remained, and in Jan. 1565 Egmond went to apprise Philip of the situation and to press for a convocation of the states-general and for a mitigation of the religious persecution. He returned without having accomplished anything.

Much disappointed, Egmond approved of the league that the lesser nobility formed in Dec. 1565, but declined to take any step savouring of actual disloyalty to his sovereign. He withdrew to his government of Flanders, where he severely repressed the iconoclastic riots. In the spring of 1567 he took the oath of allegiance demanded by the regent Margaret of Parma, but in the eyes of Philip he was a marked man. In the summer the duque de Alba arrived with an army to supersede the regent and to restore order. William of Orange fled to Germany, having warned Egmond and Horn of the dangers threatening them. Alba treacherously seized them both (Sept. 9) and imprisoned them at Ghent. The intervention of the other knights of the Golden Fleece and of several princes was of no avail. The invasion of Louis (*q.v.*) of Nassau in the north in the spring of 1568 sealed the prisoners' fate. On June 3, 1568, they were removed to Brussels, where they were condemned to death on June 4 for high treason, felony and *lèse-majesté* by the "Council of Blood." Next day Egmond was beheaded, with Horn and other nobles, in the Brussels market-square.

See T. Juste, *Le Comte d'Egmont et le comte de Hornes* (1862); R. Avermaete, *Lamoral d'Egmont* (1943); S. de Troeyer, *Lamoraal van Egmont . . . in de jaren 1559–64*, Verhand. Kon. Vlaamse Acad. v. Wetensch, Kl.d.Lett., no. xl (1961). (A. G. J.)

EGMONT, EARLS OF. In 1733 JOHN PERCEVAL (1683–1748) was created earl of Egmont, taking the title from his family home in Egmont, County Cork, although he claimed descent from the Egmonts of Flanders. The second son and heir of Sir John Perceval, Bart., he had become an Irish privy councilor in 1704, and was made Baron Perceval in 1715 and Viscount Perceval in 1723. As the partner of J. E. Oglethorpe (*q.v.*), he took a close interest in the foundation of the American colony of Georgia. Egmont is also remembered as a letter writer, biographer and genealogist. JOHN PERCEVAL (1711–70), 2nd earl, eldest son of the 1st earl, was a leader of the parliamentary opposition under George II. His strenuous efforts to gain an English title after George III's accession were successful in 1762 when he was made Baron Lovel and Holland. Egmont became first lord of the admiralty in 1763. Although an agent in the destruction of the marquess of Rockingham's ministry of 1765–66, he disagreed with the foreign policy of Rockingham's successor, William Pitt, earl

of Chatham, and resigned the admiralty in 1766. His seventh son was the prime minister Spencer Perceval, whose descendant FREDERICK GEORGE MOORE PERCEVAL (1914–) established his claim to the Egmont titles in 1939, becoming the 11th earl.

EGO, central in the organization of human personality, is that portion experienced as the "I" which perceives, remembers, evaluates, plans and in other ways is responsive to and acts in the surrounding physical and social world. Ego (Latin for "I") comprises, in Freud's term, the executive functions of personality. Ego also describes "me," the self which is known and valued. The ego gives continuity and consistency to behaviour by providing the personal point of reference which relates the events of the past (retained in memory) and actions of the present and of the future (represented in anticipation and imagination). The ego is not coextensive with either the personality or the body, although body concepts form the core of early experiences of self. The ego, once developed, is capable of change throughout life. Both the executive functions and the experienced self alter under conditions of threat, illness and changes in life circumstances.

The terms ego and self are occasionally used interchangeably, occasionally contrasted, but there is no consistent usage. The single term ego is used here to describe two sets of related phenomena, sometimes distinguished as ego and self: (1) those executive functions, involving perceiving, thinking, planning and others, through which behaviour is organized and directed (ego-as-process); and (2) the organized attitudes toward himself which comprise the individual's experienced self (ego-as-object).

Ego Development.—The newborn human infant reacts to but cannot control, anticipate or alter sources of stimulation, whether arising from the world outside or from biological processes within. One can never know whether the infant actually experiences the "big, blooming, buzzing confusion" William James proposed, but he certainly lacks the neurological and psychological mechanisms for integrated perception and action. Perception is primitive and diffuse, motor activity gross and unco-ordinated and self-locomotion impossible. Learning is limited to the simplest type of stimulus-response conditioning.

The infantile ego develops in relation to the external world and reflects, as psychoanalysis has emphasized, the helpless and dependent infant's efforts at altering and alleviating painfully intense stimuli. Perhaps if all needs were gratified at the very moment of arousal, there would never be any differentiation of an ego or sense of reality. Such a possibility being inconceivable, mechanisms evolve for maintaining the tension state while seeking means by which gratifications can be obtained, and these mechanisms in time develop into increasingly complex forms of mastery.

At the outset, perception and motility are closely tied. Immediately on stimulation there is motor discharge. The development of objective perception and directed action require the capacity to delay discharge and gratification. To delay action, while tolerating the consequent tension, is the basis of all more advanced ego functions and is prototypic of the ego's role in the later economics of personality functioning.

Infantile perception is closely related to the state of the biological drives. The world is probably vaguely apprehended as a potential source of satisfaction or of threat, and reality as such is distorted and not perceived in its own terms. Objective perception supposes psychological distance between the perceiving ego and the external sources of perception data, which allows the judgment and choice impossible under conditions of immediate discharge. The separation of stimulation and response allows the interposition of more complex intellectual activities such as are involved in thinking, imagining and planning. Instead of reacting directly, the ego has the capacity to test reality vicariously, to fantasy the consequences of one or another course of action, and to decide upon future directions to achieve its ends. The accumulation and retention of memories of past events is the necessary background for internal processes of thought and judgment. The child avoids the fire in advance of a burn, rather than simply withdrawing after the painful contact. The acquisition of language, started during the second and third years, provides a powerful tool for the development of logical thought processes as well as allowing communication and control of the environment. The more primitive and concrete preverbal thought, dependent on pictorial representation, gives way to abstract thought.

Progress from immediate to directed behaviour, from prelogical to rational thinking, is slow and moves through a number of intermediate stages during childhood. Even in physical maturity, persons differ considerably in the forms and effectiveness of ego functioning. An important dimension has been characterized as ego strength. The person of strong ego has the following characteristics: he is objective in his apprehension of the external world and in self-knowledge (insight); his activity is organized over longer time spans and he is thus able to maintain schedules and plans; he can follow resolves, and choose decisively among alternatives; he is not overwhelmed by his drives and can direct them into socially useful channels; he can resist immediate environmental and social pressure while contemplating and choosing a self-selected course. On the other hand, the ego-weak person is more like the child: behaviour is impulsive and immediate; the perception of reality and of himself is distorted; he is less capable of productive work because his energy is drained into the protection of warped and unrealistic self-concepts; he may be burdened by neurotic symptoms.

It should not be assumed that with the attainment of mature ego functioning earlier modes are no longer possible. The tendency to immediate discharge, hallucinatory wish fulfillment, distorted perception, concrete pictorial thinking and other forms of primitive behaviour may reappear in states of fatigue, intoxication, during sleep in the dream, and in the half-conscious conditions preceding and following sleep. In extreme form, the psychosis represents an enduring regression of ego functioning. However, even in a person fully conscious and physically healthy, prelogical emotional thinking can replace reality-oriented thinking, particularly following the arousal of emotional conflicts or under stress. While it is probably true that the mature person can never literally regress to the infantile state and behave precisely as he once did, under some conditions more primitive forms of ego functioning occur.

Controlled regression "in the service of the ego" is sometimes necessary in art, creative thinking, play and certain forms of spontaneous interpersonal relations. While inability to prevent a loss of concern with reality and control of impulses is ordinarily a sign of weakness in the personality, the capacity to reduce ego control at will is an indication of strength.

Emergence of Ego Identity.—In infancy, although there is consciousness, in the sense of responsiveness to stimulation, self-consciousness does not exist. Initially, there is no distinction between the body and the surrounding physical world. The infant treats portions of his body in the same fashion as he does the bars of the crib or a toy, even to hitting or hurting them. The fingers or toes are viewed with the same detachment as the movements of a light or another person. The body, as a continuous source of sensation, in time becomes represented as an early "body ego." Even to the age of two, however, the child will not recognize himself as a unique person when looking at his image in a mirror. The ego emerges as the object of self-experience as the child distinguishes himself from other objects, which requires the abilities for accurate perception, memory and the like discussed above.

The first human object for the child is the mother. His contacts with her are not only most frequent and close, but she is the first and for some years the most important source of gratification or deprivation. She is recognized as the first external agent upon whom his well-being depends. The ego is formed in relation to the mother, and only later do other social objects become important.

The experience of ego identity depends on the acquisition of memories and the development of language. Without memory there can be no continuity between experiences. The individual knows himself as an individual because he retains personally relevant experiences in which "I" did something, or something happened to "me." Out of the experienced continuities of events, places and actions there emerges the conception of the continuing

ego common to them all. After language develops, the objects of the world, including the self, become named and hence defined in a socially shared fashion. A person's name is a reference point of self-experience, since it sets him off from others, and becomes the symbol by which he is identified, addressed and placed in the organization of the family and other social groups. Throughout life a man may fight to preserve his "good name." In the magical practices of some primitive peoples, cursing a man's name is believed to do him physical harm. Ego development is also seen in the child's use of personal pronouns. Early in the second year the child describes himself variously in first, second and third person terms—"I am here. You go, John." But during the third year he becomes "I." This corresponds to an important phase in growth marked by negativistic behaviour. The child simultaneously asserts his unique identity and his new-found capacities for active mastery by standing against the world and refusing to do its bidding. The pronoun is used with complete self-reference in the sentence "I will not." Similarly, in the make-believe of early play the child releases and resumes his identity, for example, by assuming false names or hiding, and gains reassurance that he is the same person throughout. Such play also gives the child the experience of being able to control his environment, preparing him for mastery in later, more serious situations. Later, organized games provide opportunity for relating to others under differing conventions, for assuming and changing roles and for gaining perspective on the child's own and others' behaviour from the vantage of the different roles played.

George Herbert Mead stressed the importance of social interactions in the formation of the self, pointing to the fact that the early conversations and interplay between the child and others become in time internalized and the self or ego forms as the child incorporates within himself a conception of the "generalized other," representing the attitudes of others toward him and social expectations generally. As he conceptualizes others' attitudes, the concept of the self develops as distinct from the actions of others. At the same time, the child becomes socialized to the values of his group, since the concept of the "generalized other" involves knowledge of the socially shared rules and prohibitions. Evolving from the pattern of identification with parents, siblings and peers, and being conditioned by school, work and other groups later in life, the individual's self-concept often includes diverse and sometimes contradictory elements. Behaviour in the business office, with the family at home, and at play can be so different as to suggest that any person is actually many social selves. While there are certainly behaviours specific to different social roles, these are not so many different egos, but rather sets of actions adopted as necessary in terms of the ego's perception of the social requirements in the particular case. Indeed, the ability to move adaptively from one social role to another, without losing identity, is a sign of a well-integrated ego. Less adequate functioning is seen at either extreme—either inability to be anything other than "oneself," or a chameleonlike shifting of behaviour to each new social demand.

Ego in Adolescence and Old Age.—Adolescence and old age are two periods in the life cycle when previously stable ego organization must be altered because of rapid biological and social change. There are sudden changes in sexual physiology and in social expectations which make the adolescent's earlier body and self-concepts inadequate. Both the newly emergent sexuality and the demands of adult responsibility are new and threatening, and the adolescent has few clear conceptions or already perfected behaviour appropriate to them. The acceptable dependency of childhood has to yield to adult independence, but the boundaries between acceptable and forbidden behaviour are unclear. The adolescent experiences intense feelings of inadequacy, anxiety and guilt, and is given to impulsive outbursts and wish-fulfilling fantasy. Identification moves from the parental models to the peer groups, and fads and exaggerations of dress and behaviour are sought to give distinction and status to the group.

Old age is marked by a loss of sexual and productive capacity. With lessened competence, independence may revert to dependence, perhaps abruptly at retirement from work. Self-esteem is reduced and sometimes feelings of worthlessness supervene. There is a constriction of ego functions as the person becomes less venturesome, his habits become more fixed, and he approaches problems in terms of traditional solutions, rather than risk failure in novel attempts. Social as well as personal behaviour becomes more conservative.

The degree of disruption varies greatly among societies, depending on the sharpness of the transitions and the social values involved. For example, in societies which accept sexuality casually and in which there is continuity between child and adult roles there is less unhappiness in adolescence than there is in western society. Similarly, where the experience of age is valued equally with the strength and beauty of youth, and where the family organization can provide a productive role for the aged, as in many peasant and primitive societies, the transition into old age is less disturbed. In western European and U.S. society, particularly under conditions of urban living, the strains on ego organization are great.

Ego Extension and Ego Ideal.—As the object of self-experience, the ego is the organized attitudes toward himself which constitute an individual's personal reply to the question "What am I?" It consists not merely of internal experiences but includes also parts of the body, possessions, other persons, social group memberships and the philosophies or ideologies to which the person holds allegiance; in short, all those aspects of his life which are, in Gordon W. Allport's term, "propriate," experienced as of particular personal relevance, and which are important to his identity and individuality. For some, the scope of the personally relevant is broad, for others narrow, although maturity normally involves a widening range of persons and social values felt as propriate. The capacity to love another person or to identify with an ideology which transcends immediate gain depends upon this capacity for ego extension. The common use of terms such as "egotistic," "selfish" or "selfless," suggesting that having ego is synonymous with narrowed self-concern and self-love and that altruism depends on a lack of ego, is misleading. The egoless condition of infancy is perhaps the most literally selfish state, while genuine identification with the welfare of others requires a mature ego.

The ego develops as the immediacy of behaviour is reduced, and from the outset is concerned with time, both past and, perhaps more important, future. In its executive functions, the ego organizes behaviour in time sequence, gives priorities and plans future actions. As the object of experience, the ego is correspondingly oriented to the future. Within the organized attitudes toward the self are also included ideals toward which the person strives. Thus, the question "What am I?" is paralleled by "What should I be?" or "What might I become?" Psychoanalysis has given this aspect of the ego the separate designation of "ego ideal"; others have used terms such as "ideal self-concept." Behaviour is directed by internal goals the person wishes to achieve. A presently distasteful situation (a dull lecture) is tolerable if an agreeable end is anticipated (receiving a college degree). Shame occurs when the person perceives himself as falling short of his ideals, though these in turn may be unrealistically high. In childhood the ego ideal is formed largely in the image of respected family figures on whom the child would model himself. Later, however, the ego ideal depends more on abstracted values than on identification with particular persons.

Ego Defense and Psychopathology.—Although the ego is the agency of conscious, self-initiated behaviour, it is protected in these functions by unconscious processes, extensively studied and termed by psychoanalysis "mechanisms of defense," which often lead to distortions of reality-appropriate behaviour. When the person is threatened, either by the emergence of dangerous impulses or by external stress, defense mechanisms come into play. These have the double purpose of providing some degree of gratification, perhaps substitute, for motivated behaviour, while at the same time protecting self-esteem. Anxiety, originally a diffuse, primitive effect, comes to be used as a signal of danger indicating the need for defense. The ego defenses serve to reduce this anxiety by altering perception, thought or action by removing or

circumventing the dangerous condition, although the ultimate purpose is to protect the ego from disorganization and regression. While blocking greater dangers, the defensive maneuvers themselves limit ego functioning and are the source of neurotic symptoms. In the healthy person defenses appear in emergency; it is characteristic of the neurotic, on the other hand, that defenses operate continuously and in exaggerated forms disproportionate to the actual danger. The processes involved in defense are similar, though perhaps in distorted forms, to normal modes of ego functioning. It is necessary, for example, to isolate and repress in order to achieve abstract and logical thought. A major contribution of modern psychoanalysis, through the study of the defense mechanisms, has been the understanding of neurosis as continuous with normal ego functioning.

Among the commoner defenses are denial, repression, projection, reaction formation, isolation and rationalization. Denial, the refusal to accept unpleasant aspects of reality as true, is found in truest form only in young children and operates in the adult only temporarily and in extreme conditions. Repression, related to denial, represents a less blatant form of distortion and consists of excluding from memory and consciousness ideas or wishes originally associated with dangerous impulses. Although now unconscious, the repressed material remains the source of conflict and necessitates further defensive activity. Projection involves, in addition, the subsequent ascription of the unacceptable impulse to some external person or object. Though based on reality distortion, this externalizes the danger and gives the individual an outside target against which to contend. In its extreme, the mechanism of projection becomes clinical paranoia.

Although the defenses allow the personality to operate during crisis without excessive anxiety or guilt, their action opposes the ego's major functions of integrating, synthesizing and directing behaviour toward purposeful goals. Emphasis on defense has led some to the false assumption that human behaviour consists principally of developing means for avoiding painful reality or dealing positively with life problems. In the period after World War II, there was increased emphasis in psychoanalysis on the adaptive functions of the ego, which operate in relative autonomy from unconscious drives. This so-called ego psychology is also represented in the work of nonpsychoanalytic students of human personality such as Gordon W. Allport. Although the ego is capable of reality distortion and self-deceit, in the healthy person this is secondary to and in the service of adaptive and reality-appropriate behaviour.

See DEFENSE MECHANISMS; PERSONALITY; PSYCHOANALYSIS; *see* also references under "Ego" in the Index volume.

BIBLIOGRAPHY.—Sigmund Freud, *The Ego and the Id*, Eng. trans. (1927); George Herbert Mead, *Mind, Self and Society* (1934); Anna Freud, *The Ego and the Mechanisms of Defense*, Eng. trans. (1937); Gordon W. Allport, "The Ego in Contemporary Psychology," *Psychol. Rev.*, 50:451–478 (1943); Otto Fenichel, *Psychoanalytic Theory of Neurosis* (1945); Percival M. Symonds, *The Ego and the Self* (1951).

(S. J. K.)

EGOISM, a modern philosophic term denoting the ethical theory that the good is based on the pursuit of self-interest. It is sometimes misused for egotism, the overstressing of one's own worth, achievement and importance. Both come from Greek and Latin *ego*, meaning "I."

Egoist doctrines are not primarily concerned with the philosophic problem of what is the self, but start from common notions of a person and his concerns. They see the perfection of the individual attained through the furthering of his own welfare and profit, allowing that sometimes a man may not know where these lie and must be brought to recognize them. Many ethical theories have an egoist bias. The hedonism (*q.v.*) of the ancient Greeks bade each man seek his own greatest happiness, whatever this might be; Hobbes and Spinoza (*qq.v.*) in different ways hold that self-preservation is the good; those who stress the tending of the individual's conscience and moral growth are likewise egoists in this sense. In contrast with such views are the ethics governed more by man's social aspects; they stress the importance of the community rather than the individual. Under this head come theories such as the cosmopolitanism of the Stoics (*q.v.*); ethical systems that centre on the survival of a tribe or nation; and utilitarianism (*q.v.*), which seeks the greatest happiness of the greatest number. These are all forms of what Auguste Comte called altruism (*q.v.*). However, the distinction cannot always be neatly drawn. For one thing, liberal utilitarian ethics tries to show that the two apparently opposed attitudes really coincide: universal welfare is said to be best furthered by each and everyone attending to the promotion of his own best interests. Likewise, it is clear that blind and uninformed egoism will conflict with the social character of the human animal. An enlightened egoism, on the other hand, recognizes that private and individual ends are often best-served by attending to the needs and aspirations of others. This can no longer be distinguished from altruism in effect, though the motives may be different. Indeed, on grounds such as these it has at times been said that all our actions are really inspired by egoism, since what we do for others in the end ministers to our own conscience or welfare. In any case, a moral action helping others also helps the moral agent to realize his own nature, provided the action is voluntary and consciously seen as moral. At this stage, however, the term egoism becomes pointless and uninteresting for the purpose of drawing distinctions between ethical theories; if in the very nature of the case we are all necessarily egoists, it becomes idle to harp on it. Such general claims must not obscure vital distinctions between different ways of life. *See* also ETHICS.

EGRET, the name used for several graceful herons (*q.v.*), usually with white plumage and ornamental, silky nuptial plumes. These birds' habits are generally those of other herons but they have elaborate mating displays involving the plumes (*see* COURTSHIP, ANIMAL). Two moults a year are the rule, the plumes being worn only during the breeding season. The long silky plumes, most developed as a tuft on the back, owe their loose, lacy texture to the wide spacing of the barbs and the lack of barbicels. The word egret has been transferred to the plumes only, also called aigrettes, which are highly prized as ornaments in oriental ceremonial dress and were formerly used in the millinery trade. The high prices formerly paid for the plumes and the vulnerability of the birds, nesting in big breeding colonies, resulted in near extinction of egrets by unscrupulous hunters. Change in fashion and the institution of legal protection have allowed them to increase.

ALLAN CRUICKSHANK FROM NATIONAL AUDUBON SOCIETY

COMMON EGRETS (CASMERODIUS ALBUS) ON NEST

The name is most properly applied to members of the genus known as *Egretta* of the warmer portions of the globe. Included are the little egret, *E. garzetta*, of the old world, about 22 in. long, with the head plumes firm rather than lacy like those on the back; and the greater egret—known in America as the common egret (*Casmerodius albus*) and in England as the great white heron (*Egretta alba*)—of both hemispheres, about 35 in. long and bearing plumes only on the back. Among other species occurring in the old world, there are the related reef herons (*Demigretta*), which have a gray and a white colour phase. The snowy egret, *Leucophoyx thula*, of the Americas, about 24 in. long with filmy recurved plumes on back and head, is the most showy. In the Americas the name reddish egret is also used for a slate- and reddish-coloured heron with a white phase, *Dichromanassa rufescens*.

(A. L. RD.)

EGYPT (AL MISRI), a country occupying the northeastern corner of Africa and the Sinai peninsula which is usually regarded as part of Asia. Strategically situated at the crossroads between Europe and the orient and between north Africa and southwest Asia, it is bounded on the north by the Mediterranean, on the northeast by Israel, on the east by the Gulf of Aqaba and the Red sea, on the south by the Republic of the Sudan and on the west by Libya. The western frontier, defined in the Italo-Egyp-

tian agreement of 1925, starts from a point on the western shore of the Gulf of Sollum and runs in a generally southerly direction to the oasis of Melfa, leaving Jaghbub in Libya and reaching the 25th meridian E. at a point about 70 mi. S. of As Sallum. It then follows this meridian to its conjunction with the 22nd parallel N. The southern frontier was defined in the Anglo-Egyptian Condominium agreement of 1899 as the 22nd parallel N., but for administrative reasons a 20 mi. salient was created north of Wadi Halfa. To the east of the Nile, after other variations, the *de facto* frontier turns northeast from the parallel to the Red sea coast, which it strikes at Bi'r Shalatayn. The northeastern frontier was delimited in 1906; it runs from Rafah near the Mediterranean coast in a south-southeasterly direction to Bi'r Tabah on the Gulf of Aqaba. The total area of Egypt is about 386,100 sq.mi. of which only 3.6% is normally habitable, the rest being desert.

This article contains the following sections:

I. Physical Geography
  1. Geology and Structure
  2. Physiography
  3. Climate
  4. Vegetation
  5. Animal Life
II. Geographical Regions
  1. Western and Southern Deserts
  2. Eastern Desert
  3. Sinai
  4. Nile Valley and Delta
III. The People
IV. Archaeology
  1. Beginnings of Modern Egyptology
  2. 20th-Century Techniques and Excavations
V. History
  A. Ancient Period
    1. Prehistoric Age
    2. Archaic Period
    3. Old Kingdom
    4. First Intermediate Period
    5. Middle Kingdom
    6. End of the Middle Kingdom Through the Second Intermediate Period
    7. New Empire
    8. Libyan Dynasties and the Cushites (Ethiopians)
    9. Saites
    10. Persian Period
  B. Ancient Civilization and Culture
    1. Land and People
    2. Social Life
    3. Administration
    4. Army
    5. Transport and Trade
    6. Arts and Crafts
    7. Science, Medicine and Technical Processes
    8. Literature and Mythology
  C. Ancient Religion
    1. The Pantheon
    2. The Priesthood
    3. Temples and Festivals
    4. Magic and Popular Religion
    5. The Afterlife
    6. The Tomb
    7. Funeral Customs
    8. Eschatological Beliefs
  D. Macedonian and Ptolemaic Period
  E. Roman Period
  F. Medieval Period
    1. Muslim Conquest
    2. Egypt Under the Omayyad and Abbasid Caliphates
    3. The Tulunid Secession
    4. An Egyptian Caliphate: the Fatimids
    5. Egypt and the Crusades
    6. Mameluke Rule
    7. The Ottoman Period
  G. Modern Period
    1. French Occupation
    2. The Years of Anarchy
    3. The Era of Mohammed Ali
    4. Abbas I and Said (1848–63)
    5. Ismail (1863–79)
    6. The First Nationalist Movement
    7. British Occupation
    8. Abbas II (1892–1914)
    9. The Protectorate
    10. Independence
    11. Zaghlul's Ministry
    12. Ziwar
    13. Sarwat
    14. King Fuad's Coup
    15. Treaty of 1936
    16. World War II
    17. Anti-British Agitation
    18. Palestinian War
    19. Wafdist Government, 1950
    20. Mohammed Naguib
    21. Gamal Abd-al-Nasser
    22. The Suez Crisis
    23. The United Arab Republic
VI. Population
VII. Administration and Social Conditions
  1. Government
  2. Living Conditions
  3. Public Health and Social Welfare
  4. Justice
  5. Education
  6. Defense
VIII. The Economy
  1. Irrigation
  2. Methods of Agriculture
  3. Land Ownership
  4. Crop Production
  5. Livestock
  6. Fisheries
  7. Mining
  8. Industries
  9. Trade and Finance
  10. Transport and Communications

I. PHYSICAL GEOGRAPHY

1. Geology and Structure.—Structurally Egypt is almost wholly a part of the Gondwanaland foreland, the entire country being underlain by the crystalline basement complex which occasionally outcrops at the surface. Upon these old rocks of Archean age have been laid down a succession of later rocks, younger to the north. All the geological eras are represented, at least in part, with Nubian Sandstone of Secondary age covering the greatest surface area, and Eocene limestone of Tertiary Age and Pleistocene sediments of the Quaternary Age being next in importance. The oldest rock of the Archean era is the Fundamental Gneiss which occurs in southeastern and southwestern Egypt and extends across the Sudanese frontier. Above these is a complex of highly resistant crystalline rocks and metamorphosed sedimentaries (granites, serpentines, schists) mainly exposed around Aswan, the Western desert and most strikingly, as a result of massive uplift, in the mountain masses of the Eastern desert and in Sinai. Outcrops of these rocks in the Nile valley account for the cataracts. In all, these Archean rocks are exposed over nearly 10% of the present land surface. Among them are veins and pockets of gold, silver, iron, copper and other minerals and such stones as marble, serpentine, obsidian and porphyry. Deposits of the Paleozoic era are scarcely represented, the only identification being sandstones and limestones each side of the Gulf of Suez, but the denudation of the Archean basement throughout this period created huge deposits of sandstones of continental extent and in places 1,600 ft. deep. These sandstones, not homogeneous but generally unfossiliferous, are termed the Nubian Sandstones and in age must range from the Carboniferous to the Cretaceous. They cover nearly 30% of the present land surface, comprising practically the southern third of Egypt and extending far beyond its bounds both to the south and west. To the north they provide the base upon which later marine and fluvial deposits were laid down. The sinking of the land to the north during the later Cretaceous allowed an advance of the sea, and from then until the close of the Pliocene the geological record reveals a series of alternate uplifts and subsidences affecting the northern two-thirds of the country. The Cretaceous sea transgression was the most extensive, penetrating at least 600 mi. from the present coast line and leaving calcareous deposits up to 600 ft. thick. The greater part of the Cretaceous deposits have been masked by younger rocks laid down in succeeding transgressions, so much so that they are exposed over only 12% of the surface.

Eocene limestone formations, representing a submergence of the Cretaceous surface during the early Tertiary period, are the most widespread of the marine deposits in Egypt. They extend over the

central section of the Western desert and, because of a warp of the older surface, form a deep embayment west of the Nile to rather below the latitude of Aswan and form the northern part of the Eastern desert between the Nile and the Red Sea hills (Itbay) and part of the central plateau of Sinai. The whole represents a striking plateau of massive limestone cut into by the Nile valley for over 400 mi. from near Isna to Cairo. This limestone has provided most of the building stone used in the valley and delta. The Oligocene period was one of recession of the sea and the limited deposits are of estuarine and fluviatile origin, consisting mainly of sandstones and gravels devoid of fossils but including silicified wood. In the northern part of Al Fayyum depression they contain in addition the remains of large land animals including the *Arsinoitherium*, precursor of the horned Ungulata and the *Moeritherium* and *Palaeomastodon* ancestors of the elephant. These deposits are exposed in a narrow band from Bahriyah oasis to the north of Al Fayyum and in the Eastern desert between Cairo and Suez. The crustal movements of the Alpine mountain-building period began near the close of the Oligocene with the formation of the Red sea basin and uplift of its bordering mountains. A gentle subsidence followed marking the beginning of the Miocene period whereby marine limestones were deposited over the Cyrenaic plateau between the Mediterranean and the Siwah-Qattarah-Magharah depression, between Cairo and Suez, and in western Sinai. A branch of the Miocene Mediterranean sea invaded the Red sea basin and great thicknesses of gypsum and dolomitic limestone line its shores. These were much disturbed by an intense phase of the great Alpine crustal movements toward the end of the period, producing uplift and a complexity of folding and faulting and initiating the cutting of the great trench of the Nile valley in a broad syncline parallel with the Red Sea mountains anticline in the east. This was completed during the Pliocene.

The incursion of the Pliocene sea was more local, affecting the Nile syncline (Nile delta, the Wadi An Natrun to the west and part of the Nile trench) and the Red sea trough. In the Wadi An Natrun the deposits of sands and gypseous clays contain the remains of hippopotamus, elephant, giraffe and crocodile. Along the Red sea and in the Nile valley from Bani Suwayf southward they consist of fluviatile conglomerates, gravel and sands which are up to 600 ft. above present sea level. The deposits of the Pleistocene and more recent periods include the oölite limestones of the coastal strip west of Alexandria; the gravels, sands and Nile mud of the delta and Nile valley; and the drifted sands of the Western desert and of the northern part of Sinai formed by aeolian action upon disintegrating sandstone. In Pleistocene times, contemporaneous with the Ice Age in Europe, Egypt experienced greater rainfall and the torrential river flow accounts for sands, gravels and huge pebbles forming the basement of the present cultivated alluvium of the delta. This alluvium is very recent (probably being deposited during the last 16,000 years) and is derived mainly from the weathering and erosion of the volcanic rocks of the Ethiopian plateau.

2. Physiography.—Egypt may be divided into four major physical regions: the Nile valley and delta; the Western and Southern desert; the Eastern desert; and Sinai. The last three regions are arid lands; settled life is possible only in the valley of the Nile and its delta.

The Nile Valley.—This separates two very different types of country: to the west the desert is plateaulike in form, with drainage into closed depressions, while to the east is a complex of mountains and plateaus, intensively dissected, with drainage to both the Nile and the Red sea. The 750-mi. length of the Nile valley from the Sudanese frontier to Cairo is a relatively narrow groove, flat bottomed and sinuous in outline. In some parts the valley is as much as 14 mi. wide, in others it narrows until the enclosing steep bluffs are washed by the waters of the river. From the part where it enters Egypt to Aswan, a distance of about 200 mi., the Nile flows in a narrow gorgelike valley incised into the Nubian Sandstone. In this section the river is narrow and there is little or no flood plain. Granites of the Archean basement are exposed near Kalabishah, 35 mi. S. of Aswan, where the river flows through a gorge less than 250 yd. wide and for a distance of about 20 mi. toward Aswan. Just south of Aswan begins a seven-mile chain of rocky islands past which the river flows in a series of rapids constituting the First cataract. The construction there of the Aswan dam (1902), 5 mi. S. of Aswan, created a reservoir for nearly 100 mi. upstream inundating the few scattered patches of cultivation. The site of the proposed Aswan High dam is four miles upstream. When completed, this gigantic reservoir will flood the Nile valley for up to 400 mi. to the south, 150 mi. being in the Sudan. (*See* ASWAN.) The gradient of the river on this stretch is 1 in 11,000, slightly steeper than that of the remaining 745 mi. to the sea (1 in 13,200). Below Aswan the valley broadens and irrigable flood plain appears on either side. At Kawm Umbu, 25 mi. N. of Aswan, where two wadies debouch from the Red Sea hills, the valley widens to form an alluvial plain 10 mi. wide.

Near Isna a notable change of geology and physiography occurs, the Nubian Sandstone gives way to limestone formations and the valley becomes a broad flat-bottomed trough averaging 12 mi. wide between the white cliffs of the limestone plateau, which in places between Isna and Qina rise to 1,800 ft. From there the river winds in great meanders, and mud islands become a common feature of its channel. From Asyut northward the enclosing walls of the valley diminish in height, particularly on the west side, and the river shows a marked tendency to hug the eastern side so that most cultivated land lies west of the river. Northwest of Bani Suwayf there is a channel off the valley through which the Bahr Yusuf carries water to Al Fayyum depression. This depression in the limestone plateau falls from about the same level as the Nile to Lake Qarun, 148 ft. below sea level, and covers an area of 770 sq.mi. From just above Bani Suwayf to Cairo the valley narrows to about five miles with the eastern limestone escarpment prominent. (*See* also NILE.)

Nile Delta.—At Cairo the valley opens out, forming a triangular alluvial area 155 mi. from east to west and 100 mi. from south to north. The land surface slopes gently to the sea, falling about 52 ft. from Cairo, a gradient of 1 in 10,000, to all appearances a remarkably level plain. The delta replaces a gulf of the Pleistocene Mediterranean and is composed of silt brought initially by the Blue Nile and Atbara from the Ethiopian highlands. The thickness of alluvium varies with the configuration of the former sea floor but averages 65 ft. The beds of gravel and sand underlying the Nile mud occasionally rise to form sterile gravelly mounds especially in the eastern part of the delta. The Nile alluvium when freshly deposited is very soft, plastic and sticky, but on losing moisture it hardens into a tough earth; the roads in the delta are mainly formed of silt from the canals. Its main components are silica and insoluble matter (60%), ferric oxide and alumina (22%). In the north along the coast is a considerable area of salt marsh and a series of shallow, brackish lagoons—Maryut, Idku, Burullus and Manzilah. The three latter are separated from the sea only by low sand belts. Lake Maryut is separated from the sea by limestone ridges and is kept at about nine feet below sea level by pumping. This belt of *Barari* or barren land is so near sea level that drainage by free flow is impracticable and reclamation to agricultural land is slowly being achieved by washing and pumping operations. In the 1st century A.D. Strabo reported the Nile as having seven delta distributaries and a 12th-century map shows six. These have now become controlled and consolidated into two, reaching the Mediterranean at Rosetta (Rashid) and Damietta (Dumyat) known as the Rashid and Dumyat branches. The delta coast line consists of four smooth crescentic curves created by the gentle west to east longshore drift associated with the outgrowing mouths of the two Nile branches and the termination of coastal limestone ridges in the west at Abu Qir. The sea off the delta is shallow and within ten miles of the coast and is never below ten fathoms deep.

Western and Southern Deserts.—This region covers an area of 260,000 sq.mi., or nearly three-quarters of the whole country. Together the deserts form the northeastern corner of the great Libyan desert (As Sahra al Libiyah) and extend from the Mediterranean to the Sudan border and from the Nile valley west to the borders of Libya. The desert is a massive plateau of great thick-

nesses of sedimentary rock attaining its greatest altitude in the southwest, where the Jilf Al Kabir plateau rises above 3,000 ft. and an outcrop of the Archean basement rocks forms true mountains, reaching more than 7,000 ft. in the Jabal Al Uwaynat in the extreme southwest. Apart from a very narrow coastal strip of gently undulating plain, where some rain falls in winter, it is one of the most arid regions of the world. From the Jilf Al Kabir the sandstone tableland slopes gently to the north to depressions in which lie the Kharijah (Kharga) and Dakhilah oases. The northern boundary of these depressions is the much enlarged escarpment marking the southern edge of the great plateau of resistant Eocene limestone, rising in places 1,200 ft. above the depressions. In this limestone plateau are the hollows containing the oases of Farafirah and Bahriyah. The ground continues to fall toward the Mediterranean and to the depressions of the Siwah oases and Qattarah, the latter being the largest, containing salt lakes and marshes and descending to 435 ft. below sea level. These depressions, except Qattarah, are habitable because of artesian supplies of water and north of them extends the escarpment of the Miocene limestone plateau, 600 ft. high, which almost reaches the Mediterranean. The surface of the desert consists of stony or sandy plains, of vast areas of loose sandstones and of rocky limestone plateaus. One notable feature is the absence of drainage lines and another, also related to the great aridity, is the occurrence of longitudinal belts of sand dunes, especially the system of parallel dunes 200 ft. or more in height constituting the Great Sand sea extending from Siwah oases to the Jilf Al Kabir.

Eastern Desert.—This area, sometimes called the Arabian desert (As Sahra'ash Sharqiyah), covers about 86,000 sq.mi. and extends from the Nile valley eastward to the Gulf of Suez and Red sea. It consists essentially of an easterly chain of high rugged igneous mountains paralleling the coast, flanked in the northwest and west by considerably dissected limestone and sandstone plateaus. The igneous Red Sea hills occupy nearly a third of the Eastern desert; they are a prolongation of the Ethiopian highlands and extend from the Sudanese frontier to about the limestone Jabal Jalalah al Bahriyah 550 mi. to the north. There the crystalline rocks of the Archean basement have been elevated and contorted, creating a wilderness of sharp-peaked ranges. Many peaks rise more than 5,000 ft. above sea level, the highest being Jabal Sha'ib al Banat (7,150 ft.) and Jabal Hamatah (6,480 ft.). The drainage from these mountains in more pluvial periods is responsible for the complexity of wadies dissecting the surrounding sedimentary rocks, contrasting with the easily traversed plateau land west of the Nile. Although the climate is arid, a little rain falls in the mountains in most years and water is obtainable from springs and rock pools. Thus many wadies contain sparse vegetation and wells.

Sinai.—This triangular peninsula (Shibh Jazirat Sina'), with its base on the Mediterranean, has an area of 23,500 sq.mi. The core near its southern end is formed by a complex of high rugged mountains of old crystalline rocks, essentially a continuation of the Red Sea hills. The northern two-thirds of the peninsula consists of a severely dissected limestone plateau whose south-facing scarp abuts the northern flanks of these old high mountains. The Sinai mountains, a mass of sharp peaks, ridges and deep ravines, are considerably higher than any in Egypt proper. The main peaks are Mt. Catherine or Jabal Katrinah (8,651 ft.), Jabal Umm Shawmar (8,482 ft.) and Jabal Ath Thabt (7,997 ft.). The high central portion of the limestone plateau is called Jabal At Tih; its highest part, in the south (the Jabal al 'Ajmah), attains 5,335 ft. Whereas the eastern and western edges of the limestone are dissected by numerous deep rocky valleys draining to the gulfs of Aqaba and Suez, to the north the country is more open and drains to the Mediterranean by shallow wadies of which the principal is the Wadi Al 'Arish. Along the Mediterranean coast are broad tracts of sand dunes. Rainfall is scanty but greater than in the Eastern desert. Drainage water to the north is absorbed in the sand and supplies wells and palm groves in hollows between the dunes. (*See* also SINAI.)

3. Climate.—The greater part of Egypt lies within the zone of the great tropical deserts and its climate is characterized by aridity, a considerable temperature range (both seasonal and diurnal), a dry atmosphere and a maximum amount of sunshine throughout the year. The rhythm of the seasons is regular and the weather remarkably stable, broken only by disturbances associated with the passage of depressions along the Mediterranean in winter and from the interior of the Sahara in the summer. Summer and winter are the principal and longest seasons (May–September and November–March). The climate of the Mediterranean littoral and the country for a few miles inland differs markedly from the rest of Egypt since it experiences a small winter rainfall (four to eight inches), milder winters but slightly lower summer temperatures (Alexandria, January 56° F., August 78°), rather more humid air and rainless summers. Winter depressions along the Mediterranean bring to this limited area strong winds, cool weather and heavy rains. The delta shares the character of both the Mediterranean and Saharan climates, having a small winter rainfall (Alexandria, 8 in., diminishing southward to Cairo, 1.3 in.) and mild winter temperatures and dry, hot summers (Cairo, January 53°, July 81°). South of the delta within the Saharan climatic zone summer temperatures rise above 100° (Aswan mean daily maximum in June 107°), the sun burning from a cloudless sky. At night a fall of 25°–30° takes place.

From April to June sporadic depressions from the Sahara traverse Egypt toward the delta, bringing excessively dry southerly winds which, if strong, create dust and sand storms (*khamsin*) which may persist for three or four days. At Cairo they blow on about 11 days a year and relative humidity may fall to under 10%. Temperatures drop markedly during the winter, mean temperatures during January being 55°–60° F. Lower average temperatures are found only in the Red Sea hills and Sinai. The prevailing winds are northerly for most of the year, particularly during the summer, but become southerly and westerly during the winter months of Mediterranean depressions. Except on the highland and the Mediterranean littoral rainfall is meagre and very irregular. Years may pass without a drop and then a sudden storm may give one or two inches in 24 hours. In exceptional cases great damage may be done; *e.g.*, in Nov. 1925 a deluge of mud and water from the Muqattam hills devastated the Al Abbasiyah quarter of Cairo. (A. B. M.)

4. Vegetation.—The flora of Egypt is limited by the vast expanses of arid desert and by the intensive cultivation of the Nile valley and of the delta. In the deserts it is sapless and desiccated and consists of dwarf mimosas, stunted tamarisks of several varieties, coarse alfa grass and small herbs; in the deep wadies of the Eastern desert and of Sinai the vegetation is more varied and wherever subterranean water is found the date palm flourishes.

Near the Nile the vegetation is almost entirely due to irrigation and to cultivation; only a few wild plants occur, among which various vetches (*Astragalus*), the sesban shrub (*Sesbania aegyptica;* in Arabic *seisaban*), fennel (*Foeniculum piperitum*) and endive (*Cichorium endivia*) have been noted; and among wild flowers the most common are yellow daisies, poppies, irises, asphodels and ranunculuses. Goosefoot (*Chenopodium*) flourishes on the banks and islands of the coastal lakes and lagoons (Lake Manzilah, etc.) and cacti occur as a subspontaneous growth near Alexandria. The date palm (*Phoenix dactylifera*) is indigenous and about 30 varieties are grown. The doum palm (*Hyphaene thebaica*) is found from about the latitude of Luxor and to the south. The neb-neb (*Acacia nilotica*), the thorn tree of antiquity, is, next to the date palm, the tree most frequently seen by the wayside and in the villages. Other common trees are sweet acacia (*Acacia farnesiana*), with blossoms of delicate perfume, the sycamore, Christ's thorn tree (*Paliurus spina-christi*) and the carob or St. John's bread (*Ceratonia siliqua*). The famous biblus or papyrus (*q.v.*) no longer exists in Egypt, except in the botanical gardens in Cairo, but a number of types of reed are found. The lotus (*Nymphaea lotus*), greatly prized by the ancients for its flowers, is still found in the delta though never in the Nile itself. Many types of foreign trees have been introduced, partly for decorative purposes and to give shade and partly in an attempt to provide timber; of these the principal is the lebbek (*Albizzia lebbek*), a thick-foliaged mimosalike tree which thrives well, others are the poinsettia

(*Euphorbia pulcherrima*), the jacaranda, the casuarina, the eucalyptus, the weeping willow, the elm and the cypress. Many flowers common in Europe are cultivated, and roses, carnations, oleanders and geraniums have been grown from a very early period.

5. Animal Life.—The chief quadrupeds of Egypt are all domestic animals, and of these only the ass is certainly indigenous. The horse is pictured in ancient monuments, but the camel is not (although it is mentioned in several papyri) and the date of its introduction is doubtful. The wild animals are few. The principal are the dorcas gazelle, which inhabits the deserts both east and west of the Nile, the Nubian ibex, found in the Sinai peninsula and in the mountains of the Eastern desert, the hyena, the jackal and the desert or fennec fox (*Fennecus zerda*), which subsists partly on vegetable food and lives in extensive burrows in the desert. The desert hare is abundant in Al Fayyum and elsewhere; the jungle cat and the caracal lynx frequent marshy regions of the delta. The ichneumon (Egyptian mongoose or Pharaoh's rat) is common in the cultivation. The cony is found on the stony slopes of the mountains of Sinai. In the Nile valley and the neighbouring hills are bats of several species. There are two varieties of jerboa, the lesser and the greater Egyptian (*Jaculus jaculus* and *J. orientalis*). Mice and rats are common, including the fat sand rat (*Psammomys obesus*), which is found in the sand hills around Alexandria.

There are many birds both resident and visiting. The commonest birds of prey are vultures, kites, hawks and eagles. The barn owl is common; the so-called Egyptian eagle owl is less so. Game birds abound, and species include the sand grouse (*Pterocles exustus* and others); the red partridge; a red-legged partridge found in the hills about Aswan, in the Eastern desert and in Sinai; the lesser bustard; the quail (on spring and autumn migrations); the snipe; and the plover. There is a great variety of aquatic birds; ducks, geese and flamingos are found in the lakes of lower Egypt and cranes, herons, spoonbills and kingfishers are common. Myriads of pigeons are seen in every village. The hoopoe is found everywhere in the cultivation. Swifts, martins, including the crag martin, larks, finches and the bulbul (*Pycnonotus arsinoe*) inhabit either the desert or the sown areas.

Egypt has several venomous snakes including the horned viper, the echis and the Egyptian cobra. Lizards are numerous and include the desert monitor, which attains a length of four to five feet. The crocodile, hippopotamus, lion, giraffe and ostrich, depicted on ancient Egyptian friezes, have vanished and probably retreated south to the tropics many centuries ago. Butterflies are rare, but there are numerous moths. Wasps and hornets abound; grasshoppers and the praying mantis are common. Vermin of all kinds are found: cockroaches, fleas, bugs, lice, scorpions, tarantulas and centipedes.

Nile fish are of many varieties and include the Nile perch, which may scale as much as 160 lb., and buolti (*Tilapia nilotica*) and carp. Lake Qarun, which has become increasingly saline, has been stocked with marine species. (K. C. B.)

II. GEOGRAPHICAL REGIONS

From *Geology and Structure* and *Physiography* above it will be clear that Egypt may be considered as comprising five regions, relatively homogeneous in themselves but each differing from the others. These regions are the Western and Southern deserts, the Eastern desert, the Nile valley, the delta and the Sinai peninsula. It is only in the Nile valley and delta that settlement and agriculture on a large scale are possible. Viewed from the air the dividing line between the desert and the sown area is impressively abrupt, and the vivid green patchwork of cultivated fields varied by the darker date-palm groves is in refreshing contrast to the drab encircling desert.

1. Western and Southern Deserts.—A population of about 100,000 persons is to be found in the oases (Kharijah, Bahriyah, Dakhilah, Farafirah and Siwah) and in a string of small villages along the Mediterranean coast; it is only in these places that reliable water supplies are found. The rest of the country, with its stony or sandy plains, the considerable areas of dune sand, the rocky limestone outcrops, is inhospitable and repellent, even to most of the few thousand nomads who prefer to move with flocks and herds in the Eastern desert. The oases communities are self sufficient to a high degree, usually selling dates (Siwah also sells olive oil) in exchange for manufactured goods. Kharijah and Siwah are linked with habitable Egypt by a railway and motor road respectively, the other oases are more remote.

2. Eastern Desert.—This is a region of broken upland averaging 2,000 ft. in height but with some peaks more than 7,000 ft. high. The country is much dissected by narrow and deep wadies and some of the more sheltered carry a sparse vegetation sufficient to nourish the camels of nomadic Bedouin. Water resources, although sparse, are more plentiful than in the Western desert. The distortion to the strata precludes the presence of extensive artesian layers, but the greater altitude causes a small rainfall. Thus water is to be found in sheltered hollows, or beneath apparently dry stream beds, and there are occasional springs. Cultivation is practically impossible and the estimated 15,000 inhabitants are to be found mainly in villages along the Red sea coast or in mining camps. The highland core, where there are no permanent settlements, is the domain of nomadic shepherds.

The crystalline and metamorphosed rocks contain a variety of minerals, but detailed surveys are still needed. Oil is obtained from near the Red sea coast at Al Ghurdaqah and Ras Gharib and phosphates are worked at Al Qusayr and Bi'r Safajah.

3. Sinai.—The peninsula separated from Egypt proper by the isthmus of Suez is an irregular plateau rising to the south where heights of over 8,000 ft. occur. Erosion of more pluvial periods in the past has caused deep dissection of the whole upland. Rainfall is six inches per annum along the northern coast but is only two or three inches in the south. This factor, in conjunction with the great gorges, ravines and general inaccessibility of the mountainous area, explains why the sparse population is to be found almost entirely in the northern half, for the northerly debouching wadies are broad and pockets of cultivation surround wells, such as at Al 'Arish, Al Awja and An Nakhl. Elsewhere nomadism is the principal way of life.

4. Nile Valley and Delta.—Although comprising only 3.5% of Egypt, this area contains most of Egypt's population. From earliest times reliable flooding and automatic renewal of fertility has contrasted with the enclosing sterile desert and has made habitation possible. The 550-mi. narrow trench of the river from Aswan to Cairo (upper Egypt) exhibits some of the most densely settled agricultural land in the world. The gray mud-brick huts of villages huddle together every one or two miles, the surrounding land a mosaic of carefully tended small plots. Shadoofs and sakiehs lift the precious water onto the land from neighbouring irrigation channels. While there are still almost 1,000,000 ac. under the old pharaonic system of basin irrigation, the remaining 5,000,000 ac. of Egyptian farmland are served by perennial irrigation, which during the low Nile season makes use of water stored in dams to the south. The inherent fertility of the soil, no longer being enriched annually by Nile mud, is now maintained with artificial fertilizer, and very high yields of maize (corn), millet, sugar and cotton are obtained (see *Irrigation* and *Methods of Agriculture* below). The delta (lower Egypt) which fans out from Cairo to the Mediterranean is said to resemble the back of a leaf, with the river and canals obtruding like veins above the adjoining lowland. Nearly all the major towns and cities, the bulk of the country's industry and two-thirds of the total population are in this broad expanse of plain. Everywhere the pressure of population upon the land is alarming; poverty is extreme despite some of the highest yields in the world of such crops as wheat, rice, maize and cotton. The northern part of the delta (over 1,000,000 ac.) consists of reedy lagoons and salt marsh, still, in the 1960s, virtually useless. (A. B. M.)

III. THE PEOPLE

About 63% of the Egyptian population is concentrated in the delta, most of the remainder in the many villages and small straggling towns of mud brick along the Nile valley. The great majority of the people depend upon agriculture for their livelihood. The countryman, or fellah, descended from the successive Arab settle-

ments after the Muslim invasion as well as from the mixed indigenous pre-Islamic population, is dark of complexion, sturdy of physique, slightly taller than average and inured to the tedious work of constantly lifting water for irrigation by primitive means. These fellahin comprise the core of Egypt's population. They are usually cheerful, hospitable and resigned to toil, but are stubborn at times and given to jealousies and local feuds. Their life is notoriously hard and impoverished, unrelieved by leisure or luxury and slow to change. They work clad only in loincloths, and at other times wear long dark-blue or white cotton robes (*gallabiyehs*) and close-fitting skullcaps over shaven heads with turbans, usually white. The red tarboosh, of Turkish origin, was fast disappearing in the mid-20th century. Women wear black and outside the towns a few still veil their persons on the approach of a stranger. The speech is a vernacular Arabic with many dialect variations. The government is attempting, with some success, to establish a standard colloquial Arabic as the lingua franca of radio transmission. The religion of the vast majority is Islam and the way of life conforms to the general Islamic pattern. In certain customs and superstitions survivals of ancient beliefs and practices may be recognized, such as the marked reverence paid to the tombs of holy men (*sheikhs*). The general level of subsistence is low and the diet monotonous, chiefly coarse maize bread and beans; meat is a rare luxury.

Around the wells of the Eastern desert and along the edge of the cultivation a few Bedouin (*q.v.*) still pursue a nomadic way of life. In Aswan region in the south the people are more closely related to North African tribes; their speech is Nubian and their villages are more compact and characterized by domes of mud brick. The Nubians or Barabra (*q.v.*) are renowned as boatmen and provide most of the domestic servants.

About one-twelfth of the population are monophysite Christians known as Copts, a corruption of the Greek term *Aigyptoi* applied by the Muslim invaders to the indigenous population (*see* COPTIC CHURCH). The Copts are most numerous in middle Egypt from Mallawi to a little south of Akhmim. Those of the country areas are indistinguishable from their Muslim neighbours. In the towns they are found as clerks and accountants and have a reputation for manual skills; they rarely occupy the highest administrative posts. The Coptic language (*q.v.*), the last stage of ancient Egyptian, is no longer spoken but is still used in the Bohairic dialect for liturgical purposes.

Cairo and Alexandria now convey an occidental rather than oriental impression with many features reminiscent of Mediterranean cities of southern Europe. They are cosmopolitan in character and have minority groups maintaining their own languages, religions and customs. Many in these groups were born in Egypt and know no other home. They include Greeks, Italians and Maltese and to a lesser extent Armenians, Levantines and Jews. They have been important in commerce and industry and active in enriching Egyptian intellectual life, but after the Suez conflict in 1956 there was a marked exodus of all foreign communities.

For languages *see* ARABIC LANGUAGE; AFRICAN LANGUAGES.

(A. F. SH.)

IV. ARCHAEOLOGY

Ancient Egypt has always had a peculiar fascination for the western world. The Greeks felt that their own civilization was new and inexperienced in comparison with the age-old traditions and skills of this ancient land where the past still lived in the present. Visitors to Egypt marveled at the great cities with their huge temples. The *Periegesis* ("Tour Round the World") written by Hecataeus of Miletus (late 6th century B.C.) is unhappily lost. Herodotus, who wrote in the following century, gives a vivid and lively description of the country as he saw it during a visit of several months, together with the best account he could piece together of the history and beliefs of the people, based largely on the gossip of uneducated priests and dragomans. Much of what Herodotus retails is clearly popular legend, and he had no notion of the span of Egyptian history or the sequence of events long past. The priest Manetho had perhaps better sources when he wrote his *Aegyptiaca* about 240 B.C.; the surviving extracts from this work, ill-copied by later classical writers who quoted it for their own polemic ends, show how valuable a work has been lost. As it is, Manetho's list of the 30 dynasties of kings of pharaonic Egypt, in spite of the garbled forms of the names and miscopying of figures, has triumphantly survived the test of archaeology and is still retained by Egyptologists as the basis of the reconstructed history of Egypt.

Many Greeks visited Egypt, as tourists or as merchants, but it was the Romans who began collecting Egyptian antiquities, and baboons, sphinxes and statues of the Pharaohs adorned the palaces of the Roman emperors; great obelisks were transported with immense labour from Karnak (Thebes) and Heliopolis and re-erected in Rome and Constantinople. Tacitus relates that when Germanicus came to Egypt in A.D. 19 to see the sights he obtained the services of an old priest who could read the hieroglyphs and translate them into Latin.

Among the works of later classical writers, Strabo's description of Egypt in the 17th book of his universal geography is of considerable value, especially since he describes in some detail the otherwise little known topography of the delta. Geographical descriptions are also found in the *Naturalis Historia* of Pliny the Elder and in Ptolemy's geographical treatise. Both Greeks and Romans speculated on the mythology and religious beliefs of the ancient Egyptians. Plutarch's treatise *De Iside et Osiride* tells the Osiris myth more or less as it is in the native sources but adds fantastic speculation on the hidden symbolism behind the outward form. Later classical writers such as Horapollon endowed the hieroglyphic signs themselves with fanciful allegorical significance.

The last hieroglyphic inscription that has survived was that engraved at Philae by Theodosius the Great in A.D. 394. With the spread of Christianity, Coptic replaced the ancient script and the religion, long obsolescent, died out. But the tradition of the "wisdom of Egypt" survived through the middle ages and gave rise during the Renaissance to renewed speculation. Travelers once more visited Egypt, and scholars set themselves to solve the riddle of the sphinx. John Greaves of Oxford in 1646 wrote a treatise on the pyramids, and a Jesuit scholar, Athanasius Kircher, in his book entitled *Lingua Aegyptiaca Restituta* (1643), attempted with more ingenuity than restraint to divulge the hidden meaning of the hieroglyphs. Travelers in the 18th century recorded and collected as they went; in 1741 the traveler Richard Pococke, together with William Stukeley and other prominent English antiquaries, formed the Egyptian society in London.

1. Beginnings of Modern Egyptology.—The visit of Napoleon Bonaparte to Egypt in 1798, the discovery of the trilingual Rosetta stone (*q.v.*) and the subsequent decipherment of the hieroglyphs by Thomas Young and J. F. Champollion (*q.v.*), begun in 1821, at length provided material for the serious study of Egypt's past, and scholars were able to write the first histories of ancient Egypt. J. G. Wilkinson and Samuel Birch were among the earliest of these. The remarkable recording work of Karl Richard Lepsius (*q.v.*) and his draftsmen for the Royal Prussian expedition to Egypt in 1842–45 presented the world with a large collection of careful drawings and copies of inscriptions upon which historians and philologists alike could base their work. G. R. Gliddon in the 1830s was the first to talk about the new discoveries in the United States. J. G. Wilkinson first popularized the subject in England with his *Manners and Customs of the Ancient Egyptians*, first published in 1837 and re-edited by Birch in 1878 with lively illustrations taken from the copies of Lepsius, Dominique Vivant Denon and Ippolito Rosellini and Wilkinson's own sketches made in Egypt.

Meanwhile museums were receiving gifts of antiquities from travelers in Egypt. Demand created supply, and the profession of the dealer in antiquities, intimately linked in Egypt with that more illicit craft of tomb robbing, was open to the enterprising. In this way the great collections of the Louvre, the British museum and the museums of Berlin, Turin, Florence and Bologna, to name the largest, were built up. From the hieratic papyri thus discovered and the hieroglyphic monuments brought back, the philologists were reaping a rich harvest. The names of C. W. Goodwin, Heinrich Brugsch, E. de Rougé and F. J. Chabas are all

honoured in this period of the mid-19th century.

When in 1858 Auguste Mariette became conservator of Egyptian monuments to the khedive, the Egyptians embarked on a new policy for the conservation of their antiquities. The Bulaq museum was founded in a suburb of Cairo, and Mariette's excavations, largely in Giza (Al Jizah) and Saqqarah, during the 1860s and 1870s, filled the museum's cases with sculpture and antiquities, largely of the Old Kingdom. In 1851 he discovered the huge coffins of the Apis bulls in the galleries of the Serapeum. After Mariette's death in 1881, the right to excavate in Egypt was granted by concession to museums and learned societies such as the Egypt Exploration Fund (later Society) of London. The law provided for the equitable division of finds, half to the excavating body and half to the Bulaq museum. As a result, the museums of the western world could now for the first time be enriched with scientifically recorded material, while the national collection grew under the direction of Gaston Maspero and later Jacques de Morgan till it outgrew its capacity. New premises were found, first, in 1891, at Giza; then, in 1902, a large building at Qasr an Nil in the centre of Cairo was erected for the purpose, the present Cairo Museum of Antiquities.

2. 20th-Century Techniques and Excavations.—Because of the revolutionary methods developed by W. M. (later Sir) Flinders Petrie and his pupils, the old method of uncovering monuments and preserving only such relics of the past as might be termed works of art fit for museum display gave way to more systematic excavation; the painstaking examination and recording of every object, however insignificant and however fragmentary, and the study of the layers of earth in which they were found and their relative position in that earth. In their context, he showed, these things helped the archaeologist to piece together the history of the site and its inhabitants. Petrie himself studied the stylistic development and technical evolution of stone vases, pottery and other objects of common use, and his typological series, applied to prehistory, became the foundation of a relative system of chronology known as sequence dating, which, with modifications, is still used by Egyptologists. He set a standard also in the prompt publication of his results in the summer following every winter excavation.

Through these improved techniques, and the recording and study of tombs and temples, far more is known about the daily life and personal possessions of the ancient Egyptians than about those of any other country, not excepting Greece and Italy. The damper climate of Europe has not preserved many of the actual objects that men used so long ago. In the dry, almost rainless atmosphere of Egypt, baked in the hot sand, even wood is preserved intact.

The garlands worn by mourners at a funeral and the bouquets they carried still lie, withered but uncrushed, where they fell in the corridors of Theban tombs. When in 1938 an intact tomb of the 2nd dynasty was opened at Saqqarah, the excavators found the funerary repast so well preserved that dishes of beef, fish, pigeon, stew, quail, fruit, cake and bread were easily recognized, and even the footprints of the men who had borne the coffin to its last resting place nearly 5,000 years before were plainly visible on the sandy floor.

The discovery of an untouched tomb is of rare occurrence, and

ARCHAEOLOGICAL SITES, ANCIENT CITIES AND PHYSICAL FEATURES OF EGYPT

the tombs of the Pharaohs themselves, packed with treasures as they must all have been, proved too tempting for even the ancient Egyptians. In spite of their awe of the dead, neither massive pyramids nor puzzle entrances nor even secret burial places were protection against tomb robbers, and almost every Pharaoh's tomb must have been violated not many years after his death. The discovery in 1922 of an almost intact tomb, which robbers had attempted but failed to penetrate, in the Valley of Tombs of the Kings in western Thebes, created unprecedented excitement. The tomb of Tutankhamen filled the headlines of newspapers, and public interest in Egyptology, long on the wane, revived. The emptying of this tomb took nearly ten years; for the extraction and preservation of the wonderful treasures it contained the discoverers, Lord Carnarvon and Howard Carter, called to their aid a number of technical experts, and the result was the dazzling display of furniture, arms, jewels and shrines that make the Cairo museum one of the wonders of the modern world.

During the first decades of the 20th century the excavation of Egyptian sites proceeded steadily. Expeditions were financed by the governments of the United States, France, Germany, Italy and Poland, by museums and learned societies and by private individuals. Petrie, Edouard Naville, F. L. Griffith, A. M. Blackman, P. E. Newberry and T. E. Peet are among British scholars who worked in Egypt; the work of the U.S. excavators George Reisner, at Giza and in Nubia, and H. E. Winlock at Thebes should be mentioned. Unfortunately, after the controversy that followed the discovery of Tutankhamen's tomb, the Egyptian government adopted a less liberal policy whereby the Cairo museum was entitled to everything found by an excavator but might give him *ex gratia* objects that it did not need. This led to a falling off of interest on the part of museums, which were not inclined to finance projects from which they could derive little benefit, and the work of Egyptology in Egypt tended to turn from excavation to recording. The excellent reproductions of Theban paintings published by the Metropolitan Museum of Art, New York city, and the splendid volumes of the Oriental institute of The University of Chicago, are outstanding examples of this trend. World War II and its aftermath called a halt to archaeology in Egypt, but it was hoped that renewed activity, in excavation and recording, could proceed under more favourable conditions, an activity in which Egyptian archaeologists themselves took part.

Nothing has been said of the scholars who furthered the study of Egyptology in the field of "armchair" research: of the philological work of Adolf Erman, who with H. Ranke produced, after years of compilation, the great Berlin dictionary of the ancient Egyptian language; of the researches of Kurt Sethe and his colleagues in Germany; of Sir Wallis Budge, F. L. Griffith and A. H. (Sir Alan) Gardiner in England; and of H. Gauthier and G. Lefebvre in France—to mention only a few. Egyptology is now an academic subject studied in the universities. It looks to the past, but it has a brilliant future. There is much more to be discovered, and every year new techniques make further knowledge possible; the method of radiocarbon dating is but one example of possible contributions which modern science may make to the knowledge of Egypt's past. *See* also ARCHAEOLOGY: *Egypt and Western Asia.*

V. HISTORY

A. ANCIENT PERIOD

1. Prehistoric Age.—Tradition, mythology and later custom make it possible to recover but little of the story of prehistoric Egypt, of the time before written records began. Remains of Paleolithic man have been found on the gravel terraces bordering the Nile valley and in the high desert beyond; the settlements of Neolithic man were in the low desert bordering the swamps of the Nile valley, and the excavation of ancient dwellings and cemeteries of this and subsequent ages of man's development show that over a period of perhaps 4,000 years the Nile valley swamps were gradually cleared and settled by a succession of peoples from Africa and perhaps from western Asia, and that these elements fused at length to produce the civilization of ancient Egypt as it is seen at the dawn of the historic period.

At first, there are simple village communities gradually learning the techniques of agriculture, weaving, the domestication of animals and the manufacture of stone and pottery vessels. Later, larger settlements are found whose inhabitants had learned to build with sun-dried bricks, to irrigate their land, to make glass and to work copper. The terms Badarian, Amratian and Gerzean have been given by archaeologists to the successive phases of this development. Each is characterized by the manufacture of distinctive types of pottery (see *Arts and Crafts* below) and other innovations in material culture perhaps due to contact with the outside world, but there is no break in continuity. During the Gerzean period, it is clear, there was contact with the civilization of Mesopotamia, which may indicate commerce with the Persian gulf ports or even an influx of craftsmen.

Attempts have been made to reconstruct the history of prehistoric Egypt; the German scholar Kurt Sethe, studying the later distribution of religious cults and other dynastic material, supposed that at a remote period the delta had been dominated by followers of Horus while the upper Egyptians had their national shrine at Ombos, the centre of worship of the god Setekh (Seth), and that for a time the land was united under one rule at Heliopolis. Other scholars have disputed this, and explain the traditional division of upper and lower Egypt, fundamental to Egyptian history and persisting in the concept of Egypt as "the two lands," as the reflection of a contrast between the south, essentially African in character, and the delta, profoundly influenced by its neighbours in western Asia and Libya.

A few large palettes and ceremonial maceheads from Nekhen (Hierakonpolis), of the period immediately preceding the 1st dynasty (*i.e.*, the end of the 4th millennium), are carved with semi-symbolic scenes which seem to be historical records and may represent the victory of a king of upper Egypt over northerners of the delta region. From later mythology and from ancient administrative titles and elements of the royal titulary, it has been suggested that at this time the capital of upper Egypt was Nekheb, now represented by the ruins of Al Kab, with the royal residence across the river at Nekhen, while that of lower Egypt was at Buto in the delta, with the royal residence in the quarter called Pe. Nekhbet, goddess of Al Kab, represented the upper or southern Kingdom, while Udjo (Buto) presided over the delta. The two goddesses, in the form of vulture and cobra, were united in the royal diadem. A few names of kings of upper and of lower Egypt are preserved in the first line of the Palermo stone, but no annals are attached to them.

2. Archaic Period.—Names of eight kings of the 1st dynasty are known from their tombs or cenotaphs at Abydos and from other monuments. The traditional founder of the dynasty was Mena, the Greek Menes (*q.v.*), who was thought to have united the two kingdoms and founded Memphis as his capital. This king is probably to be identified with Narmer, who is shown on a Hierakonpolis slate palette wearing the two crowns of north and south; he is equated by some with Ahai, "the fighter," whose name appears in a 1st-dynasty mastaba tomb at Saqqarah and also among the series of tombs or cenotaphs at Abydos which contained royal steles. Manetho calls the dynasty "Thinite" (*i.e.*, from This, a city near Abydos); but many high officials, and possibly the kings themselves, were buried at Saqqarah, the necropolis of Memphis. A mastaba dated to the reign of Wadji, the fourth king, discovered in 1953, is particularly imposing with its decoration of bulls' heads. These mastabas, of brick with paneled façades, contained many storerooms packed with furniture and equipment, tools and hunting weapons, food and drink for the use of the deceased in the next world. As a means of dating the contents, jar labels of ivory and ebony give the principal events of the year in simple hieroglyphs and pictograms. The annals of the Palermo stone also begin with the 1st dynasty. Most of the events recorded are of a religious character, but the "smiting" of Asians is mentioned more than once.

The 2nd dynasty of Manetho, of which little is known, appears to be differentiated from the 1st on the Palermo stone; it also was Thinite and tombs of several of its kings were found at Abydos. At least two of them had tombs at Saqqarah. Peribsen

took the epithet "conqueror of foreign lands" and is associated with the cult of Set or Seth. Two kings of the 3rd dynasty, Zoser and Sanakht, had large brick mastaba tombs at Bet Khallaf, near Abydos.

The architect and learned scribe Imhotep designed for Zoser a second and mightier monument at Memphis, the great Step pyramid of Saqqarah surrounded by its complex of wonderful white limestone temples for the funerary cult. Imhotep later acquired a great reputation as a doctor and magician and was deified in the Saite period. It may be that the considerable technical advances made during the 3rd dynasty were due to his genius. The unfinished pyramid-temple complex of Zoser's successor, Sekhemkhet, though similar in design, shows increasing boldness in the use of masonry blocks.

3. Old Kingdom.—*4th Dynasty.*—Monuments and written records are henceforth more numerous. The fragments of the Palermo annals preserve a full record for six years of the reign of Sneferu at the beginning of the 4th dynasty. The events include successful raids on Nubians and Libyans, and the arrival of 40 vessels laden with timber from the forests of Lebanon. Sneferu also set up a tablet in the Wadi Magharah in Sinai where he mined copper. No fewer than three pyramids are attributed to him, one at Maydum north of Al Fayyum and two at Dahshur. Two of these pyramids are experimental in shape and appear to mark the transition between the step pyramid form and the true pyramidal shape of Sneferu's third and all later Egyptian pyramids. Pyramids did not cease to be built in Egypt till the New Kingdom, but the Old Kingdom, from the 4th to the 6th dynasties, is the age of the great pyramid builders, when the royal tomb, a pointed pile of huge stone blocks smoothly faced, was the chief monument of each successive king.

Sneferu's successors Khufu, Khafre (*qq.v.*) and Menkure (the Cheops, Chephren and Mycerinus of Herodotus) were the builders of the three great pyramids of Giza. Each was surrounded by smaller pyramids and mastabas for members of the royal family, and a funerary temple connected by a long causeway with a massive valley temple at the edge of the Nile flood (*see* EGYPTIAN ARCHITECTURE). The huge monolithic pillars of Khafre's valley temple, still preserved, were hewn from Aswan granite; between them stood 23 statues of the king, some carved from diorite brought from a desert quarry 50 mi. N.W. of Abu Simbel in Nubia. The Sphinx, carved in an outcrop of rock, stands sentinel over Khafre's causeway, a symbol of the king's might and majesty. The pyramid temples of Menkure, designed in limestone and red granite, were left incomplete, and subsequent kings of this dynasty left few monuments. In the royal mastabas there are hints of family dissension; Khufu's son Rededef, abandoning Giza, built his pyramid at Abu Roash, and Shepseskaf, the last of the dynasty, reverted to the old mastaba form for his tomb at Saqqarah. Nevertheless the 4th dynasty was regarded as the classic period of Egypt's greatness. Arts and crafts flourished under the patronage of wealthy Pharaohs, and the resources of the crown must have been almost limitless.

King Userkaf, the first of the 5th dynasty, had been a priest of Re (Ra), the sun-god, at Heliopolis; he and his eight successors worshiped Re, rather than Horus, as their ancestor, and the title "son of the sun" began to be written by them before the cartouche containing the personal name, while another "solar" cartouche, containing a name compounded with Re, followed the title "king of upper and lower Egypt." The marvelous tales of the Westcar papyrus, dating from the Middle Kingdom, narrate how three of the kings were born of a priestess of Re. Most of them built, near their pyramids, magnificent temples with obelisks dedicated to Re; the best preserved of the pyramid temples, that of Sahure, excavated by the German Orient-Gesellschaft, in its architecture and sculptured scenes has revealed an astonishingly complete artistic development as well as warlike enterprise by sea and land. The latest pyramid belonging to the 5th dynasty, that of Unas at Saqqarah, is inscribed with long ritual and magical texts. Exquisitely sculptured mastaba tombs of this time are very numerous at Memphis and are found also in upper Egypt. The annals of the Palermo stone were engraved and added to during this dynasty; the chief events recorded for the time are gifts and endowments for the temples. Evidently priestly influence was strong at the court. Expeditions to Sinai and Punt are also chronicled.

6th Dynasty.—The 6th dynasty if not more vigorous was more articulate; inscribed tombs are spread throughout the country. A tombstone from Abydos celebrates the activity of a certain Weni, during the reigns of Pepi I and his successor, in organizing expeditions to the Sinai peninsula and southern Palestine and in transporting stone from Elephantine and other quarries. Harkhuf, prince of Elephantine and an enterprising leader of caravans to the southern countries both in Nubia and the Libyan oases, flourished under Merenre and Pepi II. On one occasion he brought home a pygmy dancer from the Sudan; this drew from the youthful Pepi II an enthusiastic letter which was engraved in full upon the façade of Harkhuf's tomb.

The reign of the last-named king, begun early, lasted 94 years; his senility probably hastened the ruin of the dynasty. The royal authority had been gradually weakening; the king-god, it was beginning to be understood, suffered mortal limitations. Impoverished by vast building schemes, squandering their estates on mortuary grants and on temple endowments, the Pharaohs of the 5th and 6th dynasties found it increasingly hard to control a powerful provincial nobility now no longer dependent on royal favour but claiming office by hereditary right and residing not at court but in miniature state in their own nomes. After Pepi II an ominous gap in the monuments, pointing to civil war, marks the end of the Old Kingdom.

4. First Intermediate Period.—The 7th and 8th dynasties are said by Manetho to have been Memphite, but of them scarcely any record survives beyond some names of kings in the lists. Literary texts record a complete upset of social order and the intrusion of Asian nomads into the delta. The duration of this wretched state of affairs is unknown. The 9th and 10th dynasties of Heracleopolis Magna in middle Egypt may have spread their rule by conquest over upper Egypt and then overthrown the Memphite dynasty. It would seem that after they in turn were overthrown by their Theban rivals their monuments at Heracleopolis were systematically destroyed. The unsettled state of the country at this time is reflected in the paucity of monuments and the crude workmanship of smaller objects.

5. Middle Kingdom.—*11th Dynasty.*—At length, princes of Thebes asserted their independence and founded the 11th dynasty, which pushed its frontiers northward until finally it occupied the whole country. Its kings were named Mentuhotep and Intef, and were buried at Thebes. Their succession is imperfectly known. Nebhepetre Mentuhotep probably established his rule over all Egypt in about 2040 B.C., and attempted to reconquer Nubia; his funerary temple, excavated by the Egypt Exploration fund at Dayr al Bahri, must have been a magnificent monument. Two Mentuhoteps, S'ankhkare and Nebtawire, sent large expeditions through the Wadi Hammamat to open quarries and to trade with Punt, the land of myrrh beyond the Bab el Mandeb.

12th Dynasty.—Monuments of the Theban 12th dynasty are abundant and often of splendid design and workmanship. The Turin papyrus and many dated inscriptions fix the succession and length of reign of its eight kings very accurately. The troubled times that the kingdom had passed through taught these long-lived Pharaohs the precaution of associating a competent successor on the throne. The feudal chiefs were inclined to strengthen themselves at the expense of their neighbours; a firm hand was required to hold them in check and distribute the honours earned by faithful service. The tombs of the most favoured and wealthy princes are magnificent, particularly those of certain families in middle Egypt at Bani Hasan, Al Bersha, Asyut and Dayr Rifah, and it is probable that each had a court and organization within his district or nome like that of the royal palace in miniature. Eventually, drastic reforms, perhaps carried out by Sesostris III, appear to have centralized all authority very completely. The

names in the dynasty are Amenemhet (Ammenemes) and Sesostris (Senusret, formerly read Usertsen). Amenemhet I, the first king, who may have been vizier under the last Mentuhotep, had to fight for his throne and then reorganize the country, removing his capital or residence from Thebes to a central situation at Ith-tawi, about 25 mi. S. of Memphis. His monuments are widespread in Egypt, the quarries and mines in the desert as far as Sinai bear witness to his great activity, and it is known that he made an expedition against the Nubians. The "Instructions of Amenemhet to his son Senusret," whether really his own or a later composition, refer to his care for his subjects and to the ingratitude with which he was rewarded, an attempt on his life having been made by the trusted servants in his own palace. To secure the delta against Asian invaders, he built defensive walls on the eastern frontier.

The pyramids of both Amenemhet I and his son Sesostris I are at Lisht, near the new capital. Amenemhet II and Sesostris III were buried at Dahshur, and Sesostris II at Illahun; the workmen's town especially built for those engaged upon the construction of this pyramid was excavated by Sir Flinders Petrie in 1889 and provided the earliest example of Egyptian town planning. Amenemhet III during his long reign of 46 years built two pyramids, one at Dahshur and one at Hawwarah near the Fayyum. The development of Al Fayyum as a market garden was one of the preoccupations of the Pharaohs of this dynasty. Their plans included using Lake Moeris as a reservoir and regulating the distribution and outflow of Nile water. Several Pharaohs left monuments of considerable size in Al Fayyum, and the size and complexity of the mortuary temple of Amenemhet III, called by Greek writers "the Labyrinth," excited the wonder of visitors in classical times.

One of the great achievements of the dynasty was the subjugation of Nubia, which was now systematically conquered and brought under Egyptian control. A line of forts, or heavily fortified townships, controlled the rocky area of the Second cataract between Buhen (opposite Wadi Halfa) and Samnah 40 mi. to the south, and a trading post was established yet further upstream, near the Third cataract at Kerma. Sesostris III, who completed the organization, was later worshiped as a god in the temples of the region. The mines of Nubia were now the main source of Egypt's gold, and ivory, ebony, leopard skins, ostrich feathers, precious stones and slaves were brought up from the Sudan to the trading centre on the Second cataract and there exchanged for the manufactured goods of Egypt, trinkets and weapons.

In Asia too the influence of the 12th-dynasty Pharaohs was strong. The court of the princes of Byblos became highly Egyptianized, and Egyptian objects bearing royal names have been found on the sites of Syrian cities. The Egyptian army campaigned in Palestine, and Asian slaves, whether merchandise or prisoners of war, became plentiful in wealthy Egyptian households. Amenemhet IV's reign was short, and the dynasty ended with a queen Sobekneferu (Scemiophris), whose name is found in the scanty remains of the Labyrinth.

6. End of the Middle Kingdom Through the Second Intermediate Period.—The succeeding period is very obscure. Manetho names the 13th (Diospolite) dynasty, the 14th (Xoite from Xois [Sakha] in lower Egypt), the 15th and 16th (Hyksos) and the 17th (Diospolite), but his chronology is impossible and only a few Hyksos names are preserved. The Abydos tablet ignores all between the 12th and 18th dynasties. The Turin papyrus preserves many names, and the monuments are forever adding to the list, but it is difficult to assign them accurately to their places. Hyksos names can in some cases be recognized by their foreign aspect, the peculiar style of the scarabs on which they are engraved or by resemblances to those recorded in Manetho. It seems clear that the country was soon again divided into principalities and that Manetho's dynasties ruled in part concurrently; E. Meyer's estimate of 210 years for the whole period is generally accepted. Among the kings of the 13th dynasty (including perhaps the 14th) not a few are represented by granite statues of colossal size and fine workmanship, especially at Thebes and Tanis (San al Hajar al Qibliyah), some by architectural fragments, some by graffiti on the rocks about the First cataract.

In its divided state Egypt would fall an easy prey to the foreigner. Manetho says that the Hyksos (*q.v.*) gained Egypt without striking a blow. Their domination must have lasted a considerable time, for their monuments and scarabs are found throughout upper and lower Egypt and even in Nubia. The Hyksos, in whom Josephus recognized the children of Israel, worshiped their own Syrian deity, identifying him with the Egyptian god Setekh (Set); the memory of the "Asians" was later anathema in Egypt, yet they seem to have adopted Egyptian culture and ruled as Pharaohs. This critical episode had a profound effect on the character and history of the Egyptian people. The spirited overthrow of the Hyksos ushered in the glories in arms and arts that marked the New Empire. The 17th dynasty, in which the chief names are Seqenenre and Kamose, began the struggle, at first as semi-independent kinglets at Thebes. In his victory steles Kamose relates how he prevented his Hyksos opponent, Apophis, from communicating with the hostile ruler of Cush (Nubia), and even besieged him in his capital, Avaris (later Tanis).

7. New Empire.—*18th Dynasty.*—The epithet "new" is generally attached to this period, and "empire" instead of "kingdom" marks its wide conquests and organized rule abroad. The glorious 18th dynasty was closely related to the 17th. Its first task was to crush the Hyksos power in the northeast of the delta; this was fully accomplished when its founder, Ahmose (Aahmes or Amasis, brother of Kamose), captured Avaris. He next attacked them in southwestern Palestine, where he captured Sharuhen after a siege of three years. He fought also in Nubia, besides overcoming factious opposition in his own land. Amenhotep I (Amenophis) fought in Libya and Nubia, pushing the Egyptian frontier south to the Third cataract. After about 30 years of settled rule uninterrupted by revolt, Egypt was strong enough and rich enough to indulge to the full its new taste for war and lust of conquest. It had become essentially a military state. The whole of the administration was in the hands of the king with his vizier and other court officials; no trace of the feudalism of the Middle Kingdom survived. Thutmose I (Thothmes, Gr. Tuthmosis; *c.* 1525) thoroughly subdued Cush (Nubia and the lower Sudan), which had already been placed under the government of a viceroy; his dominion now reached beyond Napata (just below the Fourth cataract). Turning next to Syria, Thutmose carried his arms as far as the Euphrates. He made the first of those great additions to the temple of the Theban Amon at Karnak by which the Pharaohs of the empire rendered it by far the greatest existing temple in the world. He was the first of a long line of kings to be buried in the Valley of the Tombs of the Kings at Thebes.

A son by a secondary marriage, Thutmose II, succeeded as the husband of his half-sister Hatshepsut; during his short reign he put down a revolt in Nubia. At his death in 1504 B.C. Thutmose III, his son by a concubine, was only a child, and the ambitious Hatshepsut, appointed regent, soon assumed the full regal power; upon her monuments she wears the masculine garb and aspect of a king, though the feminine gender is retained for her in the inscriptions. This extraordinary woman must have had a great influence over her stepson and she was the acknowledged ruler of Egypt. Hatshepsut cultivated the arts of peace. She restored the worship in those temples of upper and lower Egypt that had not yet recovered from the neglect of the Hyksos. She completed and decorated the temple of Dayr al Bahri, embellishing its walls with scenes calculated to establish her claims, representing her divine origin and her association on the throne by her human father. Other reliefs depict the great expedition by water to Punt, the land of incense to the south of the Red sea, and the transport of two obelisks by water to Karnak. At Karnak, Hatshepsut laboured chiefly to complete the works projected in the reigns of Thutmose I and II. A date of the 22nd year of her reign was found at Sinai, no doubt counted from the beginning of the coregency with Thutmose III. Not much later, in his 22nd year, Thutmose was reigning alone in full vigour. While she lived, the personality of the queen secured the devotion of her servants and held all ambitions in check. Not long after her death there was a violent reaction. Her cartouches began to be defaced, her monuments hidden by other buildings, and the same rage pursued some

of her most faithful servants in their tombs. In the royal lists of Seti I and Ramses II, Hatshepsut has no place, nor is her reign referred to on any later monument.

The immense energy of Thutmose III now found its outlet in war. Syria had revolted, perhaps in the years of inactivity following Thutmose I's death; now the young king was ready to lead his army against the rebels. Unlike his predecessor, who merely overran one after another a series of isolated city-states, Thutmose had to face the organized resistance of a large combination, embracing the whole of western Syria and headed by the city of Kadesh on the Orontes. Six carefully planned campaigns had to be fought in order to reach and capture that city. In the 33rd year of his reign he marched through Kadesh, fought his way to Carchemish, defeated the forces that opposed him there and crossed the Euphrates into the territory of the Hurrian king of Mitanni. His annals record 17 separate campaigns in Palestine and Syria and list the immense booty and tribute obtained from that rich country. Egypt was master of an empire reaching to the Amanus mountains, and the neighbouring great powers hastened to send diplomatic presents.

In the intervals of war Thutmose III proved himself a wonderfully efficient administrator, dispensing justice in his realm, rebuilding and adorning temples with the labour of his captives and the plunder and tribute of conquered cities, or designing with his own hand the gorgeous sacred vessels of the sanctuary of Amon. In his later years some expeditions took place into Nubia. The children of the subdued Asian princelings were taken as hostages to Egypt and there educated to succeed their fathers with a due understanding of the might of the Pharaoh both to protect and to punish. Thus was an empire established on a sound basis probably for the first time in history. Thutmose died in the 54th year of his reign, perhaps the greatest of all the Pharaohs.

Thutmose III was succeeded by his son Amenhotep II, an athlete of great prowess, if his own claims are to be believed. In his seventh year he led an army into Syria, where revolt was again rife; he reached the mouth of the Orontes and returned home to Thebes with seven captive kings and much spoil. The kings he sacrificed to Amon and hanged six of the bodies on the walls, while the seventh was carried south to Napata and there exposed as a terror to the Cushites. Another revolt in Palestine was put down two years later. Amenhotep reigned 26 years and left his throne to his son, Thutmose IV, who warred in Syria and married a Mitannian princess.

His son, Amenhotep III (1417–1379 B.C.), was a mighty builder, especially at Thebes, where his reign marks a new epoch in the history of the great temples, Luxor being his creation, while avenues of rams, pylons and courts were added on a vast scale to Karnak. In Nubia he built impressive temples at Sedeinga and Soleb. His queen, Tiy, though apparently of humble parentage, was held in great honour by her husband as afterward by her son. Through the fortunate discovery of cuneiform tablets deposited by his successor in the archives at Tell el-Amarna, it can be seen how the great rulers of the day, the king of the Hittites, the king of Alashia (Cyprus) and the kings of Mitanni, Assyria and Babylonia, corresponded with Amenhotep, exchanged rich presents and gave their daughters to him in marriage. The descendants of the Syrian dynasts conquered by his father, having been educated in Egypt, ruled their paternal possessions as the vassals of the Pharaoh. A constant stream of tribute poured into Egypt, sufficient to defray the cost of all the splendid works that were executed. The so-called Colossi of Memnon, all that is left of Amenhotep's vast funerary temple in western Thebes, are seated statues of the king which originally flanked the gateway; his palace, one of the few royal residences whose plan archaeology can recover, lay farther south on the same bank. Its walls, of mud brick, were plastered and adorned with naturalistic frescoes. Near it was the vast lake built for the pleasure of Queen Tiy.

Ikhnaton (*q.v.;* Amenhotep IV), son of Amenhotep III and Tiy, was perhaps the most remarkable character in the long line of the Pharaohs. He was a religious fanatic, who had probably been high priest of the sun-god at Heliopolis and had come to view the sun as the visible source of life, creation, growth and activity, whose power was demonstrated in foreign lands almost as clearly as in Egypt. Devoting himself to the cult of the Aton, a form of the sun-god Re-Horakhte (Re-Horus of the horizon), he initiated a sweeping reform by which apparently the new cult alone was permitted. Of the old deities, Amon represented by far the wealthiest and most powerful interests, and against this long-favoured deity the Pharaoh hurled himself with fury. He changed his own name from Amenhotep ("Amon is satisfied") to Ikhnaton ("it pleases Aton"), erased the name and figure of Amon from the monuments, abandoned Thebes, and built a new capital on a virgin site at Akhetaton (Tell el-Amarna). This he dedicated to Aton in the sixth year, while splendid temples, palaces, houses and tombs were rising around him. In all local temples the worship of Aton was instituted. The confiscated revenues of Amon and the tribute from Syria and Cush provided means for adorning Akhetaton, and for richly rewarding those who adopted the Aton teaching fervently.

But meanwhile the political needs of the empire were neglected. Already in the last years of Amenhotep III's long reign the Amarna letters show that all was not well in Syria and Palestine; the Hittites were beginning to extend their hegemony over the principalities of northern Syria, and the dynasts were quarreling among themselves, while invaders from the Eastern desert harried Egyptian possessions in Palestine. In face of so dangerous a situation Amenhotep, grown elderly and indolent, seems to have done little, and his son, immersed in his own preoccupations, had little inclination for a campaign to demonstrate Egyptian might. Those of the vassal rulers who relied on Pharaoh and remained loyal as their fathers had done sent letter after letter vainly appealing for aid against their foes. It is not known whether these letters were answered or not; they seem to indicate a sharp decline in Egyptian prestige in western Asia.

Ikhnaton died in or about the 17th year of his reign, *c.* 1362 B.C. He had six daughters, who appear with him in the tomb reliefs of the courtiers, but no son. Two sons-in-law, mere boys, followed him with brief reigns, but the second, Tutankhaton, soon changed his name to Tutankhamen (*q.v.*) and, without abandoning Akhetaton entirely, began to restore to Karnak its ancient splendour with new monuments dedicated to Amon. Ikhnaton's reform had not reached deep among the masses of population; they probably retained all their old religious customs and superstitions, while the priesthoods throughout the country must have been fiercely opposed to the heretic's work, even if silenced during his lifetime by force or bribery. Tutankhamen died after nine years of reign and was buried at Thebes in the famous tomb that Lord Carnarvon and Howard Carter found still packed with its precious furniture. One more adherent of Ikhnaton, a priest named Ay, ruled for a short time.

At length a soldier, Horemheb, came to the throne as a wholehearted supporter of the old religion, and soon Aton and the whole of his royal following suffered the fate they had imposed upon Amon. Their monuments were destroyed and their names and figures erased, while those of Amon were restored. Horemheb as a practical man had to bring order into the long-neglected administration of the country and to suppress the extortions of the official classes by severe measures. He probably began also to re-establish the prestige of Egypt by military expeditions in the surrounding countries.

19th Dynasty.—Ramses (Ramesses) I, who succeeded as founder of the 19th dynasty, was an old soldier who had been Horemheb's vizier. In his brief reign of two years he planned and began the great colonnaded hall of Karnak. His son, Seti (Sethos) I, subdued the Bedouin Shasu, defeated a coalition near Beisan in Galilee and made himself master of Palestine. In subsequent campaigns cedars were felled for him by the Syrian princes, and the Phoenicians paid homage. Amor on the Syrian coast was subdued, and the Egyptians found themselves in conflict with the Hittites. The latter, however, were now firmly established in the Orontes valley, and a treaty probably ended the wars of Seti. In his ninth year he turned his attention to the gold mines in the Eastern desert of Nubia and improved the road thither. Meanwhile throughout Egypt the injuries done to the monuments by Ikhnaton were thor-

oughly repaired; the erased inscriptions and figures were restored, not without blunders. Seti's temple at Abydos and his galleried tomb in the Valley of the Tombs of the Kings stand out as the most splendid examples of their kind in design and in decoration.

Ramses II (1304–1237 B.C.) succeeded at an early age and reigned 67 years, during which he finished much that was begun by Seti and filled all Egypt and Nubia with his own monuments, some of them beautiful but most, entrusted to inferior workmen, of coarse execution. The excavation of the rock temple of Abu Simbel (Abu Sunbul), the building of his mortuary temple known as the Ramesseum, and the completion of the great hall of pillars at Karnak, begun by his father, were his greatest achievements in architecture. His wars began in his second year, their field comprising the Nubians, the Libyans, the Syrians and the Hittites. In his fifth year, near Kadesh on the Orontes, his army was caught unprepared and divided by a strong force of chariots of the Hittites and their allies, and Ramses himself was placed in the most imminent danger. But through his personal courage the enemy was kept at bay till reinforcements came up and turned the disaster into a victory. The incidents of this episode were a favourite subject in the sculptures of his temples, where their representation was accompanied by a poetical narrative, but Kadesh was not captured and, after further contests, in Ramses II's 21st year, the Hittite king Hattusili made peace, with a defensive alliance against foreign aggression and internal revolt (*see* HITTITES), sealed by the marriage of the Pharaoh to Hattusili's daughter. Ramses II paid much attention to the delta and resided there constantly. The temple of Tanis must have been greatly enlarged and adorned by him; a colossus of the king placed there was more than 90 ft. high, exceeding in scale even the greatest of the Ramesseum colossi.

Toward the end of his long reign the vigilance and energy of the old king diminished. The military spirit awakened in the struggle with Hyksos had again departed from the Egyptian nation. Mercenaries from the Sudan, Libya and the northern nations supplied the armies, while foreigners settled in the rich lands of the delta and harried the coasts. Merneptah (1236–23 B.C.), succeeding his father, had to fight many battles for the preservation of his kingdom and empire. In his mortuary temple at Thebes he set up a stele recording a great victory over Libyan immigrants and invaders in his fifth year. The last lines of the stele summarize his military achievements: Libya was wasted, the Hittites pacified, Canaan, Ascalon, Gezer and Yenoam in Galilee sacked and plundered; "Israel is desolated and has no seed, Khor [Palestine] has become a widow for Egypt." With them were "sea peoples" from the coasts of Asia Minor and the Aegean, only a few of whose names can be identified. Some had been allies of the Hittites at the battle of Kadesh.

Sherden mercenaries, distinguished by their remarkable helmets and broadswords, had long been a valuable contingent in the Egyptian army. The Lukku came from Lycia, the Teresh perhaps from Tarsus. The Akaiawasha may be Achaeans from the Aegean. Probably they were all sea rovers who were willing to leave their ships and join the Libyans in raids on the rich lands of Egypt.

20th Dynasty.—Merneptah, the 13th son in the huge family of Ramses II, must have been old when he ascended the throne. He was followed by a quick succession of inglorious rulers. A great papyrus written after the death of Ramses III and recording his gifts to the temples briefly reviews the conditions of the troublous times that preceded his reign. "The land of Egypt was in the hands of chiefs and rulers of towns, great and small slaying each other," but order was restored by Setnakht (1200–1198), the first of the 20th dynasty.

At the accession of Setnakht's son, Ramses III, the dangers that menaced Egypt were similar to those Merneptah had to meet. Again the Libyans and the "peoples of the sea" were acting in concert. The latter now included the Peleset (probably Philistines) and their kinsmen the Thekel; they invaded Syria from Asia Minor, reached Carchemish and progressed southward through Syria and Palestine while their ships gathered plunder from the coasts of the delta. This fleet was overthrown with heavy loss by the Egyptians, in whose ranks served many Aegean and Sherden mercenaries.

Meanwhile the chariots and warriors of the Peleset and their associates were advancing through Syria, their families and goods following in oxcarts and their ships accompanying them along the shore. Ramses led out his army and fleet against them and struck them so decisive a blow that the migrating swarm was dispersed; the Philistines settled, perhaps as vassals of Egypt, in the coastal plain that was to bear their name. In Ramses' 11th year another Libyan invasion had to be met and his suzerainty in Palestine forcibly asserted.

His vigour was equal to all these emergencies and the later years of his reign were spent in peace. To the temples he not only restored the property that had been given to them by former kings but also added greatly to their wealth, the Theban Amon receiving by far the greatest share. His palace was at Medinet Habu in the south quarter of western Thebes, and there he built a great temple adorned with scenes from his victories and richly provided with divine offerings. Shortly before the death of the old king, a plot in the harem to assassinate him and apparently to place one of his sons on the throne was discovered and its investigation ordered, leading after his death to the condemnation of many high-placed men and women.

Eight kings of the name of Ramses now followed each other ingloriously in the space of about 80 years to the end of the 20th dynasty. The Libyans began again their encroachments, and there was undoubtedly corruption in high places, and great distress among certain portions of the population. At this time the Theban necropolis was being more systematically robbed than ever before. Under Ramses IX an investigation took place which showed that one of the royal tombs before the western cliffs had been completely ransacked and the mummies burned. Prosecution of the criminals was ineffective, and three years later the Valley of the Tombs of the Kings was attacked and the tombs of Seti I and Ramses II were robbed.

21st Dynasty.—The authority of the last king of the 20th dynasty, Ramses XI, was shadowy. After grave disturbances and perhaps civil war, Herihor, the high priest of Amon, viceroy of Cush and generalissimo of the army, took the royal titles at Thebes during Ramses' lifetime.

From this period dates a remarkable papyrus, perhaps a historical romance, known as the *Story of Wenamun.* Sent to Syria by Herihor in order to obtain timber from Byblos for the bark of Amon, Wenamun learned to his cost that the ancient prestige of Egypt in Syria had entirely disappeared.

8. Libyan Dynasties and the Cushites (Ethiopians).—On the death of Ramses XI a prince at Tanis named Smendes (Nesubanebdet) founded a separate dynasty in the delta. The Tanite kings generally had the overlordship of the high priests of Thebes. The descendants of Herihor, however, sometimes by marriage with princesses of the other line, could assume cartouches and royal titles. Cush may have been ruled with the Thebais, but the records of the time are very scanty. The mummies from the despoiled tombs of the kings at Thebes were the object of much anxious care to the kings of this dynasty. After being removed from one tomb to another, they were finally deposited in a shaft near the temple of Dayr al Bahri, where they were discovered in modern times; others were found in the tomb of Amenhotep II. The Tanites themselves were buried in the necropolis at Tanis (San), where several of their rich tombs were found together with that of Osorkon II by Pierre Montet.

22nd Dynasty.—Libyan soldiers had long been employed in the army. Their military chiefs settled in the large towns and acquired wealth and power while the native rulers grew weaker and weaker. The rulers of the 22nd dynasty were clearly, from their names, of foreign extraction, and their genealogy indicates a Libyan origin. Sheshonk (Shishak) I, the founder of the dynasty (*c.* 945 B.C.), seems to have fixed his residence at Bubastis in the delta, married his son to the daughter of the last king of the Tanite dynasty, and secured Thebes, making one of his sons high priest of Amon. Whereas Solomon appears to have dealt with a 21st-dynasty king on something like an equal footing, Sheshonk re-established Egyp-

ANCIENT EGYPT: A CHRONOLOGICAL SYNOPSIS

NEOLITHIC AND CHALCOLITHIC PERIODS, *c.* 4500 B.C.	The earliest settlements. Domestication of animals and growth of agriculture. Contacts with western Asia. Use of copper in the later phases
"DYNASTY 0"	Predynastic kings of upper Egypt and of lower Egypt. Monuments of King "Scorpion" and (?)Narmer at Hierakonpolis. Copper working and stone cutting develop
1ST DYNASTY, *c.* 3100–*c.* 2890 B.C.	The Age of Mena (Menes). Eight or nine kings. Unification of Egypt. Royal tombs at Saqqarah and Abydos. First use of stone in building. Syrian timber imported. Development of hieroglyphic writing
2ND DYNASTY, *c.* 2890–*c.* 2686 B.C.	Ten kings, including Peribsen, Khasekhem, Khasekhemui. Rivalry of the cults of Horus and Set
3RD DYNASTY, *c.* 2686–*c.* 2613 B.C.	Five kings
BEGINNING OF THE OLD KINGDOM	
	The age of Imhotep. Zoser's Step pyramid and temple, the first buildings in stone. Pyramids at Saqqarah, Zawiyat Al Aryan and Maydum; mastabas at Bet Khallaf
4TH DYNASTY, *c.* 2613–*c.* 2494 B.C.	Eight kings: Sneferu, Khufu (Cheops), Rededef or Dedefre, Khafre (Chephren), Baufre(?), Mycerinus, Shepseskaf, Dedefptah. The age of the great pyramid builders. Pyramids at Dahshur, Maydum, Giza and Abu Roash. Mining expeditions to Sinai; trade with Byblos for timber; stone quarrying at Aswan and in the Nubian desert
5TH DYNASTY, *c.* 2494–*c.* 2345 B.C.	Nine kings: Userkaf, Sahure, Neferirkare Kakai, Shepseskare Isi, Neferefre, Niyuserre, Menkauhor, Djedkare Isesi, Unas. Rise of the Heliopolitan priesthood; Re-worship the state cult. Pyramids with sun-temples at Abu Sir. Compilation of the Palermo stone annals. Battles with the Libyans. Trading expeditions to Punt and Nubia. Beginning of the decentralization of power. Rise of the nomarchs
6TH DYNASTY, *c.* 2345–*c.* 2181 B.C.	Eight kings, including Teti, Userkare (Phiops), Pepi I Meryre, Merenre, Pepi II Neferkare, three more including a queen, Nitocris. Increased maritime trade with Byblos and Punt. Nomarchs of Elephantine lead caravans to Sudan. Growth of nomarchs' power. Larger private, smaller royal tombs. Pepi II reigns over 90 years, the longest reign in history
END OF THE OLD KINGDOM THE FIRST INTERMEDIATE PERIOD	
7TH, 8TH, 9TH and 10TH DYNASTIES, *c.* 2181–2040 B.C.	Weak kings ruling from Memphis. Upper Egyptian nomarchs setting up local dynasties. Asians and Libyans in delta. Rise of the dynasty of Heracleopolis (9th dynasty): Kheti I (Akhthoes) and 12 successors who hold the north. Nomarchs of Thebes gain adherents, and attack the Heracleopolites
BEGINNING OF THE MIDDLE KINGDOM	
11TH DYNASTY, *c.* 2133–1991 B.C.	Nine(?) kings. Three or four kings of upper Egypt called Intef or Mentuhotep, including Mentuhotep II Nebhapetre, Mentuhotep III S'ankh-ka-Ra, Mentuhotep IV Nebtawire. Mentuhotep II unifies Egypt *c.* 2040 B.C. Asians expelled from delta. Reopening of trade with Asia and Punt. Conquest of Nubia begun
12TH DYNASTY, 1991–1786 B.C.	Eight kings: Amenemhet (Ammenemes) I, Sesostris (Senusret) I, Amenemhet II, Sesostris II, Sesostris III, Amenemhet III, Amenemhet IV, Queen Sobekneferu. The golden age of art and craftsmanship. Prosperity restored to Egypt. Tombs of wealthy nomarchs. Removal of the capital from Thebes to Al Fayyum. Intensified commerce with Palestine and Syria. Conquest of lower Nubia completed. Flood control and land reclamation in Al Fayyum
END OF THE MIDDLE KINGDOM	
13TH and 14TH DYNASTIES, 1786–*c.* 1603 B.C.	Numerous kings, many of whose names are preserved in the Turin Canon
BEGINNING OF THE SECOND INTERMEDIATE PERIOD	
	Power of the later kings confined to upper Egypt
15TH and 16TH DYNASTIES, *c.* 1720–1570 B.C.	Six "Great Hyksos" and eight(?) "Little Hyksos." The Hyksos kings rule from Avaris, on the delta, one or two probably controlling all Egypt
17TH DYNASTY, *c.* 1645–1567 B.C.	17(?) kings including 11 nomarchs of Thebes and 6 kings of upper Egypt, the last of whom were Sequenenre and Kamose. Kamose wages a war of liberation against the Hyksos
BEGINNING OF THE NEW EMPIRE	
18TH DYNASTY, 1567–1320 B.C.	14 rulers: Ahmose (Amasis) I, Amenhotep (Amenophis) I, Thutmose (Tuthmosis) I, Thutmose II, Queen Hatshepsut, Thutmose III, Amenhotep II, Thutmose IV, Amenhotep III, Ikhnaton (Amenhotep IV), Smenkhkare, Tutankhamen, Ay, Horemheb. Expulsion of the Hyksos and reconquest of Nubia. Reorganization of Egypt. Growth of the temple of Amon at Karnak. The age of colonial expansion: conquest of Palestine and Syria to the Euphrates, and of Nubia (Cush) to the Fourth cataract. Diplomatic exchanges between the great powers. The Amarna age. Great wealth of Amenhotep III, the Magnificent (1417–1379 B.C.). The Aton heresy: Amon proscribed by Ikhnaton; reinstated by Tutankhamen. Horemheb restores order
19TH DYNASTY, 1320–1200 B.C.	Eight kings: Ramses (Ramesses) I, Seti (Sethos) I, Ramses II, Merneptah, 3 more kings and a queen Tawosret. Beginning of the Ramesside age. Seti I's campaigns in Palestine and Syria. Ramses II (1304–1237 B.C.) meets Hittites at Kadesh; later concludes treaty with them. Merneptah repels Libyans and "sea peoples." Dynasty ends in confusion

ANCIENT EGYPT: A CHRONOLOGICAL SYNOPSIS (*Continued*)

20TH DYNASTY, 1200–*c.* 1085 B.C.	Ten kings: Setnakht, Ramses III, Ramses IV to XI in succession. Order restored by Setnakht. Ramses III (1198–1166 B.C.) withstands pressure from Libyans and "sea peoples." Gradual decline of Egypt, at home and abroad, ending in divided rule
END OF THE NEW EMPIRE	
21ST DYNASTY, *c.* 1085–*c.* 945 B.C.	Seven "Tanite" kings: Smendes, Psusennes I; 5 others. A dynasty ruling at Tanis, while the high priest Herihor and his successors controlled Thebes. Cush independent
22ND and 23RD DYNASTIES, *c.* 945–*c.* 730 B.C.	Nine kings and five or more kings respectively, descended from Meshwesh chieftains including Sheshonk (Shoshenk or Shishak) I, Osorkon I, Takelot I, Osorkon II and others bearing these or similar Libyan names. Centres of government at Tanis and Bubastis. Control over Thebes generally maintained by 22nd dynasty but an independent line, the 23rd dynasty, established *c.* 820 B.C. at Thebes; thereafter divided rule. Records of this period scarce
24TH DYNASTY, *c.* 730–*c.* 715 B.C.	Two kings: Tefnakhte, Bocchoris. Capital, Sais in the delta
25TH DYNASTY, *c.* 730–656 B.C.	Five kings: Piankhi, Shabaka, Shebitku (Shabataka), Taharqa, Tanutamon. The Dynasty of Napata or the Cushite dynasty. Conquest of Egypt by Piankhi in his 21st year. Revival of prosperity, restoration of cults. Influence of the "God's wife" in Thebes.
	Interference in Palestine leads to conflict with Assyria; Assyrians conquer Egypt and instal Necho as governor in Sais
26TH DYNASTY, *c.* 664–525 B.C.	Seven kings: Necho I, Psamtik (Psammetichus) I, Necho II, Psamtik II, Apries, Ahmose II (Amasis), Psamtik III. The Saite age, or age of the Renaissance. Expulsion of the Assyrians. Sais the capital of a prosperous kingdom. Archaism in art. Use of foreign mercenaries including Lydians and Greeks. Foundation of Naukratis. Necho's Palestinian venture repulsed by the Babylonians. Control of Cyprus. Revolt of the army against Apries. Philhellenic policy of Ahmose
PERSIAN DOMINATION OF EGYPT	
27TH DYNASTY, 525–404 B.C.	Eight Achaemenid kings: Cambyses II, Darius the Great and their successors to Darius II Nile-Red Sea canal completed
28TH DYNASTY, 404–399 B.C.	One king: Amyrtaeus of Sais. Liberation of Egypt, with Greek aid
29TH and 30TH DYNASTIES, 399–343 B.C.	Seven kings: four kings of Mendes and three kings of Sebennytos. To Manetho's list of 30 dynasties may be added Eusebius' 31st dynasty
THE SECOND PERSIAN DOMINATION	
31ST DYNASTY, 343–335 B.C.	Three kings: Artaxerxes III Ochus, Arses, Darius III
Alexander the Great reached Alexandria in 332 B.C. Ptolemy I crowned as Pharaoh in 304 B.C.	

tian rule in Palestine. His expedition in the fifth year of Rehoboam subdued Israel as well as Judah, to judge by the list of conquered towns inscribed on his great pylon in the temple of Karnak.

Osorkon I inherited a prosperous kingdom from his father, but no further progress was made. It required a strong hand to curb the Libyan chieftains, and divisions soon began to show themselves in the kingdom.

23rd Dynasty.—The 22nd dynasty lasted through many generations; but there were rival kings, and it seems that the 23rd dynasty was contemporaneous with the end of the 22nd. These kings had little hold upon the subject princes, who spent the resources of the country in feuds among themselves.

24th and 25th Dynasties.—A separate kingdom had meanwhile been established in Cush with its capital at Napata (*q.v.*) near the Fourth cataract. Our first knowledge of it is at a moment when the Cushite or Ethiopian king Piankhi already held the Thebais. The energetic prince of Sais (Sa al Hajar), Tefnakhte, followed by most of the princes of the delta, subdued most of middle Egypt, and by uniting these forces threatened the Cushite border. Heracleopolis Magna, however, held out against Tefnakhte; and Piankhi, coming to its aid, not only drove Tefnakhte out of middle Egypt but also captured Memphis (*c.* 730 B.C.) and received the submission of a number of princes and chiefs. According to Diodorus Siculus the Ethiopian state was theocratic, ruled through the king by the priests of Amon. The account is probably exaggerated; but even in Piankhi's record the piety of the king, especially toward Amon, is very marked.

After Piankhi's return to Napata, Tefnakhte continued to rule at Sais. He and his son Bocchoris (Bakenrenef; 720–715 B.C.) compose the brief 24th dynasty. Another Ethiopian invader, Shabaka, is said to have burned Bocchoris alive. The Ethiopian rule of the 25th dynasty was now firmly established, and the resources of the two countries together might have been employed in conquest of Syria and Phoenicia. But at this very time the Assyrian empire, at the peak of its power, began to menace the west. Assyrian armies had reached the frontier of Egypt at Rafah (Raphia) in 734 and in 720 B.C. The Ethiopians continued the policy of their predecessors in lending support and encouragement to the peoples of Syria and Palestine in their resistance to the invaders. Shabaka maintained outwardly friendly relations with Sargon II. His nephews Shebitku (Shabataka) and Taharqa were energetic in opposing Sennacherib's advance against Jerusalem; but in 671 B.C. Esarhaddon defeated an army on the border of Egypt, captured Memphis with the royal harem and took great spoil.

Esarhaddon entrusted the government and collection of tribute to the native chiefs; 22 princes in all are enumerated in the records. Scarcely had he withdrawn before Taharqa returned from his refuge in the south and the Assyrian garrisons were massacred. Esarhaddon promptly prepared a second expedition, but died on the way to Egypt in 669 B.C. His son Ashurbanipal sent it forward, routed Taharqa and reinstated the governors. At the head of these was Necho (Niku), king of Sais and Memphis, the son of Bocchoris, no doubt chosen for his hostility to the Ethiopians.

Upper Egypt, however, was loyal to Taharqa. After his death his nephew Tanutamon, received at Thebes with acclamation, besieged and captured Memphis, Necho being probably slain in the encounter. But in 661 Ashurbanipal drove the Ethiopian out of lower Egypt, pursued him up the Nile and sacked Thebes. This was the last and most tremendous visitation of the Assyrian scourge. All the Cushite kings from Piankhi to Tanutamon were buried in pyramids at their ancestral home. The dynasty continued to rule, at Napata and then at Meroe (*q.v.*). But, cut off from Egypt, their civilization became ever more barbarized. They never again invaded Egypt.

9. Saites.—*26th Dynasty.*—Psamtik I (Psammetichus; 663–610 B.C.), the son of Necho, succeeded his father as a vassal of Assyria, allied himself with Gyges, king of Lydia, and, aided by Ionian and Carian mercenaries, extended and consolidated his power. Ashurbanipal's energies throughout this crisis were entirely occupied with revolts nearer home, in Babylon, Elam and Arabia; Psamtik was thus able to expel the last Assyrian garrison and proclaim himself Pharaoh over all Egypt. Under his firm and wise rule, Egypt rapidly recovered its prosperity. Greek traders entered Egypt in increasing numbers and were encouraged to settle at Naukratis, near Sais. Psamtik guarded the frontiers of Egypt with three strong garrisons, placing the Ionian and Carian mercenaries especially at the Pelusian Daphnae (*q.v.*) in the northeast, from which quarter the most formidable enemies were likely to appear. Toward the end of his reign he sent support to the Assyrians against the attacks of the Medes and Babylonians, but in 612 B.C. Nineveh fell and the might of Assyria was broken.

Psamtik died in the 54th year of his reign and was succeeded by his son Necho (610–595 B.C.). The new Pharaoh, perhaps dreaming of a new empire in Asia, prepared an expedition to help the remnant of the Assyrian forces in Syria against the Babylonians. Josiah alone opposed him with his feeble force at Megiddo and was easily overcome and slain. Necho went forward to the Euphrates, put the land to tribute and, in the case of Judah at any rate, filled the throne with his own nominee, Jehoiakim. Nabopolassar, king of Babylon, soon dispatched his son Nebuchadrezzar to fight Necho. The Babylonian and Egyptian forces met at Carchemish (605), and the rout of the latter was so complete that Necho relinquished Syria and might have lost Egypt as well had not the death of Nabopolassar recalled the victor to Babylon. Herodotus relates that in Necho's reign a Phoenician ship dispatched from Egypt actually circumnavigated Africa, and the attempt was made to complete a canal through the Wadi Tumilat connecting the Mediterranean and the Red sea by way of the lower Egyptian Nile (*see* SUEZ CANAL).

The next king, Psamtik II (595–589 B.C.), visited Syria or Phoenicia and sent an army of Egyptians and mercenaries against the king of Napata, perhaps to forestall a threatened attack. The army may have penetrated to the Fourth cataract region; on its successful return Greek, Carian and Jewish mercenaries left graffiti on the rocks at Abu Simbel. Pharaoh Apries (Hophra; 588–569 B.C.) blockaded Tyre and fomented rebellion against the Babylonian suzerainty in Judah, but accomplished little there. Colonies of Jewish refugees were given shelter in Egypt. Herodotus describes Apries' reign as exceedingly prosperous. Later, however, a disastrous expedition sent to aid the Libyans against the Greek colony of Cyrene roused the suspicion and anger of the native soldiery at favours shown to the Greek mercenaries in Egypt, who of course had taken no part in it.

Ahmose (Amasis) II (569–526 B.C.) was chosen king and Apries put to death by the native Egyptians. Nonetheless, Ahmose employed Greeks in numbers and cultivated the friendship of their tyrants. His rule was confined to Egypt (and perhaps Cyprus) but Egypt itself was very prosperous. At the beginning of his long reign of 44 years he was threatened by Nebuchadrezzar; later he joined the league against Cyrus and saw with alarm the fall of his old enemy. A few months after his death, in 525, the invading host of the Persians led by Cambyses reached Egypt and dethroned his son, Psamtik III.

10. Persian Period.—*27th Dynasty.*—The policy of the Achaemenid kings seems to have been conciliatory to national beliefs and sentiments. There are conflicting views of Cambyses' reign. The courtier Udjahorresne depicts an ideal ruler in the pharaonic tradition; Herodotus, viewing the past through the eyes of a nationalistic priesthood, draws a savage tyrant. Certainly Darius (522–486 B.C.) proved himself a beneficent ruler and, in a visit to Egypt, displayed his consideration for the religion of the country. In the great oases he developed agriculture and built temples. The annual tribute imposed on the satrapy of Egypt and Cyrene was heavy, but it was probably raised with ease. The canal from the Nile to the Red sea was completed or repaired, and commerce flourished. Papyri dated in the 34th and 35th years of Darius are not uncommon, but apparently at the very end of his reign, several years after the disaster of Marathon, Egypt was induced to rebel. Xerxes (486–465 B.C.) put down the revolt with severity.

The disorders which marked the accession of his successor, Artaxerxes, gave Egypt another opportunity to rebel. The leaders were the Libyan Inaros and the Egyptian Amyrtaeus. Aided by an Athenian force, Inaros slew the satrap Achaemenes at the battle of Papremis and destroyed his army. The garrison of Memphis held out, however; a fresh host from Persia raised the siege and in turn besieged the Greek and Egyptian forces in the delta. After two years, having diverted the river from its channel, they captured and burned the Athenian ships and quickly ended the rebellion. The reigns of Xerxes II and Darius II are marked by no recorded incident in Egypt until a revolt (*c.* 410 B.C.) interrupted the Persian domination; by 404 B.C. Egypt was again independent under Amyrtaeus.

Monuments of the Persian rule in Egypt are exceedingly scanty. The inscription of Udjahorresne, priest of Neith at Sais and from his position the native authority who was most likely to be consulted by Cambyses and Darius, tells of his relations with these two kings. For the following reigns Egyptian documents hardly exist, but some papyri written in Aramaic have been found at Elephantine (*q.v.*) and at Memphis. Herodotus visited Egypt in the reign of Artaxerxes, about 440 B.C. His description of Egypt, partly founded on Hecataeus, who had been there about 50 years earlier, is the chief source for the history of the Saite kings.

28th–30th Dynasties.—Amyrtaeus of Sais, whom Manetho makes the sole Pharaoh of the 28th dynasty, reigned for only six years. The next king, Nefaurud (Nepherites I), a Mendesian, founded the 29th dynasty. Both he and Achoris (393–380 B.C.) aided the Greeks against the Persians with grain and ships and attempted to restore prosperity to Egypt, but after the brief reign of Nefaurud II the sovereignty passed to the 30th dynasty, the last native Egyptian line. Monuments of all these kings are known, and art flourished particularly under the kings of Sebennytos: Nekhtnebf (Nectanebo), Tachos and Nekhtharheb (Nectanebo II). The first came to the throne when a Persian invasion was imminent, 379 B.C. Achoris had already formed a powerful army, largely composed of Greek mercenaries. This army Nekhtnebf entrusted to the Athenian Chabrias. The Persians, however, succeeded in causing his recall and marched against Egypt with 200,000 barbarians and 20,000 Greeks. The Egyptians suffered an initial reverse but, through the indecision of the Persian general Pharnabazus, were able to collect their forces. After a pitched battle near Mendes, Pharnabazus retreated and Egypt was free.

Nekhtnebf was succeeded by Tachos or Teos, whose short reign was occupied by a war with Persia in which the king of Egypt secured the services of a body of Greek mercenaries under the Spartan king Agesilaus and a fleet under the Athenian general Chabrias. He entered Phoenicia with every prospect of success, but having offended Agesilaus he was dethroned in a military revolt which gave the crown to Nekhtharheb; a rival pretender almost succeeded in overthrowing the new Pharaoh, but Agesilaus defeated him and left Nekhtharheb established on the throne. The opportunity of a decisive blow against Persia, however, was lost. The new king Artaxerxes III Ochus determined to reduce Egypt. A first expedition in 351 B.C. was repulsed, but a second, commanded by Ochus himself, subdued Egypt with no further resistance than that of the Greek garrison of Pelusium. Nekhtharheb, last of the native Pharaohs, instead of endeavouring to relieve them, retreated to Memphis and fled thence to upper Egypt; the end came two years later in 341 B.C.

31st Dynasty.—Ochus treated his conquest barbarously. From this brief re-establishment of Persian dominion (counted by Manetho as the 31st dynasty) few documents survive; the attempted revolt of a prince of Mendes named Khabbash failed, and according to the *Demotic Chronicle* the country suffered oppression and the violation of its gods.

See also CHRONOLOGY: *Egyptian.*

B. Ancient Civilization and Culture

1. Land and People.—*The Two Lands.*—It is probable that in predynastic and dynastic times the Egyptian landscape was very different from that of today. The Nile valley, now a flat and monotonous stretch of cultivated alluvium stretching from desert to desert, was then more undulating. The Nile was constantly changing its course, winding between swamps of vegetation and depositing silt on its banks. On the land left when the water found a new channel, man would build his dwellings, cattle would graze in the meadows and grain would be sown in fields reclaimed from the swamp. Between the villages would be marshland and tall thickets of papyrus and sedges, interspersed with pools and lagoons teeming with fish, where mongoose hunted the young water birds and hippopotamus and crocodile lurked. The desert fringing the valley, and the wadies running up into the highland, were steppe land teeming with game; ibex, oryx and other antelope species could be hunted in plenty, as well as wild cattle, barbary sheep, hyenas and even lions and panthers.

"A land won by the Egyptians, and given them by the river"; in this well-known phrase Herodotus epitomizes the absolute dependence of the ancient Egyptians on their river, and the need for toil to exploit its gifts. The Nile god, Hapi (Apis), was depicted as fat and full-breasted, with the fruits of the earth and hieroglyphs denoting life and prosperity in his hands. Farmers awaited his annual rise with expectancy but also with anxiety, for a low Nile spelled famine while an unusually high one might bring disastrous floods. The conservation of water for irrigation, and the protection of towns and villages by dikes, were the first charge of the administration, and in times of emergency the *corvée* of all able-bodied members of the community might be called out to strengthen a mudbank or clear a canal of silt.

The land of Egypt was anciently regarded as two lands: upper Egypt and lower Egypt, roughly corresponding to the two geographically distinct regions, the Nile valley and the delta. These present a considerable contrast. The delta is wide, green and intensely fertile, veined with innumerable water channels; 12 branches of the Nile were said to flow through it in ancient times. These subdivided the area into a dozen small administrative districts or "nomes," each with a city as its centre of government. Little now remains of Heliopolis, Sais, Buto, Sebennytos, Bubastis, Mendes, Tanis and Pelusium, to name the greatest of these, for their sites are now waterlogged and little excavation can be undertaken on them. Yet Tanis, Bubastis and Sais were once successively capital cities of Egypt, and the wonders of Pi-Ramesses, the residence of the Ramesside Pharaohs, are rhapsodized upon by contemporary scribes, who describe its palaces, temples, parks and orchards, and its harbours crowded with ships bringing the tribute of foreign countries from every part of the Levant.

The People.—Upper Egypt, though linked with the delta by the Nile, is a different country. The climate is hotter and drier, it has a width of only a few miles and a length of several hundred. It is isolated by the desert on either side, it looks to Africa and not to Asia for its contacts. Even the temperament and physical characteristics of the people of the Saïd today seem different from those of their fellow countrymen in the north, and the ancient Egyptian was conscious of these differences also. So different were the dialects that a man of Elephantine, they said, would be lost in the delta. Through centuries of political unification and social integration, they preserved the sense of the duality of their civilization. Pharaoh was "lord of the two lands," he wore a double crown, a composite headdress combining the white crown of upper Egypt and the red of lower Egypt. On his forehead blazed a jeweled tiara bearing the head of the cobra, symbol of the goddess Udjo (Buto) of the delta and of the vulture-goddess Nekhbet, lady of Al Kab in the south. His titles likewise emphasized his dual authority. In early times, it seemed, there had been two treasuries and two granaries, and the duplication of administrative bureaus persisted throughout Egyptian history; in the New Kingdom, and perhaps earlier, there were two viziers, each with his staff. In times of political upheaval, the country tended to split into its component halves.

At the same time, the ancient Egyptians realized that they were essentially one nation, speaking one language and sharing one material culture and, in spite of the diversity of cults, one religion. They realized the interdependence of delta and valley and the vital necessity of united rule throughout the Nile valley under a strong and centralized government. "Uniting the two lands" was the rebus, symbolic of strength and prosperity, carved on the throne of the Pharaohs, and the intertwining of the heraldic lotus and papyrus plants was a favourite theme of art.

In fact, physical anthropologists can see no significant difference between the skull measurements of dynastic Egyptians from cemeteries in various parts of Egypt, though there was probably a greater admixture of negroid blood in the population of the south and of Mediterranean stock in the north. On either side of the Nile valley stretched the desert with its nomadic tribes perhaps not unlike the modern Beja, and in the oases and on the western littoral lived the fair, blue-eyed Libyans, who wore the penis-sheath, ostrich feathers in their hair, and brightly patterned tunics. To the south, beyond the First cataract, were the Nubians, a darker strain with slightly negroid features, "people of the bow," who took service as archers in the Egyptian army. Yet farther south, the Negroes of central Africa hovered on the fringes of civilization and tended to move northward and intermarry with the Nubians.

The dynastic Egyptians were sturdy and of moderate stature. Like the modern peasants who may be descended from them, they rarely became obese in old age. They had dark hair and brown skins; the fairer, reddish hair occasionally depicted may be due, it has been suggested, to intermarriage with Libyans. They matured young and married in their teens; a few lived to be very old —Pepi II, a Pharaoh of the 6th dynasty, probably lived to be nearly 100—but infant mortality was high and expectation of life probably not much more than 36 years. In spite of perpetual sunshine and a varied diet rich in vitamins, they suffered from bad teeth and a variety of diseases, as the mummies show (see *Medicine* below). The population of Egypt in pharaonic times has been variously estimated. One scholar believes that it may have been as high as 5,000,000 in the heyday of the empire; another, on equally slender evidence, suggests that in the 11th dynasty, *c.* 2100 B.C., it may have not been above 1,500,000.

Natural Resources.—Having regard to the limited needs of ancient man, the land of Egypt was so rich in natural resources as to be almost entirely self-sufficient. Crops of all kinds could be grown on the rich alluvium, herds of cattle, goats, pigs and sheep were pastured, fish abounded, and waterfowl teemed in the reed brakes. The thick-stemmed papyrus provided fibre for paper, baskets, matting, sandals, boat-building and a thousand other purposes. On the desert steppe game abounded. Natron and salt, natural preservatives, were at hand, and the sun itself purified and preserved. The fine mud in the river bank provided an inexhaustible supply of building material and potter's clay. Acacia, palm, sycamore, fig and tamarisk furnished timber for the builder, the joiner and the shipwright. In the cliffs bordering the Nile valley there were plentiful supplies of fine limestone and sandstone for the mason, and the sculptor could obtain harder stones such as breccia, alabaster, quartzite, diorite, porphyry and schist at no great distance from the river. Copper ore was mined in the Eastern desert and in the Sinai peninsula and gold from deposits, some of which are still worked, in the Arabian desert, upper Egypt and Nubia. The Nile, that main artery linking every city and district, provided so universal a mode of transport that wheeled traffic was almost unknown; goods have always been carried more easily and cheaply by water in Egypt. From the First cataract to the sea, the river is navigable throughout its length at all seasons.

Agriculture.—Cereals do not grow wild in Egypt and the first dwellers in the Nile valley probably caught fish and game, and ate papyrus rhizomes and wild fruits and herbs. Barley and emmer wheat were probably introduced by immigrants from Palestine. In the fertile Nile valley, grain grew readily, and gradually the casual growing of a catch crop would give place to planned cultivation by a settled community living in huts, domesticating animals and at length making pottery and baskets for their needs.

The Neolithic inhabitants of the Fayyum area, living on the shores of a lake, caught fish but also grew corn, storing their grain in large basketry silos.

Peasants at work in the fields, engaged upon the seasonal pursuits of agriculture, are a favourite theme of tomb decorations. From these scenes the whole cycle of events can be reconstructed. As the inundation waters receded, leaving a layer of soft mud over the fields, the plowman and the sower went to work. The plow was at first a simple wooden affair; later a more elaborate type came into use consisting of two upright handles joined with cross-pieces and attached near the ground to a curved plowshare, shod with metal or hard wood. The long shaft was attached to a yoke at the further end, and this rested on the necks of the two draft animals, usually cows, and was fastened to their horns. Usually one plowman drove the team while another guided the plow, driving the plowshare into the ground. In front walked the sower, scattering his seed so that the animals trod it in and the share turned over the sod. In very soft ground the plow might even be dispensed with and a herd of animals driven over the ground to tread in the seed. Irrigation was sometimes necessary while the crops were growing; water was raised from river or canal with the aid of a *shaduf* or swipe similar to that in use today, consisting of a bucket suspended by a rope from a swinging beam and raised with the aid of a counterpoise on the other end of the beam. Men and women alike turned out for the harvest; the ripe ears were cut with short-handled sickles and collected in nets or baskets to be taken to the threshing floor. Oxen or sheep were driven over the heaped ears to tread out the grain, which was then forked and winnowed by raising a scooped handful high above the head and letting it fall so that the chaff blew away. The grain was then taken off to be stored in corn bins, large domed mud-brick structures with an opening at the top reached by ladder, and a trapdoor at the base through which a bushel or two could be taken as required. Another staple crop was flax, which was carefully pulled from the ground, the stems packed into bundles and taken to be hatcheled, and the fibres then prepared for the weavers. Cotton was not grown until the Persian period.

Fruit and vegetables needed perennial irrigation and were grown in orchards and market gardens crisscrossed with water channels. Fruit trees included the fig, the date and dom palm, the zizyphus and later the apple, olive and pomegranate. The *Moringa* or *bak* yielded a vegetable oil, and olive oil was produced in the New Empire. The pod of the carob was a popular sweetener, more easily obtained than honey, though scenes of beekeeping are known.

Viticulture was an important activity, especially in the delta and the oases. There were a number of royal and temple vineyards, and some in private ownership. Scenes from the tombs depict the gathering of the bunches and the merry scene at the vats, where the harvesters hold fast to ropes fastened to an overhead crossbar, in order to maintain their balance while they tread the slippery grapes. When the treading was completed, the grape skins were squeezed in a bag press to extract the last drops of juice.

2. Social Life.—*The Family.*—The painters and sculptors of ancient Egypt depict an ideally happy family life. The master of the house sits at home, or engages in outdoor pastimes, with his wife by his side; the two clasp hands, or twine their arms round each other's waists, while children embrace their father's legs or dutifully bring offerings. It was the duty of a son to love and honour his parents and to help his father, acting as his "staff of old age" when the time came. Among the papyri are many tender love poems (see *Literature* below). Women do not seem to have been secluded except perhaps those, many of them foreigners, who entered the royal harem. A wealthy man might have concubines, but it was his wife, the lady of the house, who shared his tomb and is depicted on terms of near equality with him on his tomb stele. Among the surviving papyri are marriage contracts but no bills of divorce. Women could own and inherit property and engage in trade. It is probable, however, that women did not normally learn to read and write, and none, so far as is known, could hold office in the government.

There were about six children in the average well-to-do family. Children were much loved and not a little spoiled by their parents, as the private correspondence of a Middle Kingdom official hints. On the other hand, the proper behaviour of children in the presence of their elders is often stressed in moralizing precepts. Small children went naked and wore their hair in a plaited lock on one side of the head, the rest of the scalp being shaved. Older girls wore a dress, and boys "put on the belt" (*i.e.*, assumed the loincloth) perhaps when they first went to school at the age of seven or thereabouts.

In the household of a wealthy nobleman there was a staff of servants and slaves. Some of the former, the butler, the wetnurse and the aide-de-camp or "follower," held positions of privilege in the family. But slaves, who were often Nubian or Asian prisoners of war, performed the more menial tasks; there is mention of foreign cooks and gardeners, doorkeepers and sweepers. The papyri make it clear that slaves were bought and sold, but also that they could own property and rent land on their own account and that a slave could be freed either by a declaration of manumission before witnesses or in some circumstances by adoption. Not all slaves were humanely treated, and there survive a number of references to runaway slaves and their treatment. Serfs on the big estates were bound to the land and could be transferred with the land to a new owner.

Domestic pets are often depicted in the scenes of family life. Hunting dogs of the greyhound or saluki type accompany their master to the chase, and a hunting cat is shown catching birds in the marshes. Lap dogs, cats and monkeys frequently sit under the chair of their master or mistress. The Nile goose was another favourite household pet; perhaps it earned its keep as a watchdog. Dwarfs and pygmies appear to have been especially prized in the Old Kingdom as dancers and jesters.

Houses and Furniture.—The houses of the earliest Egyptians, and those of the peasant at all times, must have been very similar to those of the fellahin today, namely a one-roomed shelter with walls of mud or of reed matting plastered with mud. The houses in the Neolithic settlement of Merimde-Beni Salaama, on the edge of the delta, were shallow oval pits scooped in the ground and rimmed with clods of earth, with a superstructure and roof of wattle or matting. In houses of the later predynastic period the roof was supported by wooden poles or bundles of reeds; the hut was rectangular in shape and sometimes provided with a sheltering wall in front of the door acting as a windbreak, the precursor of the courtyard. Models of houses with vaulted roofs and others with domes are known from the archaic period. The king's house in predynastic times was a more elaborate version of the first of these types, with an intricately paneled façade which was later reproduced in the brick walls of mastaba tombs and the stone false doors of Old Kingdom burials. (*See* also EGYPTIAN ARCHITECTURE.)

Pottery models of houses, serving as offering trays, and wooden models set in the tomb for the habitation of the dead owner give some idea of what the outside of a Middle Kingdom house looked like. There was now normally a courtyard inside the entrance gate, and a porch or pillared portico on the front of the houses. Usually the house contained several rooms; an external staircase led to the flat mud roof, which often had on it the corn bin and the bread oven and was a cool place for sleeping on hot summer nights. The walled town, Hetep Senusret, at Al Lahun contained houses both large and small, crowded together in the narrow streets. The houses of the chief officials were mansions of some dozens of rooms, while the labourers had two-room tenement dwellings, back to back.

Not very much is known of the town houses of the New Empire. Tomb paintings show that in a crowded metropolis they were several stories high. In the spacious suburbs of Akhetaton (Tell el-Amarna; *q.v.*), and elsewhere on their country estates, the nobles lived in large villas set in gardens planted with trees and shrubs, with a pavilion and perhaps an ornamental pool. The whole estate was surrounded by a high wall. The house usually had an entrance lobby and a large central reception hall or drawing room, its roof carried on pillars to a greater height than the roofs of

the surrounding rooms, so that it could be lighted by clerestory windows with stone or wooden gratings. The living rooms lay behind and around, and included bathrooms with plaster floors and a lavatory with a stone seat resting on brick containers filled with sand. The kitchens and servants' quarters, storage bins and granaries, kennels and stables were in outbuildings behind the house. Most houses obtained drinking water from their own wells.

Few royal palaces have been excavated. Usually, however grandiose, they were mud-brick structures not intended to survive, like the stone-hewn temples, to eternity. The palace of Ramses II at Qantir, in the delta, was decorated with faïence tiles bearing figures of bound captives. So also was the small palace of Ramses III outside his mortuary temple at Medinet Habu in western Thebes. The king's bedroom and bathroom and the queen's apartments were grouped around a pillared throne room from which the Pharaoh could step out onto a richly decorated balcony, supported by papyrus columns, when he made one of his public appearances. A similar balcony, called the "window of appearing," is shown on reliefs at Tell el-Amarna. There it was on a bridge which, spanning the main boulevard, linked the private royal residence with the official palace and administrative buildings. Enough remains of these buildings to suggest the richness of the ornamentation. Ceilings and pillars were painted in brilliant colours. The floor of the colonnaded hall represented a lagoon in which fish swam and water lilies floated; around it wild duck flew up from clumps of papyrus and calves gamboled. The stucco walls were painted with scenes of intimate family life of the king and queen and the little princesses, who sit on their laps or play at their feet.

The Egyptian house was not crowded with furniture. Stools and chairs were used by the wealthy, but the poor were used to sitting on the floor. Beds were likewise a luxury of the rich; simpler folk lay on the ground, on a mat or pile of linen, or sometimes on a brick dais at the end of the room. A curved headrest of wood, pottery or stone, supported on a framework or solid pillar, served as a pillow; everyone had his own, made to measure. Stools had seats of leather or plaited rush or cord. Chairs often had a curved wooden seat and a high, carved or inlaid back. A child's chair from the tomb of Queen Tiy's parents had a cushion of white and pink, stuffed with pigeon's feathers. Collapsible stools and folding camp beds with hinges have been found. Often the legs of chairs and couches terminated in a foot shaped like a lion's paw or a bull's hoof.

Clothes, linen sheets and toilet articles were stored in chests with flat or gabled lids, or in large palm-fibre hampers. Tables were rare, though small stands or occasional tables stood by guests at meals. Pot stands held the tall wine jars and water pots and flower vases filled with lotus blossoms, and a ewer and basin stood ready for hand washing between courses. Pottery braziers and fire curbs have been found, and oil lamps of the wick-in-saucer type.

Dress and Jewelry.—The typical costume of an ancient Egyptian of the Old Kingdom was a linen kilt or loincloth wrapped around the hips and tucked into the belt. The kilt could either end above the knees or be long enough to reach almost to the ankles. The overlapping flap of the short kilt was often pleated and sometimes coloured yellow. Sometimes the whole kilt was goffered and both ends tucked into the girdle so that a flap hung down in front. In statuettes and on tomb reliefs the long kilt often stands stiffly forward as if starched. In the Middle Kingdom, officials are often shown wearing a fringed mantle, possibly of wool, which completely covers the body leaving only the hands peeping out where the edges of the garment cross in front.

The classic woman's garment was a long, sheathlike linen dress extending from just above or just below the breasts to the ankle and supported by wide shoulder straps. Sometimes these robes were white, sometimes brightly coloured or patterned. Men wore on their bare chests a wide semicircular bead collar, women wore a bead choker, bracelets and anklets. Both sexes cut their hair short and wore wigs, either cap-shaped with rows of short locks or else hanging almost to the shoulders in a long bob, swept back to show the ears. In the Middle Kingdom a fuller type of wig became popular, the masculine version falling to the shoulders and dipping to a point in front on either side, while women's wigs were cut off squarely over the breasts or parted at the nape of the neck and brought forward over the shoulders in two thick plaits, their ends curled round ornamental disks.

In the New Empire and later, the dress of the wealthy of both sexes became much more voluminous and elaborate; fashion changed from decade to decade. The trend toward frivolity in costume started in the middle of the 18th dynasty. On ceremonial occasions men now wore a loose pleated robe of fine linen over an undershirt. The robe covered the whole body, had wide sleeves and was girded round the hips with a pleated scarf which flared out fanwise at the ends. Sometimes a transparent rippled shawl was worn over the shoulders. Women wore a similar robe, sometimes with a fringed mantle over one shoulder, or a full pleated dress knotted in front under the breasts. Wigs became very elaborate, falling over the shoulders in cascades of curls and plaits. Both sexes wore jewelry, the men a pectoral and bracelets, the women wide flower necklaces and rows of bracelets and anklets, with often a garland of lotus blossoms or a tiara of precious stones set atop the wig; at banquets the whole coiffure was crowned by a cone of scented pomade. Sandals of plaited fibre or tooled leather were worn "for best"; often when walking abroad they were carried. (*See* also DRESS.).

Food and Drink.—The staple diet of the Egyptian peasant must have been as simple as that of his modern counterpart: bread, beans and perhaps a salad of onions or papyrus stalks. Provision lists of temples and scenes in the tombs of offering bearers and of the preparation of food show that the better-off enjoyed a varied and ample menu. All sorts of vegetables and fruit were grown, including onions, leeks, garlic, radishes, watermelons, cucumbers, lentils, peas and beans. Lettuce and young papyrus stalks were a favourite salad. Plenty of fruit was included—figs, grapes and dates, the fruits of the persea tree and the dom-palm nut. Meat was an important item in the diet, and a variety of animals was reared for human consumption. Besides cattle of various breeds, game animals were stuffed for the gourmet's table, and in Old Kingdom tombs is depicted the forcible feeding of oxen and calves, antelopes and hyenas with loaves of bread and even with roast duck; cranes, ducks and geese were similarly stuffed with pellets of dough to fatten them. Fish were often split and dried. Milk and milk products were highly prized. Bread and cakes in a wide variety of sizes and shapes were eaten; all the activities of the bakery staff are illustrated in the tombs, from the grinding of flour to the removal of the little loaves from their conical molds. Fried cakes sweetened with honey must have resembled doughnuts.

Rich and poor alike drank beer. The successive processes of brewing are depicted: baking the coarse barley loaves, crumbling them in a vat with water, stirring and straining the fermented liquid, and the final bottling in jars closed with a lump of mud. Wine, the drink of the wealthy, was stored in large pottery amphorae labeled in ink; the docket states the vintage year and the vineyard of origin. Wine from the palace of Ikhnaton is classified by sealings on the mud stoppers as "good wine," "very good wine" and "very, very good wine."

At mealtimes the family seems to have eaten singly or in pairs, seated at small tables piled with food and waited on by servants. Children sat on cushions or on the floor. All ate with their fingers; Queen Nefertiti is shown demolishing a roast duck while her husband gnaws beef off the bone. Banquets are depicted in New Kingdom tombs; the company is entertained by a small orchestra or by dancers and acrobats. Young serving girls, clad in nothing but necklaces, minister to the guests' wants, replenishing their goblets and distributing flowers and scented headcones. Men and women tended to sit apart in the gatherings, the women chattering freely together; in the 19th dynasty married couples are depicted sitting side by side, with a considerable gain in decorum.

Games, Music and Dancing.—Both adults and children had their favourite games and are depicted playing them. Several board games are known. One was played on a circular board marked in a spiral. Another, called *senet,* was played with 12 or

14 ivory pieces on an oblong chessboard; moves were determined by throws of dice, knucklebones or marked wands. Another board game, rather like ludo, was played with dog-headed ebony pins inserted in holes. Children and adults alike threw balls of leather stuffed with chaff. Whipping top and tipcat were favourite children's games, and a number of games resembling "touch last" and "scissors-cut-paper" are shown in progress. Athletic games included pickaback and leapfrog. Various children's toys have been found; the commonest are little animals modeled in clay and crude dolls of wood or clay. Pottery rattles amused the babies. Boys used miniature weapons for their mock battles. Occasionally more elaborate toys had movable parts manipulated by strings.

Music and dancing were favourite diversions of the ancient Egyptians. Musical instruments buried in the tomb with the dead man ensured his entertainment in the next world, and reliefs depicting dancers may sometimes have that intention, though many of them represent an episode in the funerary ritual (see *Funeral Customs* below). Dancing was performed by young people of both sexes and was accompanied by rhythmic clapping and by singing. Mereruka (6th dynasty) is depicted in his tomb seated on a couch, listening to his wife singing; she accompanies herself on the harp. Blind harpists at the funeral feast sang of the fleeting joys of earthly existence and the imminence of death. Drums, tambourines and flutes are shown in the Old and Middle Kingdoms. In the New Empire, a number of new musical instruments were introduced from Asia, including the lute, the lyre and a larger, more elaborate form of harp with a long, curved neck and elongated sounding box. The sistrum, a loop-shaped rattle, was used in the worship of the goddess Hathor. Trumpets were carried to rally the army, and actual specimens, still playable, were found in the tomb of Tutankhamen. The ancient Egyptians seem to have had no system of musical notation and almost nothing is known of their music. Small orchestras of several different instruments are shown; a combination, such as lute, double flute and harp, suggests some form of harmony.

3. Administration.—*Pharaoh and His Court.*—The institution of kingship in ancient Egypt was centred around the concept, difficult of comprehension today, that Pharaoh was himself divine, a god among the gods, and thus on a plane above mankind. As the physical son of the sun-god Re, he was referred to as "the good God." He wore insignia that emphasized his divinity and he lived in constant communion with his fellow gods, performing in person their daily ritual in the temple, conversing with them in dreams and interpreting their will to his subjects. He was in fact the link between gods and men. The well-being of the king was therefore essential to the welfare of Egypt and its people. His daily routine was hedged with ritual, his person sacrosanct and his will supreme. Government was by royal decree. The king's courtiers ministered to his wants and in the 1st dynasty, though perhaps no later, may have been buried alive at his funeral to accompany him to the next world. Nobles counted it as the highest honour to be called "king's friend" or "sole companion."

The king's divinity found its most remarkable outward expression in the massive pyramid whereon he mounted to heaven at his final apotheosis. The building of a pyramid implies command of vast resources in manpower and materials. But such a monument stressed not only the might of the Pharaoh but also his mortality. Civil strife, the impoverishment of the crown and the enforced delegation of authority must have served gradually to diminish his stature and bridge the gulf that separated him from his subjects. Men dared to rob the royal tombs and even to lay violent hands on the living person of the king; Amenemhet I was one of those who suffered at the hands of an assassin. By the New Empire, the status of the king had changed; he was still god incarnate (indeed the role of Amon-Re as his corporeal father is stressed in 18th-dynasty temples) but he was also a mighty man, a warrior who, aided by the gods, went forth to war at the head of his troops. One or two kings may even have been killed in battle. Princes were trained in the military skills and lived a soldier's life, no longer hedged by tabus. They made demonstration of their physical bravery in battle and of their prowess in hunting elephant and lion or in public displays of archery or chariotry. It was at this time that the term Pharaoh, meaning the "great house (palace)," came to be used when talking of the king.

Officials.—The complexity and all-embracing nature of the king's official duties as head of all branches of the administration, as well as of the priesthood and the army, forced him at an early period to delegate his powers. At first posts were filled by princes of the blood royal, then the circle was extended to include the king's companions; in the Old Kingdom, there was a limited nobility from which the king chose his officials. By the New Empire, though there was a large number of noble families whose members might expect to fill high office, the king frequently promoted men of quite humble birth. An example is the famous architect and adviser of King Amenhotep III, Amenhotep, the son of Hapu, who was no more than a humble scribe till the age of 50 when he began to receive rapid promotion from the king. All offices were derived from royal appointment, and in theory even minor officials owed their positions to royal patronage. In practice office tended to become hereditary; young men inherited their fathers' offices and the position was normally ratified as a matter of course by the administration.

Next to the Pharaoh himself, the most important men in the state during the New Empire were the vizier and the high priest of Amon (see *Priesthood* below). So much of the burden of administration fell on the vizier's shoulders that the office was duplicated in the New Empire and possibly earlier. It seems that almost every branch of the state administration within Egypt came under the vizier's jurisdiction. The New Empire vizier was "the support of the whole land." In his office were filed records of every department of the government, private wills and contracts, and royal decrees. He tried cases, heard appeals and appointed magistrates. He was the intermediary between local authorities and the king, and had to receive their reports and transmit to them the king's commands. He controlled irrigation and agriculture, the assessment and collection of taxes, the reception of foreign tribute and the entertainment of foreign envoys. He was responsible for the upkeep of roads and buildings, the police force and the army garrisons in Egypt. With such wide responsibilities, a very large staff was necessary. Every day he was expected to report in person to the king, together with the treasurer whose duty it was to administer the state revenues and see to their storage and distribution. Other high officials were the overseer of works or chief architect, in charge of all new building enterprises, the quarrying and transport of building materials and the maintenance of communications; the royal chamberlain and his staff, who looked after the royal household; the royal herald, who had a special duty of liaison with the army abroad and the collection of foreign tribute; the army generals; and holders of various personal offices at court such as that of the tutors to the royal children, an honorary post given often to retired army officers of rank.

In the Middle Kingdom and earlier, the administration was far less highly centralized and the country was run largely by the district governors or nomarchs. These local officials, each with a large staff, levied taxes, administered justice, maintained an armed force and could call out the *corvée* in their nomes. The division into nomes may go back to prehistoric times, though this is not certain. The system, in force at least in the later part of the Old Kingdom, persisted with modifications and additions until the end. In late times there were 42 nomes, 22 in upper and 20 in lower Egypt. In Ptolemaic times a *heptanomis* of seven nomes was formed in middle Egypt. In upper Egypt the Theban nome took a leading position as the metropolitan nome. The Nile valley south of Ombos was sometimes regarded as one with the province of Nubia, though in the Old Kingdom Elephantine (Jazirat Aswan) was regarded as the "door of the south." In Ptolemaic and Roman times Nubian territory from Syene to Hierasykinos (Muharraqah) was called the Dodekaschoinos and was regarded as part of Egyptian territory, the country farther south being part of the Meroitic kingdom. The whole land of Egypt was known as Keme, the "black land." The desert was the "red land." The Greek name Aigyptos probably derives from one of the names of the city of Memphis, namely Hikuptah, the "house of the soul of Ptah."

The conquest of Nubia in the New Empire led to its organization under a viceroy, called the "king's son of Cush." The viceroy, responsible only to the king, had two deputies and a large staff. The colonial governors in the Asian empire of the New Empire Pharaohs are called "overseer of foreign countries." They had a somewhat different function, since in the city-states of Palestine and Syria vassal princes were allowed to rule, so long as they paid tribute and obeyed their treaty obligations.

Justice.—No Egyptian code of laws has been preserved. References made in the papyri to "the law relating to" such-and-such an offense indicates that records were kept of royal decrees and court decisions. Several Pharaohs had a reputation as lawgivers among the Greeks; by Ptolemaic times a body of law had grown up, known as "native law" and quoted alongside the Greek and Roman, stemming probably from kings' decisions. Several royal decrees, dealing with abuses and their reform, have survived.

The law was administered by a court (*qenbet*). For civil offenses priests and private individuals might sit as magistrates, but for serious crimes the king might appoint special commissioners as judges. One instance of this was the series of trials in which court members and the royal harem were accused of high treason; extant papyri relating to this case contain detailed depositions of witnesses and an account of the judgments. Other documents, valuable for knowledge of Egyptian legal practice, contain the proceedings of a special tribunal set up to investigate tomb robbery in the royal necropolis during the reign of Ramses IX. Punishments were mutilation and beating, or, for more serious crimes, penal servitude or execution.

Some juridical papyri have survived, including contracts, deeds of gift, wills, marriage settlements and an act of adoption. The ancient Egyptians were methodical and they liked to preserve records in writing. But the processes of law were sometimes cumbersome and a dispute might be settled by the simpler method of consulting the local oracle. The god's statue would be asked to decide between the litigants, and to indicate his decision by a movement to one side or the other.

Many late Pharaonic documents in demotic deal with legal matters, in formulas used until Greco-Roman times. The Macedonian Ptolemies introduced Greek law for the Greek settlers. The Egyptians retained their own laws. A "royal law" binding on all was introduced concerning revenue and public order. The king remained supreme judge though most legal matters were treated in courts. Diodorus mentions an ancient court of 30 judges; the *laokrites* were a civil court for Egyptians; the circuit court of the *chrematistae* held sessions for Greeks, in the chief cities; the *koinedokion* dealt with civil suits between Greeks and Egyptians. Royal inspectors supervised these courts; military courts, priestly tribunals and the tribunals of associations were outside state control. The native law of Egypt slowly merged with Greek law, and Roman imperial law reacted with both. Greeks and Egyptians became equal before the new combined "Law of the Egyptians." But the Greeks of the autonomous cities Alexandria, Ptolemais and Naukratis retained Greek laws; Roman citizens were bound by imperial law. An adapted Roman provincial law also grew up. By A.D. 250 imperial law prevailed; legal documents in demotic ceased.

State Revenues.—These were derived from three main sources: the royal monopolies, which probably included booty, tribute and foreign trade; the crown estates; and taxation. From his huge income, the king had to maintain the palace and its staff, the whole bureaucracy, the army and conscript and slave labour; to furnish the gods with offerings, build temples, provide for the upkeep of dikes and canals and reward faithful service. The treasury was filled with the goods acquired as the result of the exercise of royal monopolies. Quarries and mines, property of the crown, were opened by official expeditions and sealed again on departure. Trading expeditions were similarly controlled, and frontier posts, heavily fortified and garrisoned, kept out unauthorized immigrants and ensured that imported goods went to the royal treasury (see *Transport and Trade* below). A few natural products grown in Egypt may also have been crown monopolies. Campaign booty generally took the form of prisoners and captured equipment, and also livestock; tribute was levied on the conquered, and annual consignments of goods, slaves and many kinds of natural produce were sent to Egypt. The king and the temples owned vast estates. In accordance with the doctrine of divine kingship the whole land in theory belonged to the Pharaoh, who was responsible for its fertility. It is probable that although land was continually donated by the king to his officials as recompense for their loyalty, and enormous estates were given to the temples, yet the theory of royal ownership was never altogether allowed to fall into oblivion; and under a strong Pharaoh land could be recalled into state control or even revert to the crown at the expiration of its term of ownership. In Saite times an official was rewarded by a land grant from the king which was registered as temple property, though he enjoyed the revenue from it. By this ingenious means the Saites avoided the danger of large feudal landowners becoming independent of authority.

The crown estates are frequently mentioned in papyri, on jar dockets and in letters. They included arable land, vineyards, market gardens and papyrus swamps in which fishing and fowling were profitable industries. Each estate had its own staff; some were worked by slave labour and others let as smallholdings. Tenant farmers paid the harvest tax, of perhaps one-seventh of the yield. It is possible that temple estates were also regarded as crown property. It was the duty of every Pharaoh to make offerings and gifts to the gods in order to secure their good will, and the inventory of the great Harris papyrus offers proof of the huge size of temple estates in the New Empire. These estates were not normally exempt from taxation, and like the crown lands could be let out to tenants.

Land grants from the king as a reward for services are often mentioned. Land in private ownership could be transferred by sale or inheritance. On attaining high office, a man acquired an official estate, the perquisite of that office and inalienable. Each city had its tax office, and among the fiscal staff were those concerned with the cadastral survey. When the corn began to ripen, the fields were measured by the "holder of the cord" and the "stretcher of the cord," and on the data obtained an assessment would be made, of bushels or sacks of corn per *aroura* (100 square cubits, about $\frac{2}{3}$ ac.). The Wilbour papyrus is a document of this sort, although the abbreviated notations and the obscurity of some technical terms render it difficult to understand. The collection of the corn tax at harvest time was controlled by the "scribe of fields," who received the farmers' loads at a threshing floor on the river bank; the corn was measured, loaded onto barges and taken to the state granaries. Owners of livestock also paid a proportion of their herds as tax, on the basis of an annual cattle census.

In the matter of perishable foods the expenditure of the state revenues was closely linked with their collection. All officials in state employ received salaries in the form of a daily ration of food and drink, graded according to rank and responsibilities. The ration was normally distributed from local administrative centres as loaves and cakes, cuts of meat and jugs of beer. The state maintained in each centre a factory containing slaughterhouse, mill, brewery, kitchens and bakehouses where the proceeds of the fiscal collection, wheat, vegetables, fruit and livestock, were prepared for consumption. Special bonuses were paid to officials on completion of missions, to courtiers during festivals or to soldiers as incentives. In the case of temple revenues, foodstuffs were first prepared in the temple kitchens as offerings to temple deities, and then removed from the altars and distributed as daily rations to the temple staff.

4. Army.—Unlike the Assyrians, the Hittites and other nations of antiquity, the Egyptians were not a warlike people. Economically self-sufficient, and preoccupied with the enjoyment of the fruits of their industry, their ideal was to live peaceably. The country had desert on either side and was not constantly harried by hostile forces on its northern and southern borders. In the Old Kingdom there was no large standing army. The king had his bodyguard, and a larger force could be recruited locally as required for warlike forays or for quarrying, mining or trading enterprises.

In these the expeditionary force might have to fight or only to drag stone; in either case it would be disbanded when the task was completed. In times of unrest, when local chieftains disputed for the throne, there was civil war and bloodshed, but reunification again brought the tranquillity that men desired. The Second Intermediate period, however, saw the rule of foreigners so hated that a bitter war was fought to achieve their expulsion. Thereafter, a kind of military craze swept the country. The autobiographies of noblemen and royal inscriptions sound a new, martial note, bravery in the field is extolled and Pharaoh goes out to war, like the storm god himself, in his chariot to do battle at the head of his troops. This enthusiasm did not last. By the end of the dynasty the backbone of the army seems already to have been the mercenary troops, Nubians, Libyans and Sherden from the Aegean, and in spite of a brief revival when a national militia was raised to defend Egypt from the peril of the "sea peoples," it was henceforward a mercenary army upon which the Pharaohs relied. Herodotus tells of the Hermotybians and Kalaisirians, who, he says, were settled in separate colonies in the delta; the meaning of these words is obscure, but it is known that Psamtik I relied on Greek and Carian mercenaries and that there were military colonies in Daphnae and elsewhere, and a settlement of Jewish veterans at Elephantine.

Old Kingdom soldiers were armed with battle axes, lances, throwsticks or clubs, or with bows and arrows. Nubian soldiers were archers. Scaling ladders were used in the assault of a walled town, and the walls were mined by crowbars. Similar methods of attack were used in the 19th dynasty against Canaanite fortresses —the Egyptians never developed the use of heavy siege engines such as those devised by the Assyrians. Regular troops of the royal militia are represented in 5th-dynasty reliefs, and in the 6th dynasty a large expedition sent against the Bedouin of Palestine contained troops levied from every nome, each under the command of the local governor. In the Middle Kingdom broad daggers, spears and large round-topped shields of skin, stretched on a wooden framework, are shown. The unit seems to be a troop of 40 men. The nomarch raised his own force and might be asked to man an expedition sent to mines or quarries, or a sea-borne trading venture down the Red sea. Some 3,000 soldiers on one of these expeditions received two jugs of water and 20 rolls of bread as their daily ration. Wells and cisterns were dug en route and left guarded for use on the return journey.

After the Hyksos war, the army was expanded and reorganized under centralized control and with new weapons introduced from Asia: daggers with a round-tipped hilt, the khepesh (scimitar), the laminated composite bow and, most important of all, the war chariot, by this time a formidable weapon in all the armies of western Asia. There was as yet no cavalry; the horses were small and were ridden only by grooms. Chariots were light, beautifully made frames of bentwood, on large four-spoked (later six-spoked) wheels, with leather panels and a quiver for arrows suspended diagonally at the side. They were drawn by two horses and occupied by two men, a driver and a warrior. During the battle the charioteer stopped, looped the reins round his waist, and held a shield above the archer.

The infantry was divided into companies of 250 archers and spearmen, made up of five platoons of 50 men, each commanded by an officer. Until the time of Horemheb the regular army in Asia seems to have consisted of two divisions, probably of 2,000 men apiece; Seti I added a third, and Ramses II had four, as well as auxiliaries and special troops, with which to face the confederate Hittite army at Kadesh on the Orontes. Soldiers now wore defensive armour, helmets and breastplates of leather sewn with overlapping metal scales. In the Saite period, tunics of linen were sewn with crocodile skin. Greek metal armour was probably only worn by princes, and must have been almost intolerable in summer heat.

When fighting overseas the army employed ships for transport and supplies, but until the late New Empire there was no navy in the accepted sense of the word. Once or twice fighting on the Nile is mentioned, but in connection with a land action. The only sea battle illustrated is that fought by Ramses III against the sea pirates: the Egyptian warships are distinguished from those of the enemy by their curved hulls, lion-headed prows and bell-shaped crows'-nests, from which the lookout could direct the steersman through the melée of hand-to-hand fighting. Most of the enemy ships are capsized; they have square hulls and a bird's beak on the prow. War gave an impulse to shipbuilding, and records surviving from the archives of the royal dockyard at Memphis in the reign of Thutmose III give lists of stores supplied to the boat sheds. A large staff of military scribes was attached to the army, to register recruits, arrange for the transport of troops, attend to the commissariat and so forth. High-ranking officers had to be men of considerable ability, able to solve practical mathematical problems involving engineering tasks and to be conversant with the language and geography of the province within their command. They were often men of proved ability promoted from the ranks, rather than the sons of high officials.

5. Transport and Trade.—In ancient Egypt, as in rural Egypt to this day, the universal beast of burden was the donkey. Riding is not shown in tomb reliefs, but it must have been the general form of human transport and a nobleman is depicted in a litter on donkeyback. Corn was carried to the granaries in large paniers, and Bedouin in a painting in a Bani Hasan tomb carry their worldly goods on the backs of asses. The camel was not known as a pack animal in Egypt until Ptolemaic times, and the horse carried neither men nor goods. Wheeled carts drawn by oxen were occasionally used, but heavy loads such as blocks of stone were usually dragged on sledges to the river edge, the ground in front of the runners being kept slippery by watering. For long distance, water transport was preferred. Many kinds of river craft are depicted, as well as scenes of boatbuilding.

The simplest form of boat was a reed float, or light skiff made from a bundle of papyrus stems tied tightly at either end. On these the Egyptians went fishing and fowling in the marshes, and though they were easily upset by a hippopotamus or overturned by hostile boatmen in a scuffle, they were easily righted and could be carried on the back. Some of these reed boats were large enough to take several men and even carried a bipod mast and sail, the mast being set forward for balance. The Nile is swift and wide, and sail was necessary even for short trips; but since on the Nile the prevailing wind blows from the north, and the Egyptians could not tack, they had to use oars or paddles downstream.

When heavy cargoes or crops in bulk had to be carried, a larger vessel was needed. Probably as early as the 1st dynasty wooden ships were being built. A remarkable example of an actual 4th-dynasty boat was discovered in 1954 in its rock grave alongside the pyramid of Khufu at Giza and patiently reconstructed piece by piece. It is clinker-built, of long planks of cedar joined edge to edge by lashing, the rope in each case being passed several times round bars gouged in the inner surface of adjacent planks. There is no true keel and there are no ribs, the hull being strengthened against water pressure by transverse beams and the framework of the deck. A spacious two-roomed cabin, a canopy for the steersman, a set of paddles and two steering oars, which would have been lashed to upright posts in the stern in lieu of a rudder, completed the equipment. The boat is 143 ft. long and has a beam of 20 ft.

Large ocean-going vessels were in use in the Mediterranean from at least as early as the 3rd dynasty, trading in timber with the port of Byblos and known as "Byblos boats." They were given additional strength in the form of a hogging truss of double rope extending from stem to stern and tightened with a Spanish windlass. Similar vessels undertook the Red sea voyage to Punt (perhaps Somaliland). Whether Egyptian ships, hugging the coast, ever got as far as the Aegean or to the Persian gulf or the coast of India is not known, though there is slight evidence that they did. The golden mounts of a bed bearing the name of a 5th-dynasty Pharaoh have been found in a tumulus at Dorak, near the Sea of Marmara in northwest Turkey, and early Egyptian objects have been found in Crete. Certainly Sumerian vessels appear to have landed on the Red sea coast of Egypt in very early times. In the Middle Kingdom a vessel 150 ft. long, with a crew of 120 sailors, is mentioned. New Empire ships were longer still,

and a very wide sail, suspended transversely between very long spars and supported by ropes running from the boom to the masthead, replaced the square or oblong sail. The mast was now stepped amidships for easier maneuverability. Huge barges, especially built for the transport of large blocks of stone, had no sails. Hatshepsut's great barge carried two obelisks end to end on its deck, and was pulled from Aswan to Karnak by 30 rowing boats, each manned by 30 or 32 oarsmen. Royal barges, and the barks of the gods used in river processions during the annual festivals, must have been splendidly painted and gilded and hung with tapestries.

Coinage was not used in Egypt till the last native dynasty, and all trade was conducted on a barter basis, though the value of goods was usually stated in terms of weight of metal. Gold rings or deben weights of copper were mediums of exchange. Commercial documents concerned with the purchase of goods, including slaves, have been found, though they are not numerous. It is probable that trade was highly organized, though little is known of business methods, of banking, loans and contracts, such as the tablets of Babylonia in their thousands have preserved.

Foreign trade, at any rate in the earlier periods of Egyptian history, was a royal monopoly. Merchants went abroad only on the business of the king, and heavily guarded frontier posts controlled the entry of foreigners who came to barter their produce for the coveted manufactures of Egypt. Certain Sumerian artifacts have been found in a context of the later predynastic period in Egypt, and, although there is no reciprocal evidence of Egyptian objects found in Sumer, there must have been some kind of commercial contact, whether direct or indirect, between the two countries. From very early times, before history began, seafaring ships must have brought to the delta ports logs of coniferous timber such as cedar, pine and juniper from the Lebanon. The port for this trade was Byblos. Egyptian objects of early dynastic date have been found on the site of Byblos (biblical Gebal), and in the tombs of the 1st-dynasty kings at Saqqarah the roofs and floors were composed of great beams of pine 20 ft. long which cannot have come from trees grown in Egypt.

In the south, caravans led by the local governors of Elephantine, the most southerly Egyptian town, set out on journeys to the Sudan that lasted months and from which were brought back ebony and ivory, animal skins, precious stones and other treasures of inner Africa. Incense played an essential part in the temple ritual, and the aromatics trade was of great importance to the Egyptians; most highly prized were frankincense and myrrh, to obtain which caravans traveled through the Wadi Hammamat east of Coptos to the Red sea coast and thence took ship, generally building their own boats for the journey, down the coast to the "divine land" of Punt. The resins of terebinth (turpentine) and pine were much-used substitutes.

In the New Empire, Egypt became an imperial power and traded on an equal footing with other countries of western Asia. The diplomatic correspondence found at Tell el-Amarna contains many references to the exchange of gifts between potentates. The Kassite kings of Babylonia and the Hurrian kings of Mitanni sent richly carved and inlaid couches, chariots wrought with electrum, Asiatic copper and the prized lapis lazuli obtained by the Babylonians from Badakhshan in distant Afghanistan. Ostensibly these consignments were presents from one king to his "brother," but the donor in each case expected an equivalent gift from Egypt in terms of bullion. Western Asia at this time was greedy for gold. Its kings vied with each other in the decoration of their palaces, and they looked to Egypt as an inexhaustible source of supply. "In my brother's land," wrote one, "gold is as common as dust," and in that belief they demanded payment for their gifts.

Wall paintings in the tombs of New Empire viziers, whose charge it was to receive foreign envoys, depict the reception of tribute from the vassal states of Syria: elaborate vases of chased or embossed metal, carved ivories, ornamental weapons and other objects in the manufacture of which the Syrians had attained great skill, and exotic animals such as a bear and a baby elephant for the royal zoo. Several of the more elaborate pieces in the tomb of Tutankhamen are of northern origin. These countries also furnished the Pharaoh with many of the raw materials he required: olive oil, bitumen, honey, copper, wine, tin, lead and a little iron; cedarwood, deal, maple and boxwood for joinery, and elm, ash, oak and beech for chariot making. Nubian imports included tusks of ivory and logs of ebony, ostrich feathers and eggs, panther skins and giraffe tails, heavy ingot rings of gold from the mines of Cush, incense, hides, oxen and Negro slaves.

Other visitors to Egypt depicted in the viziers' tombs are Cretans with long, curled hair and patterned loincloths, bringing ambassadorial gifts from the court of Knossos or from Mycenae on the mainland: elegant bull rhytons, conical goblets and other *objets d'art* familiar to Aegean archaeologists, as well as metal ingots. It may be assumed that they did not go away empty-handed.

In the 7th century B.C., under the patronage of the Saite kings, Greek merchants began to market their wares in Egypt, and Naukratis in the delta, not far from Sais, was given a commercial monopoly. During the Saite and Persian periods Egyptian trade received encouragement: a canal was cut from the Nile to the Red sea near modern Al Isma'ilyah (Ismailia), the sea route to India was opened and Phoenician sailors, by command of Darius, achieved the circumnavigation of Africa.

6. Arts and Crafts.—*Paleolithic and Neolithic Periods.*—Flint implements of the *coup de poing* or hand-ax type are found in the gravel terraces bordering the Nile valley. Flakes chipped off in the process of shaping them provided scrapers, knives and borers. The Upper Paleolithic tools of northeastern Africa can be differentiated from those of Europe. Small saw-toothed flakes identified as sickle blades and crude grindstones suggest that Upper Paleolithic man reaped wild cereal crops and knew how to bake bread. Arrowheads of bone and flint indicate that he was still largely a hunter. In the long centuries, perhaps millenniums, of the New Stone Age, the next step further was taken: men turned agriculturalists, domesticated animals, grouped their dwellings in village communities and started fabricating pottery vessels in which they cooked and stored their food, weaving linen cloth and making mats and baskets. Stone weapons and implements were now well fashioned, the axes polished, long knife blades and arrowheads skilfully chipped and shaped. In Al Fayyum, *c.* 5000 B.C., Neolithic pottery was crudely shaped from coarse clay, but in the subsequent Tasian period (so called from Dayr Tasa in middle Egypt where the first Late Neolithic cemeteries were excavated) well-baked pots of polished reddish-brown clay (called black-topped ware because half the pot was carbonized by subjecting it to dense smoke) or beakers of black clay with incised decoration were buried with the dead. Other grave goods included alabaster and limestone palettes for grinding malachite and galena for eye paint, little ivory cosmetic spoons, bone and shell fish-hooks, small shell disk beads and ivory bracelets. In these simple artifacts are to be found the beginnings of Egyptian art.

Badarian Period.—Settlements and cemeteries dating to *c.* 4000 B.C. have been found around Al Badari, south of Dayr Tasa, which appear to represent a later stage of the Tasian culture. These again were an agricultural and pastoral people, but their skills were improving: the bowls and jars of black-topped ware are now very finely made, their thin walls decorated with a ripple burnish. Carved spoons and combs of ivory, and crude statuettes of women were placed in the grave, the latter probably to accompany the dead. Palettes were made of slate. A few copper beads and a copper awl found in a cemetery suggest that metalworking, though still rare, was not unknown to the Badarians. In jewelry they had made remarkable advances: various hard stones were carved and drilled to make beads, and they had mastered the elaborate technique of applying a vitreous blue-green glaze to steatite.

Amratian Period.—The predynastic Egyptians of *c.* 3600 B.C. lived much the same village life as their Badarian ancestors, but their crafts and techniques had advanced. Numerous settlements and cemeteries have been excavated. A wide variety of pottery vessels, all of them with a dark-red slip burnished by rubbing and many of them decorated with bold linear designs in white paint, have been found. Geometrical figures, crosshatched in imitation of basketry, and occasional figures of crocodiles, hippopotamuses or men are seen on this pottery of what used to be called the First

Predynastic period or, by German archaeologists, Negada I. The cutting of hard stone had developed and stone vases, usually cylindrical or barrel-shaped with a small foot, and disk-shaped maceheads of polished granite or breccia appear. Slate palettes have the shapes of animals or fish. A number of rather crudely modeled little statuettes of men, women and animals, probably again funerary, are found, but the ivory carvers' art was in advance of the potters' in this respect; one or two remarkable ivory figurines have been found, and ivory combs are decorated with figures of birds, ibex, ostrich and other creatures, all typically African motifs.

Gerzean Period.—The characteristics of the Amratian culture have been discovered only in upper Egypt, that is to say from Al Amra southward. On top of the level in which they are found lie remains characterized by quite different pottery wares, stone vase shapes and so on. These characteristics occur alone in the cemeteries of the Gerzean civilization in the latitude of Al Fayyum, and it is therefore assumed by many, though not all, archaeologists that the Gerzean culture represents a different evolution, due perhaps to closer contact with Asia, which developed independently in lower Egypt and perhaps had its centre in the delta. The bearers of this civilization, it is suggested, at some time in the 4th millennium spread southward and swallowed up the Amratians, imposing upon them their material culture and probably also their rule.

Characteristic of the Gerzean period is pottery of a buff colour with remarkable painted decorations in dark red paint. The themes are often scenic: the Nile is indicated by wavy lines and the desert hills by solid triangles; in the centre of the picture is usually a large boat with cabins or shrines and many oars, and flamingos and ibexes emphasize the topic of river and desert. Most of the boats have an emblem or standard and may have a cult or funerary significance. A remarkable advance was in the use of a tubular drill with abrasive for stonecutting. A great variety of materials was used, including hard stones such as basalt, diorite and schist, and characteristic shapes included large globular bowls with tubular handles by which they could be suspended. Maceheads were not disk- but pear-shaped, the form probably derived ultimately from Sumer. Flint knives are ripple-flaked, masterpieces of the knapper's art. Metallurgy had now advanced, and harpoons, spearheads, daggers and other implements and weapons are found, though not yet in large quantities. Metal was still a precious commodity.

The last period of Egyptian prehistory, *c.* 3400 B.C., saw yet further advances. Contact with western Asia may have inspired building in mud brick, the use of cylindrical roller seals, the use of certain styles of ornamentation and the first attempts at writing. The hieroglyphic script that was evolved as a result of the transmission of this idea, however, had little in common with the early pictographs of Sumer and nothing with the wedge signs into which these developed. On ivory objects such as combs and knife handles, and on large ceremonial slate palettes, lively figures of men and birds and animals were now carved in relief. Large commemorative maceheads and palettes found at Hierakonpolis are covered with sculptured scenes which seem to represent symbolically the acts of rulers of upper Egypt who, wearing the insignia of Pharaohs, triumph over their delta enemies or perform rituals. These may be regarded both as the earliest historical documents and also as the earliest form of hieroglyphic writing (*see* Egyptian Language).

Archaic Period.—The essential skills had now been acquired. With the unification of Egypt, perhaps by Narmer, who is depicted on his splendid state palette wearing crowns of both upper and lower Egypt, the Egyptians made rapid strides. With unprecedented resources in men and materials at their command, the kings of the 1st dynasty built at Abydos and Saqqarah large tombs of recess-paneled brickwork. The contents of these tombs are astonishing in their quantity, variety and virtuosity. Gold and copper were lavishly used, vases of great delicacy were carved from alabaster and schist, furniture was well-jointed and fitted with carved ivory bulls' legs. The typically Egyptian art of applying blue glaze to a granular quartz matrix, conventionally called Egyptian faïence, had already reached a high level. Pottery was now replaced by stone for fine vessels, but the introduction of the potter's wheel in the archaic period led to mass production of undecorated household wares. Leatherwork, ropemaking and textile weaving were highly developed crafts. The king was accompanied to the tomb by provisions of all kinds—tools and weapons in quantities, and ingots of copper to make more in the next world if the supply should run out, as well as bread, meat and great jars of wine for his refreshment. Timber was imported by sea from Syria for the great beams that lined and roofed his grave. The portcullis blocks in the entrances to some of these tombs show the finished skill of an experienced mason and bear witness that monoliths weighing up to 20 tons could be quarried and transported. The Egyptians of the 1st dynasty had a well-developed system of writing, both carved in stone and wood and written on papyrus, a mastery of carving in the most refractory materials, and a wide range of copper tools, from saws six feet long to the finest needles. Many of their products were of a high artistic order and many of the motifs which were to become characteristic of pharaonic painting and sculpture had already made their appearance.

Old Kingdom.—The sudden and extraordinary development in stone architecture which characterized the reign of Zoser is generally associated with the name of Imhotep, the king's chief architect, who bore among his titles of rank the simple words "sculptor, woodcarver." He was later regarded as the patron of artists and of science and learning, and it may not be fanciful to attribute to his genius various technical advances in stone and wood carving. The earliest known large royal statue is the magnificent seated statue of Zoser, his wig spread like a lion's mane over his shoulders. With the 4th dynasty begins the long series of private tombs decorated with reliefs, originally brightly painted, representing the dead man at home with his family, hunting or fishing, or watching the peasants on his estate performing their domestic or agricultural tasks. These reliefs, which depict with astonishing vigour and realism the daily life of the people of ancient Egypt, are a constant source of delight and instruction to those who study them, for the artist, though bound by conventions of style and proportion, often permits himself lively and unexpectedly unconventional details and depicts animals and birds with masterly realism.

Life-size statues of both royal and private personages were now carved for the tombs, in wood or in stone, and painted, the eyes being inlaid with quartz or crystal and obsidian, to give a life-like effect. Realism was essential. These statues were carved primarily in order to replace the dead man's body, if it should unhappily decay or be damaged in the tomb. The result is a wonderfully convincing series of portraits, even if the owner could afford only a statuette a foot high for himself and his wife. The decoration of temples with relief sculpture continued throughout the Old Kingdom. Hardly anything is left of the reliefs of the 4th-dynasty kings, but in the mortuary temples of the 5th and 6th dynasties are portions of beautifully carved historical and ritual scenes. Two life-size copper statues, found at Hierakonpolis, of King Pepi I and his son (6th dynasty) exemplify the skill of the metallurgist (see *Metalwork and Jewelry* below).

Middle Kingdom.—Toward the end of the Old Kingdom, art and craftsmanship declined as the direct result of a growing insecurity and weakening of the central government. A period of civil war ensued. The steles of the First Intermediate period reflect the low water into which Egypt's fortunes had fallen; they have none of the delicate carving and masterly design of Old Kingdom reliefs, the figures are thin and clumsily incised and there is little surface modeling. With the reunification of the country under the 11th-dynasty Thebans, art and craftsmanship began to revive rapidly. It reached its full flowering in the 12th dynasty, when a remarkable school of portrait sculpture arose, characterized by the vigour of its mastery over intractable materials and some originality of treatment. Among the most remarkable groups of royal portraits of the 12th dynasty are the heavily maned sphinxes found at Tanis in the delta, which have been plausibly attributed to Amenemhet III.

Relief sculpture is rare in the Middle Kingdom and is mostly confined to the walls of temples. The workmanship is of a very

high quality and the hieroglyphs beautifully designed and spaced. The nobles' tombs, which in the Old Kingdom were carved in relief, still contain scenes of daily life and of funerary rituals, but these are now sometimes painted on the stucco or limestone surface of the wall. The art of the lapidary reached its apogee in the 12th dynasty. Jewelry is minutely executed in gold, sometimes with filigree or finely granulated ornament. The jewelry from the princesses' tombs at Dahshur and Illahun included pectorals of cloisonné inlay in carnelian, green feldspar and the rare lapis lazuli, and delicate diadems of twisted gold wire with tiny gold flowers. Scarabs and beads are beautifully carved in amethyst, carnelian, crystal or glazed steatite, and small cosmetic jars are carved from alabaster or the friable obsidian (volcanic glass). Small figurines of animals and little bowls, molded in faïence and decorated in dark paint with patterns of lotus, were popular at this period.

New Empire.—A second deterioration in the political situation, leading to divided control and eventually to the domination of foreigners, set a period to the prosperity of the Middle Kingdom. Art again declined under the 13th dynasty, and as far as is known, no great masterpieces were produced during the whole of the Second Intermediate period. There are, however, indications that under the Hyksos kings, whose monuments were obliterated with ruthless thoroughness by their successors, Egyptian craftsmen continued to find royal patrons. It is plain that the expulsion of these strangers and the enthusiasm which brought Asiatic conquests in its wake gave new impetus to the arts. The finest products of Syrian and Aegean craftsmanship came to the Egyptian court, and the knowledge of new techniques, brought by captive artisans, was quickly absorbed. Exotic musical instruments, parade weapons of metal inlaid upon metal, vases elaborately gadrooned or embossed, vessels like lotus flowers carved in alabaster with transparently thin walls, light chariots with panels of painted leather were now made by Egyptian craftsmen for their royal and noble masters. The richness, indeed ostentation, of the paraphernalia of court life can only be glimpsed in the objects buried in the tomb of the young Pharaoh Tutankhamen. Not only shrines, vessels, statues and models used in the funerary cult but also the things he used in life were buried in this tomb: his footstools, thrones and beds, his clothes and jewelry, his chariots and weapons, his walking sticks and his fans. Modern craftsmen who have studied these objects marvel at the accuracy and intricacy of their manufacture. The techniques of the 12th dynasty are employed with far richer and more elaborate effect. The taste however is not always impeccable, decoration often mars or obscures the outline, and the design of some of the more elaborate pieces borders on vulgarity.

A desire to break away from the restraints of convention was responsible, at just this period, for that remarkable episode or interlude in art, politics and religion which is called the Amarna period. The sculpture and wall painting of the earlier part of the 18th dynasty maintain the strict canon of proportion laid down in the Old Kingdom and little modified through the centuries. Figures could be portrayed only in certain stereotyped poses. A pose having been selected, the block or wall surface was squared out in preordained proportions and carved according to precedent. Only the features might sometimes, not always, betray the traits of personality, so that in spite of the conventional stance, left foot advanced, shoulders squared and hands clenched at the thighs, it is occasionally possible to glimpse the individual. In the brilliantly painted scenes that decorated noblemen's tombs at Thebes and elsewhere, and on the walls of New Empire temples, the stereotyped portrayal of the human figure—head in profile, shoulders squared to the front, hips and legs in profile—was not departed from except in the case of the figures of slaves, peasants and foreigners. But the sculptors who worked at Ikhnaton's court, inspired perhaps by the king's new slogan, "living in truth," broke away from some of these conventions and portrayed the king and queen and their children in informal attitudes and situations. For the first time, Egyptian royalties are human. Ikhnaton offers his wife a flower; he kisses his baby daughter, while two little princesses play on Nefertiti's lap. The family's physical peculiarities, elongated skull, pointed chin, thin neck and limbs and heavy thighs were reproduced without flattery, indeed exaggerated to the point of caricature in the earlier experiments. With experience came moderation, and many of the later examples of the Amarna studios are among the most admired products of Egyptian art; *e.g.*, the celebrated limestone head of Queen Nefertiti in Berlin. Motifs taken from nature, from flowers, plants and birds, are used in frescoes, in jewelry and in the shapes of toilet articles and the decoration of pottery. Polychrome glass vases appear and polychrome faïence in bright shades of blue, purple, yellow, chocolate and apple-green.

The Ramessides and the Decadence.—Under the kings of the 19th dynasty, vulgarity came to the fore. After Seti I's beautiful work at Abydos, architecture became grandiose and clumsy, Ramses II's taste for the colossal produced the Rameseum, Abu Simbel and the hypostyle hall in the temple of Karnak. Few Ramesside statues match the work of the 18th dynasty, and relief sculpture, now deeply sunk cavo-relievo, lost delicacy, though the vigour and freedom of the battle scenes, in which the king is depicted slaying hordes of his enemies, is unsurpassed. Craftsmen continued to carve beautiful small objects in stone and wood, and polychromy was still favoured. Asian influence became marked, and motifs were borrowed from Syrian art, just as Semitic ideas and Semitic cults gained popularity among the mixed crowds that thronged the residence city in the delta. The deterioration continued in the 20th dynasty. Vulgar display concealed growing poverty of ideas; gold becomes gilding, and it is everywhere. The yellow-varnished coffins, with their very elaborate relief decorations in gesso showing mythological scenes and figures, are characteristic of this period. There is little sculpture, and the manufacture of small objects degenerates into carelessness.

Ethiopians and Saites: the Renaissance.—New vigour was breathed into the failing talent of Egypt's artists under the encouragement of the Napatan Pharaohs. Themselves enthusiasts for the great days of the past, they sought to revive the ancient skills and recapture the old inspiration, and a new school of sculptors arose at Memphis who took as their models the style of the Memphite reliefs and statues of the Old Kingdom. Officials of the 25th and 26th dynasties are portrayed wearing the archaic loincloth, and reliefs in their tombs are copied, sometimes figure for figure, from Old Kingdom reliefs. A noble named Ibi, who died during the reign of Psamtik I (*c.* 600 B.C.), copied scenes from the tomb of a namesake Ibi at Dayr al Jabrawi, 200 mi. farther north and nearly 2,000 years earlier in time. The workmanship of the Saite sculptors was impeccable. They achieved mastery of the hardest stones, imparting a wonderful finish to their lifelike portraits in basalt, schist and granite. Some of their work has great beauty. But something was lacking; in these neo-Memphite works there is a difference that does not escape the modern critical eye, and as time goes on this difference becomes accentuated. An artificial smile, perhaps influenced by archaic Greek sculpture, intrudes upon the native gravity, producing a curiously blank expression; the Memphite smile gives way to the Saite simper. It continued through the Ptolemaic period in Egypt and became characteristic of the royal portraiture of that age. In the 5th and 4th centuries Egyptian tomb reliefs and vase decorations exhibit increasingly the influence of Greek art. The art of the last native kings is still Saite but of a curiously delicate, refined nature that distinguishes it from all earlier periods.

7. Science, Medicine and Technical Processes.—*Astronomy and Time Measurement.*—The practical nature of science among the ancient Egyptians is exemplified by their attitude toward the heavenly bodies. The celestial world above them stimulated their imagination but produced myths no less fantastic than those of other early peoples. In no known Egyptian writing is there any speculation on the nature or size of the heavenly bodies or the prediction of any celestial event. The position of the stars was noted and they were arranged in constellations associated with various gods of the pantheon. Yet it was realized that the circuit of sun and moon, and the nightly risings of the stars, provided a means of measuring time. Within the civil calendar of the 30-day month, ten-day intervals or decades were recognized. The star,

called a decan, which appeared just before sunrise during a certain decade, would appear about 40 minutes earlier during the next decade and so on, so that each decade became represented by a different star. The consecutive risings of decans were used in the Old Kingdom to denote the "hours" or divisions of the night, approximately 12 covering the interval of darkness between dusk and dawn. On the ceilings of the tombs of Ramses VI and IX at Thebes are charts of the decans. Each division of the year has a representation of a seated figure, and for each hour of the night the position of the decan relative to that figure is given. The figure doubtless represents an observer. Observations were made by a priest on the temple roof holding a transit staff made of a slit palm leaf; his assistant stood some distance away holding a plumb line. The passage of the star across the plumb line was seen through the slit in the transit staff.

As the night was divided into 12 hours, so also was the day. Thus our 24-hour day derives from the primitive stargazing of the Egyptians. Day time was measured by the sun, through the use of an upright gnomon as shadow clock or sundial, or by water clocks, none of which appear to have been very accurate. The latter consisted of a stone bowl with a graduated scale on the side; water gradually ran out through a small hole in the base. Successive observations of the Dog Star (Sothis or Sirius) made possible the realization of a true or solar year. (*See* also CALENDAR: *Ancient Egyptian.*)

Mathematics.—In spite of the reputation the Egyptians had for wisdom among the Greeks, they do not seem to have made the same remarkable advances as did Babylonian scientists and mathematicians. Mathematics developed for strictly practical ends. In a country where landmarks are liable to be eliminated yearly by the Nile flood, geometry was a civil necessity. The amazing accuracy of the setting out of the Giza pyramids is proof that the Egyptians were skilled in linear measurement. The areas of triangular and trapezoidal fields could be calculated. Problems posed in the mathematical papyri, of which the Rhind papyrus is the most important, deal with everyday tasks of calculation: From a certain amount of grain, how many loaves can be baked? Given a ramp of length x and height y, how many bricks will be needed?

The notation was a simple decimal one. There were different signs for units, tens, hundreds, thousands and so on up to 1,000,000; 22 would be written $1 + 1 + 10 + 10$, and to write 999 required no fewer than 27 signs. Multiplication was reduced to a series of duplications—that is to say, the Egyptian could only double. Thus to find five times a number he could only double it, double it again thus arriving at four times, and add the number once, making it five. Division was the same process in reverse: to divide 13 by 4, one started with the divisor 4, doubled it and added another 4, thus demonstrating that 3 times 4 made 12: 13 subtract 12 is one, so the answer is $3\frac{1}{4}$. Unit fractions were expressed from early times, but with the exception of $\frac{2}{3}$ no fraction could be written that had not the numerator 1, and compound fractions were therefore set down in terms of added unit fractions; *e.g.*, $\frac{2}{2\frac{3}{4}}$ was written $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{1\frac{1}{2}}$.

In spite of the limitations imposed by this clumsy equipment, the ancient Egyptians found themselves able to solve most of the practical problems that came up for solution. They were familiar with the concepts of squaring and square roots. They correctly determined the area of a triangle and the volume of a truncated cone. In computing the area of a circle as the square of $\frac{7}{8}$ of the diameter they arrived at approximately the value of π. The mason measured the slope of a pyramid by its cotangent; *i.e.*, the vertical height was divided by half the side of the square base. He plotted the curve of an arch by offsets measured at regular intervals from a base line.

Medicine.—With their largely vegetarian diet and ample sunshine, the Egyptians might be expected to have been one of the healthiest nations of antiquity. They may indeed have been so, but pathological examination of mummies has shown that they, too, suffered from a diversity of complaints: water-borne diseases, then as now, took their toll, and fly-borne eye afflictions often led to blindness. Venereal diseases were unknown but there are known cases of tuberculosis, appendicitis, rheumatism and gout, hydrocephalus and clubfoot. Ramses V died, it seems, of smallpox.

Several medical papyri have survived, the most important of which are the Papyrus Ebers and the Edwin Smith Surgical papyrus. The former, chaotic in its arrangement, is a strange mixture of magical formulas and old wives' remedies. Besides prescriptions for curing various symptoms, it contains specifics for ridding the house of such pests as flies, rats and scorpions. The *materia medica* prescribed for illness are often unpleasant (*e.g.*, the blood of mice, and flies' excreta), but the beneficial effect of certain drugs had been noted, and it is clear that side by side with the traditional medico-magical treatment of disease had grown up a considerable body of scientific knowledge based on empiric therapeutics. The Edwin Smith papyrus is in a different category; it deals with surgery of the bones and outer tissues, working from the head downward. Cases are arranged systematically, under headings: name of complaint, examination, diagnosis, verdict. Notes are added on some of the cases, the exact nature of a lesion is stated and the cause is discussed. Egyptians themselves derived much of their learning in medicine from the first dynasties, and the practice of mummification must early have led to a detailed knowledge of human anatomy. The papyri show that the position and function of the stomach and intestines, the action of the larger blood vessels and the relation of the pulse to the heart were realized. Egyptian physicians and surgeons had a great reputation among the Greeks and other peoples of the ancient world, and specialists were sent from Egypt to treat the sick in the courts of distant potentates. Medical libraries were attached to most of the larger temples and Galen, Dioscorides and other Greek physicians mentioned drugs, prescriptions or methods of diagnosis they learned in the temples of Ptah or Imhotep at Memphis.

Quarrying and Stoneworking.—Egypt is particularly rich in building stones. Fine-grained limestone, found along the Nile valley from Giza to Isna, was the usual material for temple building till the 18th dynasty, after which the sandstone of the south came into use. Black and red granite, or syenite, came from the region of Aswan (Syene) at the First cataract. Yellow and red quartzite, a fine red sandstone called "wonderstone" by the Egyptians, diorite from the desert west of Tushka, alabaster (calcite) from the hills behind Amarna, basalt and graywacke (schist) were used for shrines, statues, obelisks and other purposes.

Much can be learned of quarrying methods from marks left in the quarries or on blocks left partially extracted. In a cliffside limestone or sandstone quarry a low gallery was cut near the top, high enough for men to work; on the stone floor under their feet the blocks were marked out, and narrow trenches driven between them with picks and chisels. The blocks were then separated horizontally by metal wedges driven in from the side. Hard stones may have been drilled or sawn with the help of an abrasive; wetted wooden wedges may have been used in certain circumstances. The surface was dressed by pounding with dolerite balls such as were found in the Aswan quarries. There also a giant obelisk lies half disengaged from its matrix; at a critical moment in the process of extraction it had flawed.

Once transported to their destination the blocks would be leveled top and bottom, moved up earth ramps to the wall face, and maneuvered into position. The final dressing, carving and painting of the wall face were executed from the top downward, as the earthen scaffolding was gradually removed.

Stone vases were hollowed out by means of a tubular drill or a bow drill with a flint bit. In addition to the hard stones mentioned above, ornamental stones such as breccia, porphyry, serpentine and steatite were used for pots large and small; red porphyry was quarried in the Mons Porphyrites (Jabal Dokhan) in the 1st dynasty, but not again till Roman times, when it provided the imperial purple of Byzantium.

Precious stones used by the lapidary included carnelian, jasper, green feldspar, turquoise, amethyst and the imported obsidian and lapis lazuli. Gems were valued for their colour, not their sparkle, and were not faceted.

Metalwork and Jewelry.—Copper ores occur in the Eastern desert of Egypt and in the peninsula of Sinai. Tools and other objects hammered from native copper are found in Egypt from the

Badarian period, and the smelting of ores soon followed. State-organized expeditions with armed escort went to Sinai, horizontal shafts were driven into the rock, following the veins of ore, and vast slag heaps near the mines show that it was smelted on the spot. The remarkable life-size copper statues of Pepi I and his son (*c.* 2300 B.C.) are thought to have been hammered over a wooden core and finished by chasing, a technique learned perhaps from Sumer; the parts were fixed to the core with copper nails. There is today no known source of tin in or near Egypt, and its absence may explain the late appearance of bronze there; it is first found in the Middle Kingdom (*c.* 1900 B.C.) and does not become common till *c.* 1500 B.C. It is not known whence the Egyptians obtained their tin; early examples of bronze may have resulted from the smelting of naturally mixed ores of copper and tin, rather than an artificial alloy. Tin was not in general use in Egypt till the 5th century B.C. Iron too made a relatively late appearance, with the exception of occasional worked meteorites. The tomb of Tutankhamen contained, as valuable rarities, an iron dagger with a gold hilt and crystal pommel, and an iron headrest. Once the superiority over bronze of "steeled" iron (*i.e.*, hardened and toughened by repeated hammering, heating and cooling) was realized, it became more generally used for tools and weapons, but it was not popular in Egypt till *c.* 600 B.C.

The earliest copper implements were cast in open molds, but later both closed molds with a sandy core and the cire-perdue process were used for casting elaborate objects such as statuettes, the limbs and appendages of which were affixed separately by riveting. The tomb of the vizier Rekhmire (*c.* 1450 B.C.) contains a painting illustrating the casting of bronze or copper doors for a temple. The metal is being melted in crucibles over an open fire with a forced draft from four skin-bag foot bellows; each crucible is then lifted by means of withies and the molten contents poured through one of a series of 17 funnels into the mold, which is oblong in outline and probably of terra cotta. The air vents are not shown.

Other techniques known to the ancient Egyptian metalworker were *repoussé* work, tracing and, in late times, engraving. Bronze was occasionally inlaid with other metals. Gold was soldered by the use of an alloy of silver or copper with gold. In the 12th dynasty this technique was used to produce minutely fine patterns in granulation, each tiny grain being separately sweated onto its background. Fine spirals and coils of gold wire were soldered in filigree designs. Cloisonné work, in which small pieces of faïence or semiprecious stones are enclosed in little open-fronted cells with narrow gold walls, also reached perfection in the 12th dynasty. This technique was used for the necklace and the vultures' outstretched wings on the innermost of Tutankhamen's magnificent gold coffins.

Carpentry and Cabinetmaking.—Wood can be rough-hewn with stone tools, but the development of metal tools and the importation of fine-grained foreign woods enabled fine woodwork to develop as early as the Archaic period (*c.* 3000 B.C.). Because of their rarity, great economy had to be exercised in the use of these woods, and small pieces were jointed by dovetailing or by mortise-and-tenon joints. Flaws were covered by skilful patching, and imperfections were concealed by plaster. Veneer was another method of effecting economy.

The timber was cut with an ax and planed with a small adz. For cutting planks a pull saw was used: the wood was lashed to an upright post and sawn downward. Holes were bored with the bow drill, which was in use from early times, but the lathe was scarcely known. Polishing with a sandstone rubber completed the finish. Parts of furniture were orig1nally lashed together with cord or leather thongs, but carpenters soon learned a variety of joinery techniques, including the use of dowels or split pegs of hardwood, and sometimes bronze nails, though these were rarer. Ingenious methods of hinging and fastening the lids of boxes were evolved. Gaming pieces of ivory were often kept in a box with a sliding lid; other lids were fastened with a cord round two knobs, which could be sealed, and occasionally a box is found with a turning knob operating a simple locking device.

Inlay was glued to its base; the use of several coloured woods or of ebony and ivory together was popular in the New Empire. A fragment of a 3rd-dynasty coffin was of six-ply, each layer pegged so that it lay across the grain of the piece below. In the New Empire curved pieces of wood were increasingly used in furniture making; the wood was either bent by heat to the required shape or specially grown for the purpose. The light chariots of the period are masterpieces of joinery; in the tomb paintings wheelwrights are seen at work on the spoked wheels, and in one surviving chariot the felloes are each of a single piece of wood, bent into a circle. Reliefs also show the making of elaborate shrines decorated with carved openwork, gilt and painted, and of carved statues and statuettes in ebony, yew and other hard woods.

Textiles and Basketry.—A fragment of coarsely woven linen has been found in a Neolithic settlement in Al Fayyum. Flax remained the usual material for woven textiles throughout Egyptian history: cotton is not found in Egypt till the 5th century B.C., though the Assyrians grew it earlier. A simple horizontal loom is represented on a Badarian dish; in the Middle Kingdom, women weavers are depicted at work on a similar loom, and in a tomb model of a weaving shop from Dayr al Bahri all the stages of weaving from the preparation of the rove are shown. Spinning and warping are shown on many tomb paintings. The vertical two-beamed loom appears in the 18th dynasty; two men sit abreast on a bench before it, and the cloth appears to have been a yard or more wide.

Examples of patterned textiles, fringed cloths, looped fabrics of the turkish towel type, and embroidered fabrics have been found. The richest collection of Egyptian textiles comes from the tomb of Tutankhamen, in which there were coloured robes in tapestry weave, embroidered tunics, patterned braids and plain bleached linen of remarkably fine weave. A limited number of vegetable dyes—safflower and indigo, madder and henna—were used. Examples of twined, coiled and plaited basketry have been found. Elaborate basket weaves were developed in the New Empire, when not only mats but sandals, bags, baskets and sieves were made of papyrus, reed or palm. Large coiled baskets have patterns in coloured checker work not unlike those made in Nubia today.

8. Literature and Mythology.—Egyptian literature has survived largely in the form of long texts in the hieratic or demotic script on papyrus or on ostraca (bits of pottery or limestone bearing inscriptions). Some of these literary writings may be classified as romances of a popular kind, often with a historical or mythological background; others are poems, either secular or religious; yet others are collections of moral precepts; and a few are model letters. Many of them evidently served as school texts, to be copied and recopied by scribes, and of these there is often more than one version, usually in a fragmentary condition, so that passages from one may serve to fill lacunae in another. It must be remembered that what chance has preserved can be only a fraction of the whole body of Egyptian literature.

Among popular stories that have survived may be mentioned that contained in the Papyrus Westcar, the "Story of King Khufu and the Magician," which tells a lively story of miraculous happenings at the court of the monarch as a prelude to the no less miraculous birth of triplets to a priestess of Re who were destined by prophecy to become future kings. The story is probably political propaganda, but it contains the essential ingredients of a fairy story and probably circulated as such. The "Story of the Shipwrecked Sailor," about a sailor who found himself a castaway on an island in the Red sea near the fabled land of Punt and was entertained by a monster serpent with eyebrows of lapis lazuli; the "Tale of the Two Brothers," which starts as a narrative but seems to become involved with the Osiris myth; and the "Legend of the Doomed Prince," of whom it was foretold that he would die by the crocodile, the snake or the dog, and tried in vain to escape his destiny, are other examples of this genre. The story of the "Eloquent Peasant," whose pleading in the end secured him justice, and the "Tale of Sinuhe" are both set in a Middle Kingdom background and contain no supernatural element.

The "Tale of Sinuhe" is one of the few texts preserved virtually complete. Sinuhe was a fugitive from the Egyptian court who escaped the frontier guards and lived for years in Palestine among

the Bedouin. At length, feeling the approach of old age, he yearns for his native land and sends a letter to Pharaoh petitioning pardon. Moved by this plea, the king sends him a pardon and welcomes him home, and the story ends happily with his reception at court. A historical background is also assumed for the story, perhaps a burlesque of the voyage of Wenamun, who was sent to Byblos to buy timber for the sacred barque of Amon and found himself treated in cavalier fashion by the local ruler.

The so-called "Maxims and Instructions" belong to the genre of didactic or wisdom literature (*see* WISDOM LITERATURE), a popular form of expression throughout the ancient orient. Advice is put into the mouth of an old man admonishing a younger on the rules of good behaviour, or of a king who gives advice on wise rule to his heir. The "Maxims" of Ptahhotep, of Ka-gemni, of Ani, of Amenemope and others must have been learned by heart as part of a young man's education and widely quoted in conversation; they contain many parallels with Hebrew literature.

If thou art one of those sitting at the table of one greater than thyself, take what he may give, when it is set before thy nose. Thou shouldst gaze at what is before thee. Do not pierce him with many stares. . . . Let thy face be cast down until he addresses thee and thou shouldst speak only when he addresses thee. Laugh after he laughs and it will be very pleasing to his heart and what thou mayest do will be pleasing to the heart. . . .

If thou art one to whom petition is made be calm as thou listenest to the petitioner's speech. Do not rebuff him before he has swept out his body or before he has said that for which he came. A petitioner likes attention to his words better than the fulfilling of that for which he came. . . . (From J. B. Pritchard, ed., *Ancient Near Eastern Texts Relating to the Old Testament,* 2nd ed., pp. 412–413; Princeton University Press, 1955.)

Similarly moral in tone, but with a different flavour, is the so-called pessimistic literature, frequently engraved on the walls of New Empire tombs. The harper at the funeral feast sang of the transience of life:

Follow thy desire while thou yet livest!
Put myrrh upon thy head, clothe thyself in fine linen
Increase yet more thy happiness, and let not thine heart languish.
Follow thy desire and thy good
Fashion thine affairs on earth after the command of thy heart
That day of lamentation will come to thee, when the Still of Heart does not hear.
Mourning does not deliver a man from the netherworld.
(Refrain) Make holiday! Do not weary thereof!
Lo, none is allowed to take his goods with him,
Lo, none that has gone hence has come back!

(From Miriam Lichtheim, "The Song of the Harpers," *Journal of Near Eastern Studies,* vol. iv; The University of Chicago Press, 1945.)

The most remarkable text of this sort is the celebrated dialogue between the "Man Who Was Tired of Life" and his soul, in which he seems to be meditating on suicide. The last stanzas praise death as a happy release from this miserable life:

Death is in my sight today,
As when a sick man becomes well again,
Like going out-of-doors after detention. . . .
Death is in my sight today
Like the perfume of lotuses,
Like sitting on the shore of the Land of Drunkenness. . . .
Death is in my sight today,
As when a man desires to see home
When he has spent many years in captivity.

(From R. O. Faulkner, "The Man Who Was Tired of Life," *Journal of Egyptian Archaeology,* 42:29–30; 1956.)

In contrast to the pessimism of these poems, Egyptian love poetry sounds a happy note:

Do not flutter, O my heart!
The beloved comes to thee,
But so do people's eyes.
Don't let them say of me
"That woman is fallen in love!"
Hold firm when thou thinkest on him,
O my heart, and flutter not!

(From S. Moscati, *The Face of the Ancient Orient,* in Eng. trans., p. 128; Vallentine, Mitchell & Co. Ltd., 1960.)

Poems in praise of the sun-god, the god Amon and other deities are found engraved on the walls of temples and tombs and on private funerary steles. The well-known hymn to the Aton has been likened to Psalm civ. Temple inscriptions contain long ritual texts, often with a mythological content, and that celebrated hotchpotch of magic and mythology, the *Book of the Dead,* contains many references to the world of the gods, but few complete myths have survived in literary form and Plutarch's version must be consulted for the most famous legend of all, that of Isis and Osiris.

C. Ancient Religion

When faced with the multiplicity of representations of deities in the wall paintings and relief sculptures of ancient Egypt, of the ceremonial objects employed in their cult, and of references to them in the various liturgies, hymns, myths and sacred dramas preserved in the ancient writings, the layman may well feel a bewilderment bordering on despair; no religion, it seems, was ever so complex, so contradictory, so baffling. He must, however, remember that he is looking, from a great distance, at a period of 3,000 years of history, before which had lain thousands of years of prehistory, in which beliefs and cult practices evolved and changed, new influences made themselves felt, regional differences were reconciled and theological speculation and mythological invention were active. The catholic tolerance of the Egyptians conserved the old with the new and did not see incongruity in holding simultaneously several conflicting beliefs. This the student of Egyptian religion must seek to understand. In the short space of this article only a brief outline can be given of some of those beliefs.

The primitive religion of the Egyptians appears to have been an animistic nature worship localized in the various districts and villages of the delta and the Nile valley, each of which had its tutelary spirit indwelling in some animal, bird, or reptile, some tree or plant or object, which played a prominent part in the life of that locality. These spirits of nature were either life-giving and friendly, like the domesticated cow or ram, or menacing, like the hippopotamus, crocodile or cobra. In either case their favour must be solicited with set words and actions, they must have houses built for them and offerings must be made to them. Their life cycles were bound up with the life of the village, and the imagined story of their birth and growth, their struggles, triumph and death and their rebirth, were often enacted by the community in an annual sacred drama.

In each locality, the god has his temple, or "divine castle," staffed by his "servants," the priests, who ministered to his daily wants. In return for these services, the god protected his people, ensured the fertility of their fields and herds, and saw to their well-being; in default, he might visit his wrath upon them in the form of plague, famine or other disaster.

1. The Pantheon.—Many local cults were of great antiquity and some of them retained primitive characteristics throughout their history. The totemic origin of cult in Egypt, if our explanation is the correct one, serves to illuminate that aspect of Egyptian religion which called forth derision from the Greeks and Romans and evokes a feeling of distaste or even repugnance today, namely the worship of animals, or theriolatry.

Thoth (*q.v.*), whom the Greeks identified with Hermes, was portrayed as an ibis-headed god and his sacred animal was the baboon. Possibly this may be here a fusion of two zoomorphic deities, in the neighbourhood of Hermopolis, the seat of his cult. The inquisitive search of the ibis among the water plants, and the chattering ape, may each have suggested Thoth's later function as god of science and learning and patron of scribes.

Anubis (*q.v.;* Anpu) of Kynopolis in the 17th nome was a dog- or jackal-god associated with the cult of the dead. Together with the other canine deities, Wepwawet, the "opener of the ways," and Khenti-amentiu, he presided over the desert of the west, the necropolis land where wolves and jackals lurked, and all were regarded as protectors of tombs. Anubis had a special association with mummification. Khenti-amentiu, an early god of Abydos, later became merged with Osiris (*q.v.*).

The crocodile-god, Sebek (Gr. Suchos), was worshiped principally in Al Fayyum, the capital of which was known to the Greeks as Crocodilopolis; there in late times was a pool of sacred croco-

diles. The hippopotamus was a goddess, Taurt or Taweret (Gr. Thoueris), the "Great One," who was specially reverenced by women during pregnancy; so also was Heqet, the frog-goddess of birth. Bast (Bastet), the cat-goddess of Bubastis in the delta, was mistress of love, of matters feminine and of fashion, while Sekhmet (Sekhet), the "Powerful One," was one of a number of awe-inspiring lion-goddesses.

Snakes were a constant peril, and though the horned viper was not worshiped, the goddess Wadjit or Udjo—commonly called Buto (*q.v.*) after the name of the town where her cult originated—had the form of a cobra. She was the tutelary deity of lower Egypt, as Nekhbet, the vulture-goddess, was of upper Egypt, and the king wore the heads of both cobra (uraeus) and vulture on his brow as symbols of the unification of the two lands. Another cobra-goddess was Ernutet, who must be propitiated at the harvest. Hawk- or falcon-deities were, understandably, sky-gods. The sun-god Re (Ra) himself is always depicted as hawk-headed, and it was in the form of Horus (*q.v.*), the hawk, the ancient sky-god of Nekhen (Hierakonpolis), that the king was thought to mount to heaven at his death. Horus had many cult centres; he is sometimes known as Har-wer (Haroeris), "Horus the elder," to distinguish him from Horus the child, son of Osiris and Isis. His annual battle with the forces of evil, often embodied in the figure of Setekh (Set), was a popular theme of mythology (*see* also EDFU).

Re and Horus combined in hawk form as Re-Horakhte, or "Re-Horus of the horizon." Re was thought to sail the sky in his boat, and at night to traverse the underworld, battling as he went with the forces of darkness. Mentu or Month, the war god of Hermonthis, was also depicted as hawk-headed wearing the sun disk on his head.

Another sky-deity, Hathor, bore a more unexpected shape as a cow-goddess. The centre of her cult was Dandarah (*q.v.*), and she was associated in cult with Horus of Edfu, as her name, "Castle of Horus," indicates. She appears to have been a primitive mother-goddess and her cult probably absorbed that of other cattle deities. There were bull-gods in various parts of the country: Apis (*q.v.*), or Hapi, was probably associated with the god Ptah as early as the 1st dynasty if not before; Merwer (Mnevis), at Heliopolis, was probably as old, and other delta bull-cults are known. The hawk-headed Month had a bull-epiphany as Buchis in four cult-centres south of Thebes. There were several ram-gods, the best-known being Khnum of Elephantine, god of the cataract region, and the Theban Amon (*q.v.*), whose name, the "Hidden One," seems to derive from an epithet of Min, god of Coptos, whom he closely resembled. Amon, like Min, was never depicted in animal form, but the ram was his emblem.

Most curious of these animal-gods is Set or Setekh, whose peculiar characteristics—long snout, prick ears, greyhound's body and upright tail—can be identified with no known animal. Setekh was popular with the Hyksos kings, who identified him with their Asian deity at Avaris in the delta; he fell from favour at their expulsion, was reinstated by the Ramessides and execrated once more in the Tanite period, becoming the murderer of Osiris in late legend.

In the case of the bull-gods, at any rate, a single animal was held to be the incarnation of godhead and venerated as such. Apis, Mnevis and Buchis had special markings and when each died and was mummified and given burial in its stone sarcophagus, search was made for a suitable calf born at the time of its death, in which the god might have been reincarnated. This may originally have been the case with other theriomorphic deities, but in late time the whole species was regarded as sacred in the neighbourhood of the cult centre: Bubastites would not kill cats, Mendesians venerated their rams and Hermopolitans all ibises. Thus arose the custom of mummifying sacred animals when they died and burying them in special cemeteries or catacombs such as have been found at Abydos, Hermopolis, Crocodilopolis, Elephantine and elsewhere.

In some centres, the local object of worship was some standard or fetish. Often the original significance of this object was lost in the mists of prehistory. Such was the pillar or "Dad" at Heliopolis, originally apparently a lopped tree trunk with projecting stumps of branches; another was the pole with a waterskin suspended on it, the emblem of Imiut, who was early merged with Anubis. The goddess Neith at Sais was represented by two crossed arrows and a shield.

In addition to these ostensibly native cults, rooted in the soil of Africa, other deities were hospitably admitted to the pantheon who seem to have affinities with the religion of western Asia. Some of these—and all of them in human form—were frankly foreign intruders: the great Semitic goddesses of Syria, Anath, Anahita, Qedesht, "the holy one," and Astarte, and the war-gods Resheph and Baal-Mikal were popularly worshiped in the cosmopolitan east delta cities under Ramesside patronage. Others appear to have entered Egypt very early: Osiris, as a vegetation god, seems to have had a close connection at all periods with Byblos in Syria, while Ptah, whose Semitic name means "the Opener," may possibly (though this has been doubted) have been the god of immigrants who settled in Memphis and founded the 1st dynasty. It has even been thought that cosmic deities like Iah the moon-god, the air- and earth-gods, and even Re himself may originally have been Asian, not African, in origin. Bes (*q.v.*), the dancing dwarf who made children laugh, and whose beard and Silenuslike appearance betray his foreign origin, was introduced during the Middle Kingdom.

Perhaps it was under the influence of these anthropomorphic gods that the ancient zoomorphs received human shape; most of them throughout their history retained their animal heads, skilfully grafted onto human shoulders by the artists who drew them.

The development of Egyptian religion followed closely the political development of the country. A small local deity, with the success of the ruler of that locality, might become the most influential god in a larger area and finally achieve national status when his worship became the state cult. Such was the case with Horus of Behdet (Edfu) in predynastic times, of Re in the later Old Kingdom and most strikingly of all of Amon of Thebes, who from obscurity rose to be the patron of the 12th-dynasty Pharaohs, and again in the New Empire when he became *Amonrasonther,* "king of the gods." The vicissitudes of the god Setekh's cult have been described.

But such changes did not fundamentally alter Egyptian religion. It was not in the nature of the Egyptians to abolish or eliminate; they assimilated, rationalized, added the old to the new, and archaic cults were merged in later rites. At Memphis, Ptah absorbed Sokaris, the god of the neighbouring necropolis whose name may survive in the modern Saqqarah. Considerable ingenuity was displayed in the task of integrating disparate elements and reconciling conflicting beliefs. Where several localities were combined in one city, the local deities might be combined in a divine family, usually a triad. Thus in Thebes Amon was paired with the vulture-goddess Mut and a local moon-god, Khonsu, was said to be their son. In Elephantine Khnum had as his consort Satis, goddess of the cataract; their daughter was Anukis, from the neighbouring island of Sehel. Ptah of Memphis had Sekhmet the lioness to wife, and to complete the triad, the demigod or deified hero Imhotep, the architect and vizier of King Zoser, was regarded as their son.

Around these deities and many more, myths took shape and theologians projected into the local cult their own philosophical speculations. There were a number of different theories about the origin of the cosmos: three main cosmological systems are known. In Hermopolis they believed in the "ogdoad," eight primeval cosmic deities who had emanated from Nun, the water of chaos. The Heliopolitans derived the universe from the god Atum, who, emerging from the primordial hill, begat of himself Shu and Tefnut, air and moisture; Shu in turn begat Nut, the sky-goddess, whom he is often represented as holding aloft, and Geb, the earth-god, who lies at his feet. In the latest, the Memphite theology, it was Ptah who had created the world by thought and spoken word, and eight creator gods were conceived of as contained within the god as aspects of his creative activity; together they were spoken of as the divine ennead. To the theologians was due also the deification of abstract qualities or forces; *e.g.,* the goddess

Ma'at, who personifies truth and cosmic order, and the gods Hu and Sia, who personify creative utterance and intelligence.

These mythopoeic tendencies continued and complexity increased. Osiris, a delta vegetation god, became associated with the dead kings; in the Middle Kingdom, if not earlier, his cult was transferred to Abydos (*q.v.*), the burial place of kings. In the afterlife, every Egyptian now expected to become Osiris. His origin was not thereby lost. As vegetation gods in Syria died at harvest and were revived in spring, so Osiris was thought to have died at the hands of his enemy Setekh. Isis (*q.v.*), his wife, mourned for him, and his dismembered body was reassembled by her piety and that of their son Horus the child (Gr. Harpocrates) who finally fought with Setekh and avenged his father. The legend is known in several forms, from representations of episodes in the drama of his death and resurrection enacted annually at Abydos, from literary papyri, and from Plutarch's account of the myth in its latest form. In the Roman period it was Isis, the mother, whose mysteries were celebrated in secrecy and achieved popularity in Rome itself. Sarapis (*q.v.*), a composite figure combining Osiris with Sokaris, the mummified hawk-god, and Apis the sacred bull, was also worshiped throughout the Roman empire. In Egypt the worship of Isis survived on the island of Philae among the devout Nubian Blemmyes until Justinian closed her temple in the mid-6th century A.D.

With so great a multiplicity of religious beliefs, the educated Egyptian of the New Empire might have been expected to develop a concept of abstract divinity, of the "divine" in its various manifestations. There are indeed many references to "the god." But the significance of this should not be overemphasized. In personal matters one might turn to a particular deity to whom intimate prayers could be addressed and toward whom one had a feeling of personal devotion. But henotheism is not monotheism, and in spite of all that has been written to the contrary it would seem that the ancient Egyptians never developed any tendency toward the worship of one God. The sole possible exception is the heresy of Ikhnaton and his followers, perhaps only a handful of individuals in the priestly school at Heliopolis, who for a brief time attempted to prefer the sun disk Aton above all other gods and even resorted to iconoclasm. That this concept was quite foreign to Egyptian thought is shown by the utter failure of Ikhnaton's enterprise and the prompt return to polytheism after his death.

2. The Priesthood.—The general name for priest, *we'eb*, is a word that means "pure." Bodily purity was ensured by ablution, washing the mouth with natron, and shaving all hair from head, face and body. Priests were circumcised and wore clean white linen garments. Dietary tabus are mentioned by Greek writers, and probably involved abstaining from the flesh of the local cult animals. In the New Empire, but probably not before, the priesthood formed a separate caste; priestly office was normally inherited from father to son, and in the hierarchy of each temple there were ranks from the humblest temple servitors to the high priest and his deputies, who in the case of the state cults were men of enormous wealth and influence. A temple of average size might have a staff of between 20 and 50, but the large national temples had many thousands of priests; and the domains of the temple of Amon at one time constituted perhaps one-twelfth of the whole arable land of Egypt.

The *we'eb*-priests performed the more menial tasks in the temple and constituted the rank and file of the temple hierarchy. The "servants of the god" were higher, it seems, in the scale of privilege, and between or perhaps above them was a category of priests called the "fathers of the god" whose function has not been explained. Specially trained priests chanted the liturgies and others with special functions included the "master of mysteries," the "scribe of the altar," and those, like the *sem*-priest and the embalmer, especially concerned with funerary rites. The staff of the temple was organized in four *phylae* or watches, each presided over by a phylarch. Each watch served for a month at a time, in rotation. A serving priest was called "he who is in his month." Certain priests wore special dress; for instance the high priest of Heliopolis, who bore the title "chief seer," had a mantle of leather studded with metal stars, while the pontiff of the temple of Ptah at Memphis wore a curious pectoral or archaic design and the plaited sidelock that was normally the prerogative of princes. The *sem*-priest wore a leopard skin over his tunic.

There were priestesses, whose function it was to sing and shake the sistrum, especially in connection with the cult of Hathor. These were not professionals but the wives of noblemen who accounted it an honour to take part in the rites of the goddess.

The daily ritual is known in some detail from documents and temple reliefs of the New Empire. After preliminary ablution in the sacred pool, the officiating priest, deputizing for the Pharaoh, entered the sanctuary, scattering incense smoke from a censer; he broke the seal of the shrine, unlocked it, and prostrated himself before the god's statue. Then he undressed it, washed and anointed it, clothed it in fresh garments and painted its eyes with eye paint. He laid food and drink before it, and finally again locked the door of the shrine, obliterating his footsteps. At every stage in the ritual he pronounced the appropriate formula.

3. Temples and Festivals.—A temple, in essence, was the god's house, and the primitive temple seems to have been little more than a reed hut containing the sacred fetish, with an enclosed courtyard in front of it and a pennant or standard on a pole to indicate that the god dwelt here. Flagstaffs continued to be erected in front of the great pylons of later temples, and the hieroglyph for "deity" or "sacred" was a flag on a pole. These remained the essential parts of any Egyptian temple, but inner courtyards were added, with additional pylons before them, and pillared halls gradually deepening in gloom till, in the centre of the temple, the sacred image reposed in its shrine, in mystery and darkness.

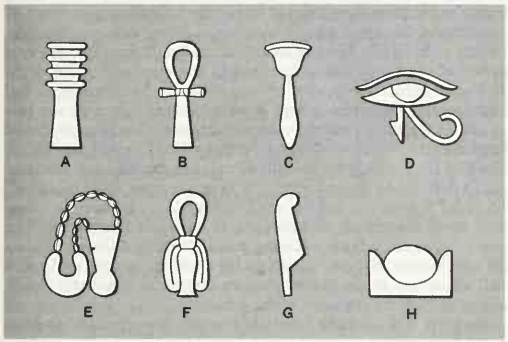
It seems clear that the general public had no place in the daily ritual of the temple and participated in none of the sacred rites. The privileged might be permitted to bring their offerings into the forecourt, but only priests had access to the inner halls, and only the most senior members of the fraternity were permitted to enter the holy of holies. On the occasion of a festival, however, the god's statue was borne out into the streets on the shoulders of priests. On these occasions, perhaps once a year, there was public rejoicing and feasting. The city was "bestrewn with faïence, gleaming with natron, garlanded with flowers and fresh herbs. The prophets . . . are clad in fine linen, the king's suite are arrayed in their regalia. Its youth are drunk, its citizens are glad, its young maidens are beautiful to see; rejoicing is round about it, festivity is in all its quarters, there is no sleep in it until dawn." (H. W. Fairman, "Worship and Festivals in an Egyptian Temple," *Bulletin of the John Rylands Library*, vol. xxxviii, no. 1, 1954.)

One of the greatest of the festivals was the "Beautiful Feast of Opet," lasting nearly a month, when the god Amon left his temple at Karnak and was escorted upstream in a decorated barge amid great excitement, to visit his consort at Luxor. Soldiers, dancers, acrobats and musicians thronged the Nile banks, street vendors set up booths, and the whole population of Thebes went on holiday and turned out to watch the processions. Pilgrims from every part of Egypt flocked to the mystery play at Abydos, where the death and resurrection of Osiris were enacted. (*See* also Feast and Festival: *Egypt.*)

4. Magic and Popular Religion.—In a world thronged by supernatural beings closely concerned with the destiny of man, magic played a vital role in everyday life. Through the medium of those priests who were skilled in the magical arts, men might ward off evil and even coerce spirits into doing their will. Incantations might drive out disease, and spells and amulets were worn on the body to protect it against injury. The manufacture of amulets was a thriving trade, especially in the late period. They were made in a variety of materials including coloured stone, faïence, metal and occasionally wood and glass. The earliest are little animal heads, perhaps the protective animal of the clan, worn by the Badarians. Some were made for suspension, others were laid on the mummy at burial to protect its various members. One of the commonest is the scarab, an oval stone carved on the back in imitation of the dung beetle (*Scarabaeus sacer*), the flat underside being engraved and used as a seal, and often mounted as a signet ring. Scarabs appear first in the First Intermediate period and remained popular throughout Egyptian history; prob-

ably every Egyptian of standing had his own personal seal. The amuletic nature of the scarab was never forgotten, in spite of its practical use, and large "heart scarabs" engraved with a passage from the *Book of the Dead* were placed on the mummy. The dung beetle was connected in popular belief with the creator sun-god, whose epithet *Khepri,* from the verb *kheper,* "to be, come into existence," was written ideographically with a picture of the scarab or *kheper* beetle; the creature is sometimes depicted rolling the solar globe through the sky with its hind legs, just as the dung beetle is seen to roll its ball of dung (*see* SCARAB).

There were a great many other amulets which protected or conveyed qualities of wisdom, strength or health to the wearer. Among them the commonest are shown in the illustration.



GENERALIZED FORMS OF EGYPTIAN AMULETS: (A) PILLAR OF OSIRIS, CONVEYS STABILITY; (B) ANKH OR CRUX ANSATA, CONVEYS LIFE AND IMMORTALITY; (C) PAPYRUS COLUMN, CONVEYS YOUTH, VIGOUR AND FRESHNESS; (D) EYE OF HORUS, CONVEYS HEALTH AND HAPPINESS; (E) MENAT, COUNTERPOISE OF A NECKLACE, GIVES PROTECTION, ESPECIALLY TO WOMEN; (F) GIRDLE TIE ("BLOOD OF ISIS"), GIVES PROTECTION; (G) OSTRICH FEATHER, CONVEYS TRUTH, ORDER AND RIGHT; (H) RISING SUN, CONVEYS LIFE AND RESURRECTION

Figures of the gods, or of genii such as the "four sons of Horus," were worn as amulets to give the protection or attributes of that god to the wearer. There were calendars of lucky and unlucky days, and an oracle could be consulted before any serious enterprise was embarked upon. Sympathetic magic was practised in the manufacture of clay or wax images of enemies, labeled with their names; the figure was then smashed or melted, to ensure the destruction of that enemy. In the late period, small household images of deities were set up in private houses, and the gods of the common people—*e.g.,* Bes and other grotesques such as Taurt, the hippopotamus-goddess, and the dwarf children known as "sons of Ptah"—became popular members of the household cult. Statuettes of Harpocrates, the vanquisher of crocodiles, scorpions and snakes (often called Cippi of Horus), were a popular protection. Demigods, men like the architect and doctor Imhotep and Amenhotep son of Hapu, were worshiped by those who found them more approachable than the great divinities of the official pantheon. The increasing importance of what may be called personal religion culminated in the mystery cults of the Greco-Roman period, by means of which, after initiation, man was able to establish intimate contact with the deity.

5. The Afterlife.—Surrounded on all sides by relics of his ancient past, preserved by the remarkable conservative qualities of climate and soil, it is small wonder that the ancient Egyptian believed in life after death. No people has believed so fervently in personal survival or expended so much time and so large a portion of its resources in preparation for it.

The Egyptians approached death with mixed feelings. They were a happy people who enjoyed life, and like all mankind they dreaded the cessation of earthly existence. Funerary steles exhort passers-by as "ye who love life, ye who hate death." The dead were even thought to exercise powerful influences for good or evil over the living, letters written to dead relatives have been found in which the writer begs the spirit of the dead to cease from troubling him. Yet the general picture of death was of a happy consummation, a continuation of life on a higher, happier plane forever "in peace, in peace, on the beautiful roads of the west" where "a fair, fair welcome awaits him who has reached the west."

A number of beliefs seem to have coexisted about the nature of the immortal element in man. The soul was conceived of as a bird, the *bai* or ba, or as the shadow, or as the ka, the double or counterpart of man which at birth was created, together with the corporeal self, by Khnum on his potter's wheel, and which consumed the offerings set out for it and received the prayers of the dead man's relatives. Another supernatural aspect of man's personality, the akh, was a spiritual power attained only after death. It is uncertain to what extent an attempt was made to clarify and separate the functions of these spiritual entities, to which must be given the general name of "the soul." It was believed that the soul, after death and burial in the tomb, and assisted by the proper rites and magical formulas, continued to live in the tomb the life which its owner enjoyed upon earth, and that it still retained the shape and functions of the body. Faced with the problem of corporeal decay, the Egyptians, at least from the 2nd dynasty onward, attempted to preserve the body by artificial means, and the art of embalming or mummification developed, the object of which was to preserve, as nearly as possible, the appearance of the deceased as it had been in life. (*See* MUMMY.)

6. The Tomb.—As soon as a new king came to the throne, he selected a site for his tomb, his "house of eternity," and began work on it. Nobles of high rank, with the king's permission, spent years in the preparation of their own tombs. Private tombs differed from those of royal personages both in size and in design. They consisted essentially of two parts, the burial chamber, usually at the bottom of a rock-cut shaft which was filled in after the burial service; and the superstructure, which remained accessible to visitors. The latter was brightly painted with scenes from the life of the deceased, often with autobiographical inscriptions relating significant episodes in his career, ending with an exhortation to the living to pray for him and maintain his funeral offerings.

In the Old Kingdom these tombs were usually of the mastaba or bench type with a chambered stone superstructure. Early mastabas have a slab or offering table in front of the façade on which food and drink could be laid daily for the dead man or his ka. Later in the Old Kingdom a tomb chapel adjoined the burial chamber, with a "false door" through which the ka was supposed to come out to take the offerings. To protect the body from possible damage, from robbers or malevolent spirits, it was encased in a coffin and sometimes an outer stone sarcophagus; wooden or stone statues reproducing the features of the dead man were placed in the tomb to replace the mummy in case it should perish in spite of these precautions. In the Middle Kingdom and New Empire and afterward, the private tomb was usually a series of rock-cut chambers cut in the hillside, in the innermost of which representations of the owner and his wife were carved, sitting side by side.

Since the dead were conceived of as continuing their earthly existence in the tomb, elaborate equipment was buried with them, its richness and variety depending on the resources of the occupant. Furniture, chests of clothes, vessels for eating and drinking, weapons and articles of personal equipment were buried for the use of the dead. In royal burials of the 1st dynasty, men and women slaves were slain at the king's tomb and buried by his side to serve him in the next world. Later, models of servants were put into the tomb which by means of magical formulas could be brought to life. Statuettes of servants baking, or brewing, or grinding corn are found in Old Kingdom tombs, and in the Middle Kingdom whole groups of little figures at work in model kitchens or workshops are included together with model fishing boats, houses, troops of militia and so forth. During the New Empire, painted representations of these activities were thought to serve the same purpose.

The decoration of the walls of a tomb also included processions of offering bearers. In order to ensure that fresh food and drink and fresh candles were continually supplied to the dead man for

his daily needs, the tomb was usually furnished with an endowed estate, the produce of which provided the offerings and supported in perpetuity a mortuary priest, often a member of the family, called the ka-priest or "priest of the double." This estate was inalienable and in theory the endowment of such an office ensured that the mortuary cult of the deceased would be continued in perpetuity. In the case of the Pharaohs, whole towns with their serfs were set aside for the provision of a permanent mortuary cult attached to a royal pyramid. (*See* also EGYPTIAN ARCHITECTURE; PYRAMID.)

7. Funeral Customs.—When a noble or a member of his family died, the body was taken to the embalmer's workshop to be prepared for the tomb. It was first eviscerated by means of an incision in the abdomen, and then dehydrated with dry natron (soda) and aromatic resins for 70 days. It was then laid straight, hands folded on chest, and bandaged. Over the first layer of wrappings resin was daubed and on this the funerary jewelry, bracelets, anklets and amulets, was laid. Further bandaging, with the aid of linen pads to fill out the shape, completed the mummy, which was then covered with an inner anthropoid cartonnage, gilded and painted with mythological scenes (*see* MUMMY). Sometimes several coffins then enclosed the body, the outer being of wood.

The internal organs, liver, lungs, intestines and stomach, were separately mummified and enclosed in four vessels known to Egyptologists as canopic jars, each having as a stopper the head of one of the tutelary deities known as the "four sons of Horus." The four jars were usually boxed in a canopic chest.

All was now ready for the funeral. On the appointed day, the mourners arrived to escort the mummy to its last resting place. On a bier drawn by oxen, and accompanied by the wailing widow and another female relative impersonating the goddesses Isis and Nephthys, the funerary cortege proceeded to the bank of the river and thence by boat to the west bank. There, again on a sledge, the coffin was dragged to the tomb and there set on end while the *sem*-priest performed various ceremonies with special ritual implements designed to "open the mouth" of the mummy so that it might regain its powers of speech and be able to partake of the food and drink offerings presented to it. Libations were poured from a series of vessels, and the mummy was then lowered to the burial chamber amid the wailings of the relatives and a group of female mourners who tore their clothes and poured dust on their heads. A long line of servants carried into the tomb the canopic chest, and packed the antechambers with the funerary equipment which was to accompany the dead to the next world. Among this paraphernalia would be a box containing some dozens, or even hundreds, of little mummiform figures of faïence or pottery called ushabtiu (*ushebtiu, shauabtis*) or "answerers" because they were to answer the roll call in the hereafter, in place of the deceased, should he be called upon to work in the heavenly *corvée*. "If the deceased So-and-so shall be summoned to do forced labour, to plow fields or build dikes, or to carry sand from east to west, 'I will do it, here am I!' shalt thou say." With enough of these little slaves to take his place, a man was sure of a life of uninterrupted leisure and enjoyment in the hereafter. (*See* also FUNERARY RITES AND CUSTOMS.)

8. Eschatological Beliefs.—Though life after death was a continuation, in the tomb, of life on earth, a number of different theories, never entirely reconciled, grew up about the royal hereafter. Pharaoh was early thought to continue his existence in the world of the stars, ascending to heaven to join the gods. With the increase in the influence of Re in the royal cult, the Pharaoh was imagined as joining the sun-god and traveling with him in his solar boat through the sky by day, and in another through the perilous darkness of the celestial underworld by night, protected by magical formulas which in the New Empire were inscribed on the walls of his tomb. Solar boats were buried near the king's mastaba or pyramid. These two doctrines, stellar and solar, are found in the "pyramid texts," the oldest surviving body of religious writings, inscribed in the pyramids of the 5th- and 6th-dynasty Pharaohs, which contain many relics of primitive belief and ritual. Also embodied in these texts is the doctrine of the Osirian hereafter, in which the king after death became one with Osiris. As time went on, the nobles who were granted the privilege of a tomb endowment were equally permitted to participate in the royal hereafter, and every noble Egyptian was believed to face, after death, the ordeal of the Osirian judgment and the term "the Osiris So-and-so" became almost equivalent to saying "the late Mr. X."

According to this doctrine, on entering Amentit, the underworld, the dead man had to face his judges for the "weighing of the heart." In the presence of Osiris and the 42 gods who acted as assessors, the dead man's character, in the form of his heart, was placed on the balance by the god Anubis, while the scribe Thoth acted as recorder. In the other pan of the balance was the feather or image of Ma'at (Truth). Sins weighed down the heart, and a man's evil actions were thought to tip the balance against him. If condemned, his heart was thrown to a waiting monster, the "devourer of hearts." He who was found to be without sin, or "true of voice" (justified), was admitted to the company of the gods. It was therefore the prayer of every man to be cleansed of sin, and funerary literature is full of such prayers. Part of the text of one of the many funerary papyri placed in New Empire tombs, known as the "Book of Coming-Forth-by-Day," which we call the *Book of the Dead*, is the so-called "negative confession," the dead man's declaration of innocence. "I have not given short measure," he says, "I have not caused any to hunger, . . . I have not taken milk from the mouth of babes . . . I have not dammed flowing water . . ." and so forth. Having made his declaration, he adjures his own heart: "Set not thyself to bear witness against me; speak not against me in the presence of the judges, cast not thy weight against me before the Lord of the Scales."

Such are a few of the infinitely complex concepts of life after death found in the religious papyri and in the tombs. Many of the eschatological beliefs of Hellenistic times must stem from Egyptian ideas, and mummification of the dead continued into Coptic times. The Gnostic beliefs revealed in the Hermetic writings, in the figure of Hermes Trismegistos (*q.v.*) and the papyri from Khenoboskion reflect in its latest form the simple conviction of the predynastic Egyptian that death is not the end. Whether the poor in pharaonic Egypt had a similar expectation of immortality is not known; the few mean objects buried in their graves lead us to suppose that at least they had some hope of it.

(M. S. DR.)

D. Macedonian and Ptolemaic Period

Alexander the Great occupied Egypt late in 332 B.C. without opposition, the inhabitants greeting him as a liberator from Persian tyranny. He visited Memphis and sacrificed to Apis in accordance with his policy of respecting local religions. After founding Alexandria he journeyed to the temple of Amon at the oasis of Siwah and thereby acquired in Egyptian eyes confirmation of his new position as Pharaoh of the land. In the arrangements he made before departing for the interior of Asia he showed again his concern to hold the people's good will; the Persian satrap was replaced by two Egyptian governors. Taxation, however, was entrusted to a Greek, Cleomenes of Naukratis, who soon obtained, with Alexander's consent, absolute control of the country; under him Alexandria was completed.

On Alexander's death (323) the Macedonian Ptolemy, son of Lagus, was appointed satrap to govern alongside Cleomenes. He lost no time in ridding himself of Cleomenes and began at once to treat the land as his destined kingdom, though it was not until 305 that he took the royal title. The dynasty he founded ruled Egypt to the death of Cleopatra in 30 B.C. (*see* PTOLEMIES). The history of Egypt during this time, as distinct from the history of the dynasty itself and the court at Alexandria, is known only imperfectly.

At first the country was regarded by the Ptolemies as no more than a base upon which to establish their power (indeed, to some extent this remained true of the later Ptolemies also). Like their Seleucid rivals in Asia, they relied entirely upon Macedonians and Greeks for the army and the administration. These settlers, together with other newcomers, such as Syrians, Jews and Anato-

lians, represented only a tiny minority among the several millions of native Egyptians.

The Ptolemies treated the Egyptians' religion with respect, and permitted the native priesthood to retain some part of its old power. They themselves went through the pharaonic rituals and built many temples to the native gods. On their side, many Egyptians of all classes learned Greek and took Greek names. Yet this "hellenization" was quite superficial. Intermarriage between Greek settlers and Egyptian women did not bring Hellenism to the natives; rather, it was the first step in a gradual weakening of the cultural identity of the Greeks, who increasingly adopted the native religion and customs.

After the defeat of Antiochus the Great at Raphia (217), when Egyptian soldiers had a large share in the victory of Ptolemy IV Philopator, nationalist feeling among the Egyptians grew rapidly; armed risings broke out, mainly in upper Egypt, and continued intermittently till 85, when Ptolemy IX destroyed Thebes. Even so, discontent never led to a general insurrection of the sort that had overthrown Persian rule in the 5th and 4th centuries. The cause of these risings is to be found partly in hostility to alien rule, but chiefly in the weakness of the later Ptolemies and their failure, despite good intentions, to ensure that administration at local level was just and efficient. (R. H. Si.)

E. Roman Period

The Roman domination of Egypt lasted for almost seven centuries, from Cleopatra's suicide in 30 B.C. to the triumphal entry into Alexandria of the Arab general Amr ibn al-As in A.D. 642. As in the Ptolemaic era the ruling dynasty and class consisted of foreigners; but now Egypt was merely a member of a larger political unit ruled by an absentee emperor through his deputy the prefect. Absolute control of Egypt being, moreover, an important constituent of the emperor's political ascendancy in Rome, through the corn tribute distributed free to the populace of the capital, Augustus in practice treated Egypt as his personal estate. The prefect, though only of equestrian rank, commanded its three legions, dispensed justice and replaced the Pharaoh in native eyes. Senators were not allowed to enter the country without explicit permission. Augustus' policy regarding Egypt was proved wise by events. In A.D. 69 it was an invitation from its prefect, Tiberius Julius Alexander, that determined Vespasian to make his successful bid for the purple. Unsuccessful attempts to imitate Vespasian were made by the generals Avidius Cassius in 175 and Pescennius Niger in 193, and by the prefect Mussius Aemilianus *c.* 260; but the candidature of Marcus Aurelius Probus in 276 did succeed. Diocletian had a severe struggle *c.* 296 with another pretender, Domitius Domitianus. Thereafter the problem was resolved (until the time of Justinian) by separating civil from military authority. The civil administration of Egypt, subdivided into three provinces, remained in the prefect's hands. All troops were put under the command of the dux or duke of Egypt, an officer independently responsible to the emperor.

The Roman power did not give peace from civil war, but it did confer immunity from foreign invasion. Egypt suffered a short-lived annexation in 270 when Zenobia, queen of Palmyra, profited by the anarchy to add it to her kingdom. In 616 it was lost to the Persians for ten years. Otherwise the country suffered no more than (in the 3rd and 4th centuries) an occasional raid from the Ethiopian tribes of Blemmyes and Nobatae. The boundaries and defense arrangements remained practically as settled by Augustus. The southern frontier was drawn at Syene (Aswan), with a fortified zone and a neutral area between it and the Ethiopians. A project of the second prefect Aelius Gallus to subdue mainland Arabia proved abortive. The commerce of the Red sea forts was developed vigorously under Augustus and Tiberius, and caravan routes through the Eastern desert were organized to support them.

The entrepôt trade from the east was indeed one of the three principal sources of Egypt's wealth. The others were its manufactures (Egyptian craftsmen ranged from glassblowers to physicians and astrologers) and its bumper harvests, gift of the annual flood of its river. Roman policy aimed at exploitation of this wealth for the benefit of Rome. The bulk of the land, "public" land by name, paid a high customary rent in kind; a constantly growing category of estate land was exploited direct by the emperor; land held as private property paid money taxes, as did craftsmen and industries. All individuals except Roman citizens paid poll tax. An army of officials, few of them professionally trained or remunerated, administered this complex system, the paperwork for which (census returns, declarations, assessments, etc.) has survived in the form of innumerable papyrus documents. In the civil service Rome encouraged the growth of a system of "liturgies": that is, individuals nominated by authority were compelled to perform administrative duties for fixed periods, the nominator's property and that of the "liturgist" serving as guarantee in case of failure.

Roman policy was in the main honest and not deliberately oppressive; but the results of too-concentrated and ill-managed a fiscal exploitation were always vexatious, and finally disastrous. The remedy for individuals was to run away from their homes, either to Alexandria or to the desert. Tax collectors' reports show abuses rife as early as the emperor Claudius (A.D. 41–54), when one person in ten in the village of Philadelphia was reported missing from his home for nonpayment of taxes. Even under the Flavians, Trajan and Hadrian (the supposed good period), the vices of the system continued. In good times there was an appearance of well-being; in bad, when the Nile failed, when plague occurred (in some villages in the 2nd century it carried off 40% of the population), when military requisitions were increased to meet a crisis, there was no margin, and the result was misery and revolt. Serious outbreaks are known in the years 153 and 172.

Misery fell heaviest on the Egyptian-speaking "public peasants." Above them in the social scale came the Greco-Roman classes of the nome capitals such as Oxyrhynchus, descendants of the Greek settlers of the 3rd century B.C. They still treasured the classics of Greek literature, learned Homer, read Sappho and Menander and frequented the gymnasium and theatre. These were the men who bore the brunt of the day-to-day Roman administration. Above the nome capitals ranked the few purely Greek cities, Alexandria, a city of magnificent public buildings, still contributed to art and learning, through Greek scholars of the calibre of Harpocration, Ptolemy and Appian and Jewish sages such as Philo. Its preponderantly Greek population, however, was from the outset hostile to Rome, failing to obtain the grant of a city council, and suspecting Rome of favouritism toward its Jewish minority. Anti-Semitic outbreaks in A.D. 38–40 and in 66 left a legacy of permanent hatred and a literature of opposition to Rome known as the *Acts of the Pagan Alexandrian Martyrs*. In 115–117 a fanatical Jewish revolt ended in the near extermination of inland Jews.

The organs of administration in Egypt, at first idiosyncratic, gradually became assimilated to those of the rest of the empire. In 200 Severus, visiting Egypt, gave senates to the nome capitals, and placed on the senatorial order the responsibility for local administration and tax collection. In 212 in a famous edict partly preserved on papyrus, Caracalla gave Roman citizenship to the whole empire, Egyptians included. The process was completed under Diocletian, who introduced a comprehensive tax-quota system based on units of productivity, the *caput* and *iugum*, and imposed the levies by 15-year periods termed indictions. Latin replaced Greek as the language of higher administration. In the 4th century Egyptian autopract landowners, given the right to collect their own taxes and pay them direct to the central authority, could not be prevented from extending their patronage, and the system of the colonate, of service according to status, was recognized officially in 415. The papyri give a lively picture of the authority, pomp and arbitrariness of the autopract family of the Apions at Oxyrhynchus.

The Alexandrian church is reported in the 2nd century to have been much disturbed by Gnostic heresies (a system of direct intuition of God), the tenets of which have been brilliantly illuminated by the find in the 1940s at Naj' Hammadi of a series of Gnostic treatises translated into Coptic (*see* Gnosticism). But the strength of orthodox congregations inland is also shown by finds of fragments of the Greek New Testament written not later

than the 2nd century. The Alexandrian church was remarkable for the catholicity of outlook and thought of its leaders Pantaenus, Clement and Origen, who made Christianity intellectually acceptable to their contemporaries. It was persecuted by the emperor Severus in A.D. 200, then, together with Christians generally, by Decius in 250, again in 257–260, and most severely under Diocletian and Licinius from 302 to 311. An indirect result of persecution was the birth of monasticism (*e.g.*, in Scete), at first practised by solitary anchorites, then in cenobitic monasteries under a rule of which that established by Pachomius *c.* 323 was the prototype. In the country districts Christianity appealed as a last refuge for native Egyptians, and in the form of "Coptic," the ancient Egyptian language, written in characters mainly borrowed from Greek, enjoyed a renaissance (*see* COPTIC LANGUAGE). In Alexandria the leaders of the church wielded an international power, and doctrinal conflicts such as those over the nature of the Trinity tended to take on a political aspect merely because of Alexandrian opposition to Rome and Constantinople. A like factiousness set the Egyptian church in schism after the Council of Chalcedon in 451, since it espoused the monophysite doctrine of the single nature of Christ after the incarnation. (*See* ALEXANDRIA, GREEK ORTHODOX PATRIARCHATE OF.)

The end of Roman rule came suddenly, heralded by the appearance of Arab armies before Pelusium early in 640. (E. G. T.)

F. Medieval Period

1. Muslim Conquest.—Although Syria had fallen into the hands of the Arab conquerors in A.D. 636, the Byzantine authorities failed to anticipate an attack on Egypt by putting the Sinai approaches in a state of defense. Amr ibn al-As (*q.v.*), the companion of Khalid ibn al-Walid in the Syrian campaign, began the invasion of Egypt with only 4,000 men under his command and without positive permission from the caliph Omar. Marching by way of Al 'Arish and the coastal road in the early spring of 640—the precise date is not established—he besieged and took Pelusium (Al Farama), 30 mi. S.E. of the present Port Said, and from there pressed on to Bilbis (Bilbays), on the caravan road between Pelusium and the Nile, where he conducted an arduous siege and stormed the town. He then raided the valley as far as Al Fayyum, recrossed the Nile and, with the help of reinforcements sent him by the caliph, defeated a Byzantine army between Heliopolis ('Ayn Shams) and the great fortress of Babylon (Babilyun) in July 640. Babylon, which lies at the southern end of modern Cairo, was then invested by the Arabs. At the end of a long siege the garrison capitulated under an armistice, and Amr, having eliminated pockets of Byzantine resistance in the delta, appeared before the walls of Alexandria.

Without siege equipment the defenses could neither be scaled nor breached, so he left a holding force there and made a raid on upper Egypt. The emperor Heraclius, who had banished Cyrus, the patriarch of Alexandria, for his mishandling of the defense of Egypt against the Arabs, died in 641; his successor, Constans II, reversed his predecessor's sentence on Cyrus, restored him to the patriarchate and sent him back to Alexandria to make peace with the Arabs. By the terms of the treaty the Greek residents were given a year to depart. At the end of this period, in Sept. 642, the city was peacefully transferred to the Muslims, and subsequent attempts by Byzantine naval expeditions to recover the city failed. The story that Amr caused the great library at Alexandria to be destroyed lacks adequate documentary proof.

The virtual surrender of one of the richest provinces of the Byzantine empire poses the question of why it happened. Over 900 years of contact with Ptolemaic, Roman and Byzantine civilizations had failed to influence the mass of the people, who represented an even older culture. The Greco-Roman way of life never penetrated beyond the small urban ruling body and throughout remained an alien influence. The adoption of monophysite Christianity by the mass of the inhabitants was more than a break with the state doctrines of the Byzantine official world; it was a spiritual secession from the west. In the psychological realm an exhausted Christianity, crippled by over 200 years of theological dispute, was at a disadvantage against the triumphant young faith of Islam. On the material plane the incompetence of the Byzantine leadership was aggravated by the absence of a unified military command.

But although the people of Egypt may have concluded that the Muslims could not be worse masters than their former Byzantine oppressors, there are no grounds for supposing that the Copts actively welcomed the Arabs as liberators. Until the terms of the armistice of Babylon had been made public they could not have known what the character of the Arab rule would be. Further, the Arabs, who knew neither Greek nor Coptic, would have been unable to reassure the inhabitants of their intentions or to spread Muslim propaganda among them. Finally, though the Copts may have detested the rule of Constantinople, it is unlikely that they would have welcomed conquerors whose religion was opposed to their own. The attitude of the Copts toward their new masters seems to have been negative, born of uncertainty and indecision. The Arabs treated the Copts no better or worse than they treated the non-Muslim *dhimmis* elsewhere. In return for paying tribute in money and kind the Christian inhabitants were exempted from military service and were allowed to practise their religion and administer their community affairs through their own chiefs.

Mystery surrounds the identity of the Christian leader whom the Arab historians called Muqawqis, a possible corruption of the Coptic word *khaukhios,* Caucasian, a reference to Cyrus' origin. This person was commonly supposed to have signed both the armistice of Babylon and the treaty of Alexandria. It is less improbable that there were two different persons; the first a Coptic commander of Babylon who negotiated an armistice of primary concern to the Arabs and the local Coptic inhabitants, and the second, Cyrus himself, whose treaty of Alexandria regulated the international relations between the Arab power and the government of Constantinople.

Amr's choice of Al Fustat (derived from the Greek *phossaton,* "boundary") as the Arab capital was dictated by military considerations. It grew out of the Arab camp for the siege of Babylon and is situated just north of the Byzantine fortress. Like other Arab centres established in Syria and Iraq it progressed from a temporary to a permanent camp and lastly to a city. But the Arabs, unlike the Byzantines, had no mind for town planning, and there in Al Fustat occurred the usual stages of Arab urban development. Round a core of permanent structures—mosques, palaces and administrative offices—grew up a vast confusion of houses and huts, sited to no plan and periodically ravaged by fire and pestilence. The confusion would thus remain until some prince with sufficient resources would seek to escape from the squalor round him by rebuilding the capital on a new and more healthful site, where the squalor would in time follow him and the process would begin over again. Al Fustat remained the capital of Egypt from 641 till 969. The only surviving monument to the founder is the mosque of Amr, much altered by later restoration. In Abbasid times the governors of Egypt lived at Al 'Askar, a northern suburb of Al Fustat. But Al Fustat was not immediately abandoned, though it was almost destroyed by fire in 1168, and from then onward was in slow decline. The place is now called Misr al Qadimah (Old Cairo).

Amr cleared the sweet water canal that connected the Nile with the former Byzantine port of Klysma, called by the Arabs Al Qulzum, a mile from the modern Suez, and thus made possible the passage of vessels laden with grain from the Nile valley to the ports of Arabia. After a short while Egypt was divided into two governorates: Abdullah ibn Sa'd was allotted upper Egypt and Amr lower Egypt, though for a time Abdullah was sole governor of the whole country during Amr's absence. In 641 Abdullah had made his first raid—it was no more than a reconnaissance—beyond the southern border of Egypt. In 651–652 he raided the Nubians in force, bombarded Dongola, the seat of the Christian kingdom, with catapults, and destroyed the church. The Nubians then sued for peace, which was granted them on liberal terms against an annual exchange of gifts.

Friction arose between Amr and Abdullah. The caliph Othman wished the governorship to be divided, leaving Amr in command of the army and Abdullah, who was a partisan of Othman, in control

of the finances, the more lucrative office. Amr took umbrage at this and retired from public life. Othman was murdered in 656, and it is significant that his murderers started from Egypt on their fatal mission to Medina, for Egypt was deeply involved in the political turmoil which then rent the caliphate. Amr returned to Egypt in 658 as the nominee of Mu'awiya, the rival of Ali, for a second term as governor, and stayed until his death about 663 when he was over 90. (*See* also CALIPHATE: *The First Four Caliphs*.)

2. Egypt Under the Omayyad and Abbasid Caliphates.—From the time of Amr to 968 Egypt was ruled by governors sent from the capital of the caliphate. During this period there were no fewer than 111 governors, of whom a few, like Amr, occupied the post more than once. Frequent postings of governors led to recurrent changes in the direction of affairs. Of the 21 governors who ruled during the period of the Omayyad caliphate all were loyal to the government of Damascus except one who supported the pretender Abdullah ibn al-Zubayr, a nephew of 'A'isha, the prophet's wife, in his luckless struggle with the Omayyad caliph Yazid I, and he had to be forcibly driven from his post by the caliph Marwan I. Marwan's son Abd al-Aziz, brother of the caliph Abd-al-Malik, was one of the stronger Omayyad governors. During his 20 years in office Abd al-Aziz made Egypt practically independent and devised an administrative system which lasted for many years. When the Omayyad caliphate was overthrown, Marwan II, the last caliph of the line, fled to Egypt where he was killed in 750.

A persistent feature of Arab rule was the absence of planning in the economic and fiscal fields. Amr's Nile–Red sea canal, opened to fulfill an immediate need, was blocked by order of the caliph al-Mansur in 761–762 because the caliph wished to deny grain to Medina, where his uncle was in rebellion against him. A vain effort to reopen the canal to navigation was made during 1135–38 in the reign of the caliphs al-Rashid and al-Muktafi. The same substitution of the whims of governors for considered policy is seen in the treatment of the Christian subjects. The caliph Omar expressly forbade the building of churches in towns created by the Muslims, yet the governor Abd al-Aziz allowed the building of a church at Helouan (Hulwan). Another trait that persisted until the end of the Abbasid caliphate was the extreme jealousy with which the caliphs regarded any tendency on the part of their regional officers to become popular either with the soldiery or with the inhabitants.

At first the occupying power confined its activities to the collection of tribute and the maintenance of garrisons. The Arabs lived apart from the people; they were originally forbidden to acquire land, and there was no concerted effort on the part of the Muslim conquerors to convert the Copts. The old Byzantine administrative divisions of the country, as well as the Byzantine taxing system, were maintained; and the distinction, which was to be characteristic of classical Islamic polity, between the taxation of Muslims and non-Muslim tolerated faiths, whereby the former paid taxes mainly on land and the latter mainly by a poll tax, did not then exist. As time went on the character of the Arab rule changed from one of mere occupation to colonization and to a progressive assimilation of the native inhabitants. Conversions to Islam, particularly of Christian landowners, had the effect of depriving the treasury of revenue, and as various expedients to prevent this failed, the theory eventually prevailed that all landowners, Muslim and Christian, paid land tax, and members of the tolerated sects paid a personal tax also. The continuance of Arab rule and the gradual exclusion of Christians from employment in the state services marked a conscious endeavour by the ruling body to Islamize the country. The use of the Arabic language had been steadily gaining ground; in 706 it was made the official language of the government offices, though the occasional use of Greek, the old language of administration, continued for some time longer.

The Arab conquest transformed Egypt from an outer province of a western power to a central region of Islam. The possession of Egypt made possible the extension of Arab dominion along the north African coast into Spain and, to a lesser extent, up the Nile into Nubia. The earliest Arab campaigns in north Africa were only raids until in 670 Okba ibn Nafi, one of the caliph Mu'awiya's officers, founded Kairouan and used the place as a base for further raids on the Byzantine settlements to the west. With the coming of Hasan ibn al-Nu'man al-Ghassani as governor of North Africa (Ifriqiya) about 698, a permanent Muslim stronghold was established in the territory and it was made independent of Egypt. No comparable development occurred on the southern border of Egypt, where Aswan long remained the southern limit of permanent Arab occupation. The Christian kings of Dongola were generally in treaty relations with Cairo. 'Anbasa ibn Ishaq, the last Abbasid governor of Egypt, was faced by a revolt of the Beja of the eastern desert who invaded upper Egypt and did much damage. An army commanded by Mohammed ibn al-Qummi followed them into the Itbay mountains east of Qus and in 858 defeated them. Food for the Egyptian army was supplied from ships on the coast near 'Aydhab (Old Suakin).

The concentration of Abbasid power on the Tigris, remote from Egypt, made supervision difficult. The Abbasid governors on an average changed even more frequently than their Omayyad predecessors, an indication that the central government placed little faith in its agents in distant Egypt and saw to it that they did not hold office long enough to become too experienced or too popular. Remote control led unerringly to a decline in Egypt of the political authority of Baghdad. The caliph Harun al-Rashid, so successful elsewhere, allowed Egypt to relapse into civil turmoil where rival factions strove for power. His reign of 23 years (786–809) saw 24 governors in Egypt. The Copts rose several times against the government, not usually because of any religious issue but because of a feeling of injustice over taxation. In 831 the Copts even joined with the Arab Muslims in a revolt, a combination so menacing to the Abbasid power that the caliph al-Ma'mun himself visited Egypt in 832 while his general, Afshin, crushed the rebels, massacred the Coptic prisoners and sold their families into slavery. Never again was there a serious Coptic revolt. The Arabs fared little better. The leading place in the ruling institution which they had held since the days of Amr was taken by foreign mercenaries and they were elbowed out into the wilderness.

It was from this time that Egypt was caught up in the transition that was going on throughout the Abbasid empire. The province was assigned as a fief to officers of non-Arab ancestry who themselves resided in Baghdad and employed stewards to administer Egypt on their behalf. Abdullah ibn Tahir, one of the caliph al-Ma'mun's generals, was the first of these fief-holders to govern Egypt by delegation. Government, always predatory in character, degenerated into a system of extortion by agents who had no interest in long-term improvement of the country's economy. The "Turks," who became a powerful and finally dominant element in the ruling body, were not necessarily all Turks in the ethnic sense but were Turkish-speaking soldiers—slaves who included Kurds, Mongols and other northern peoples in their ranks. Collectively they were known as Mamelukes, soldiers of slave status who were owned (Arabic, *mamluk*) by the sovereign or the amirs, who would themselves most probably be of Mameluke origin. For their institutions and organization *see* MAMELUKE.

3. The Tulunid Secession.—The weakening ties between the central authority in Baghdad and the far-flung provinces of the Abbasid dominions threw the whole caliphate into disorder. Spain and North Africa detached themselves without great difficulty, and next came the turn of Egypt. In 868 Ahmad ibn Tulun, son of a Turk from Ferghana, was sent to Egypt to join the governor's staff, and soon made himself master of the province. Feeling no loyalty to the Baghdad caliph, he established himself as sovereign of Egypt and ignored the caliph al-Mu'tamid's request for tribute. By this action the surplus revenues of Egypt were retained in the country, and from the accumulated wealth Ahmad and his successors strengthened the fighting forces, developed agriculture and embellished Cairo with public buildings, of which the great mosque of Ibn Tulun is the most striking survival.

The Tulunid dynasty nevertheless rested on weak foundations; it depended for its existence on a military caste of Turkish and Negro praetorians, for Arabs had long since ceased to be enrolled. In 877 Ibn Tulun easily occupied Syria and brought that country

into an association with Egypt that was maintained, with intermissions, until the Ottoman conquest of 1517. This dynasty, having no roots in either Egypt or Syria, soon fell into decay. Khumaruya, Ahmad's son and successor, was a voluptuary, and his successors barely maintained even a semblance of authority. Under Ibn Tulun's fourth and last successor, another son, called Shayban, the Syro-Egyptian state reverted in 905 for 30 years to the caliphal authority of Baghdad.

Another dynasty of central Asian origin (they claimed descent from the old kings of Ferghana) were the Ikhshidids, who, having succeeded to power in Egypt, incorporated Syria within their dominions, to which they added Mecca and Medina. The dynasty was short-lived. The founder, Mohammed ibn Tughj, on whom the caliph bestowed the princely title of *ikhshid*, was the sole ruler of any merit. His two sons who followed him were puppets in the hands of a remarkable vizier, Abu'l Misk Kafur, who was neither Arab nor Turk but a black slave from Ethiopia or the Sudan. He had the luck to be sold to Mohammed ibn Tughj, who recognized his talent, made him tutor to his children and promoted him to be an officer. Kafur showed outstanding ability and in 940 captured Aleppo for his master. On his deathbed Mohammed ibn Tughj appointed him guardian of one of his sons, and Kafur was the real ruler during the nominal rule both of this son and of Ali, his brother and successor. Kafur's magnificence was legendary, but it was not all empty pomp, for he was a genuine scholar, a patron of scholars and is said to have been a poet. Al-Mutanabbi, greatest poet of the day, for a time claimed him as a friend. His patronage, however, caused his clients to give him an overkindly biography and greater fame than he perhaps deserved. The Ikhshidid rulers held power from 935 to 969 but left no conspicuous monuments. In 969 their last weak ruler, a boy named Abu'l-Fawaris Ahmad, grandson of the founder, was overturned by the Fatimid power. The Hamdanids, a dynasty founded by Hamdan ibn Hamdun with its centre at Mosul, drove the Ikhshidi governors from northern Syria.

4. An Egyptian Caliphate: the Fatimids.—The Ikhshidids suffered from the political weaknesses of their Tulunid predecessors; neither dynasty rested its power on other than alien force. Thus the Ikhshidids were unable to resist the onslaught of the Fatimids (*q.v.*) from the west, a new force which did not rely wholly on the sword. For the Fatimids were more than an aggressive military power; they were religious propagandists fired by missionary zeal. They burst into Egypt at a time when the country was debilitated by famine, caused by a succession of poor Nile floods, and disorganized by the rulers' incompetence in economic affairs.

From the time of the Arab invasion Egypt had been governed by Sunnite Muslims. The Fatimid rulers and the Safawids of Persia were the only two important Shi'ite dynasties of Islam. Because of the religious sanction claimed by the Fatimid sultans, they occupy a special place in Egyptian history, for they were not only secular princes but claimed to be caliphs as well. Under Fatimid auspices Egypt was no longer a dependency of another caliphate or a precariously placed rebel state but a caliphate in its own right which for a time eclipsed its waning Sunnite rival at Baghdad.

The Fatimids traced their descent through the imam Ismail and al-Hasan to the caliph Ali and Fatima, the prophet's daughter, but the authenticity of their claim, though conceded by such authorities as Ibn Khaldun and al-Maqrizi, was widely questioned in the Sunnite world. The founder of the family, Sa'id ibn Husain, drove the Sunnite Aghlabids from North Africa and had himself proclaimed ruler, adopting the title of Imam 'Ubaydullah al-Mahdi. From his capital near Kairouan he extended his rule from the frontiers of Morocco to Cyrenaica. In 914 he advanced along the coastal road into Egypt, captured Alexandria and raided the delta. 'Ubaydullah's successors continued the founder's expansionist policy until in 969 a Fatimid general, Jawhar al-Siqilli (the Sicilian), destroyed the remnants of the Ikhshidi forces and entered Al Fustat.

Jawhar at once began to transfer the capital to a site three miles to the north of the old city, to a spot which the court astrologers named Al Qahirah, "the victorious," for Mars was in the ascendant on the naming night. The construction of the city of Al Qahirah (corrupted into Cairo by the Italians) was so far advanced that by 973 the new capital was already functioning. Among the public buildings Jawhar erected the mosque of Al Azhar which later became a famous seat of Islamic learning. The Fatimid dominions now comprised North Africa and all the former Ikhshidi territories, the holy places of Arabia and Yemen. In 969 a political mission was sent up the Nile. Abdullah ibn Ahmad al-Uswani, head of the mission, wrote an account of the Christian kingdoms of Alwa and Maqurra beyond the Fatimid frontier.

Of the Fatimid caliphs who ruled after their occupation of Egypt, two were sovereigns of note. Abu Mansur Nizar al-Aziz, whose peaceful reign lasted from 976 to 996, opened a period of imperial splendour remarkable for the construction of mosques and public works including irrigation canals, and for the toleration accorded to Christians. Al-Aziz was followed by an enigmatic and controversial figure, Abu Ali al-Mansur al-Hakim, who, though reverenced by the Druzes and regarded as an imam by the Isma'ili community, has been accused by Sunnite historians of horrible atrocities. Al-Hakim did, however, kill several of his viziers and demolished Christian churches, including that of the Holy Sepulchre in 1009, an action which gave impetus to the preparations for the first crusade. Further, he forced indignities upon both Christians and Jews, making them wear humiliating tokens of identity. But it should be added that in his fanatical bigotry he was an exception to the general Fatimid toleration toward the non-Muslim population. If it is remembered that the founder of the family, Sa'id ibn Husain, was probably descended from a leader of the Isma'ili brotherhood, it will not be surprising that al-Hakim ended by declaring himself an incarnation of God, an assumption accepted by the Druzes. Al-Hakim died by violence, though the cause of his murder is uncertain. The caliphs who succeeded him were rulers of mediocrity who surrendered their powers to viziers.

The structure of the Fatimid state followed in outline the usual Islamic pattern but with differences in detail. The rulers maintained a general control but left the routine administration to viziers and councils. The chief minister was the vizier, who was responsible for the civil side of the administration, including the all-important department of finance. Traditionally the vizier was a civilian, but in 1074–75 the feeble al-Mustansir appointed a soldier, the Armenian Badr al-Jamali, to be both vizier and commander in chief. The absence of administrative continuity, a cardinal weakness in medieval Islamic government, followed from the rapid and often capricious appointment and discharge of viziers, while the increasing power of the military in the state undermined the authority of the civil departments.

Political deficiencies did not prevent a signal development in other aspects of human activity under the Fatimids. Even at this early period the trade of the Red sea was sensitive to the European market for oriental goods, such as spices and other luxuries, the demand for which was being stimulated by the crusades. Alexandria, the outlet for the Egyptian transit traffic to Europe, via the Red sea, grew prosperous. The Fatimid age was one of great achievement in architecture. The mosque of Al Azhar in Cairo became a centre of Shi'ite theological training and propaganda. It survived the transition from Shi'ism to orthodoxy under the pious Saladin and became, as it still is, the principal centre of Sunnite Islamic teaching in the world. The mosques of al-Hakim, and al-Jiyushi, and the Aqmar mosque, testify to the artistic greatness of the Fatimid architects. Three gates in the walls of Cairo, Bab al Futuh, Bab al Nasr and Bab Zawayla, built in the Byzantine style, are the grandest Fatimid military works that survive. (*See* also Islamic Architecture: *Egypt.*) The age was rich in the arts of woodcarving, metal and glass working and ceramics. (*See* also Islamic Art.)

The Fatimid era vividly impressed itself upon contemporary Islam and left a strong imprint on Egyptian civilization. The inability of the Fatimid rulers to rival the high culture of the great Arab rulers of Spain, and their rather lukewarm, but not negligible, patronage of the secular sciences and literature,

BY COURTESY OF (TOP AND CENTRE PANELS) THE METROPOLITAN MUSEUM OF ART, NEW YORK, (BOTTOM PANEL) THE BOSTON MUSEUM OF FINE ARTS

EGYPTIAN HANDICRAFT OF 1580–1200 B.C.

Top: Carnelian plaques, XVIIIth dynasty, from bracelets or armlets, inscribed with name of King Amenhotep III (1411–1375 B.C.) The lower plaques are ornamented with scenes from the king's life. The centre plaque, of dark sard, is ornamented in pierced work with a human-headed sphinx holding before it a cartouche bearing the name of the king. Gold setting is modern. Centre: Turquoise blue glass vase found at Thebes, XVIIIth dynasty; gold statuette of the god Amon, in the likeness of King Thutmose III (1501–1447 B.C.); faience cup in the form of a lotus flower, ornamented in low relief. XIXth–XXth dynasties. Bottom: Wooden coffin painted with hieroglyphics and symbolic figures relating to the soul's journey to the world of the dead

EGYPT

BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

CARPENTERS' TOOLS OF THE NEW KINGDOM, *c.* 1550–*c.* 1350 B.C.

1. Bronze ax with original wooden haft
2. Bronze pull saw
3. Bronze hand saw
4. Bronze adz with original handle
5., 8. and 11. Bow drill with two bits
6. Model adz from foundation deposit
7. Oil horn
9. Bronze chisel with wooden handle
10. Bronze bradawl
12. Slate hone

BY COURTESY OF (1–6) THE TRUSTEES OF THE BRITISH MUSEUM; (7) UNIVERSITY COLLEGE, LONDON

EGYPTIAN WEAPONS, 3000–1250 B.C.

1. Copper (or bronze) halberd riveted to a silver handle; 12th dynasty, about 2000 B.C.
2. Bronze dagger with gold handle; 18th dynasty, about 1450 B.C.
3. Copper dagger with ivory handle and silver rivets; 12th dynasty, about 2000 B.C.
4. Bronze *khepesh* sword; 19th dynasty, about 1250 B.C.
5.–6. Bronze ceremonial battle-axes; 18th dynasty, about 1450 B.C.
7. Two-lugged copper ax; 1st dynasty, about 3000 B.C.

BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM (6) METROPOLITAN MUSEUM, N.Y.

JEWELRY AND GLASS, 2000–1350 B.C.

1.–4. Necklaces of gold, carnelian and amethyst. 12th dynasty; about 2000 B.C.

5. and 7. Parts of necklaces of gold and carnelian. 18th dynasty; about 1400 B.C.

6. The pectoral of Senusret II, gold inlaid with lapis lazuli, turquoise and carnelian—the design expressing the wish that the sun-god may grant an eternity of years of life to the king. 12th dynasty; about 1890 B.C.

8. Beads and pendants of faience. 18th dynasty; about 1350 B.C.

9.–11. Vases of polychrome glass. 18th dynasty; about 1350 B.C.

EGYPT

BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

EGYPTIAN FAIENCE AND GLAZED POTTERY FROM THE I. DYNASTY TO THE ROMAN PERIOD (3200 B.C.–A.D. 150)

1 and 2. Faience hippopotami. 12th dynasty, *c.* 2000 B.C. 3. Faience bowl with embossed animal friezes. 30th dynasty, *c.* 350 B.C. 4. Falcon for inlay, 30th dynasty. 5. Vase *Glazed pottery* Roman period, A.D. 150. 6. Part of a kneeling figure in inlaid glaze. 20th dynasty, *c.* 1180 B.C. 7 and 8. Faience lily and thistle for inlay. 18th dynasty, *c.* 1370 B.C. 9. Faience bowl. Early 18th dynasty, *c.* 1500 B.C. 10. Mouse, polychrome. 12th dynasty, *c.* 2000 B.C. 11. Ape, blue, discoloured. 1st dynasty, *c.* 3200 B.C. 12–14. Lotus cups, blue glaze. 18th–22nd dynasty, *c.* 1400–800 B.C. 15. Throw-stick from tomb of King Ikhnaton. 18th dynasty, *c.* 1365 B.C.

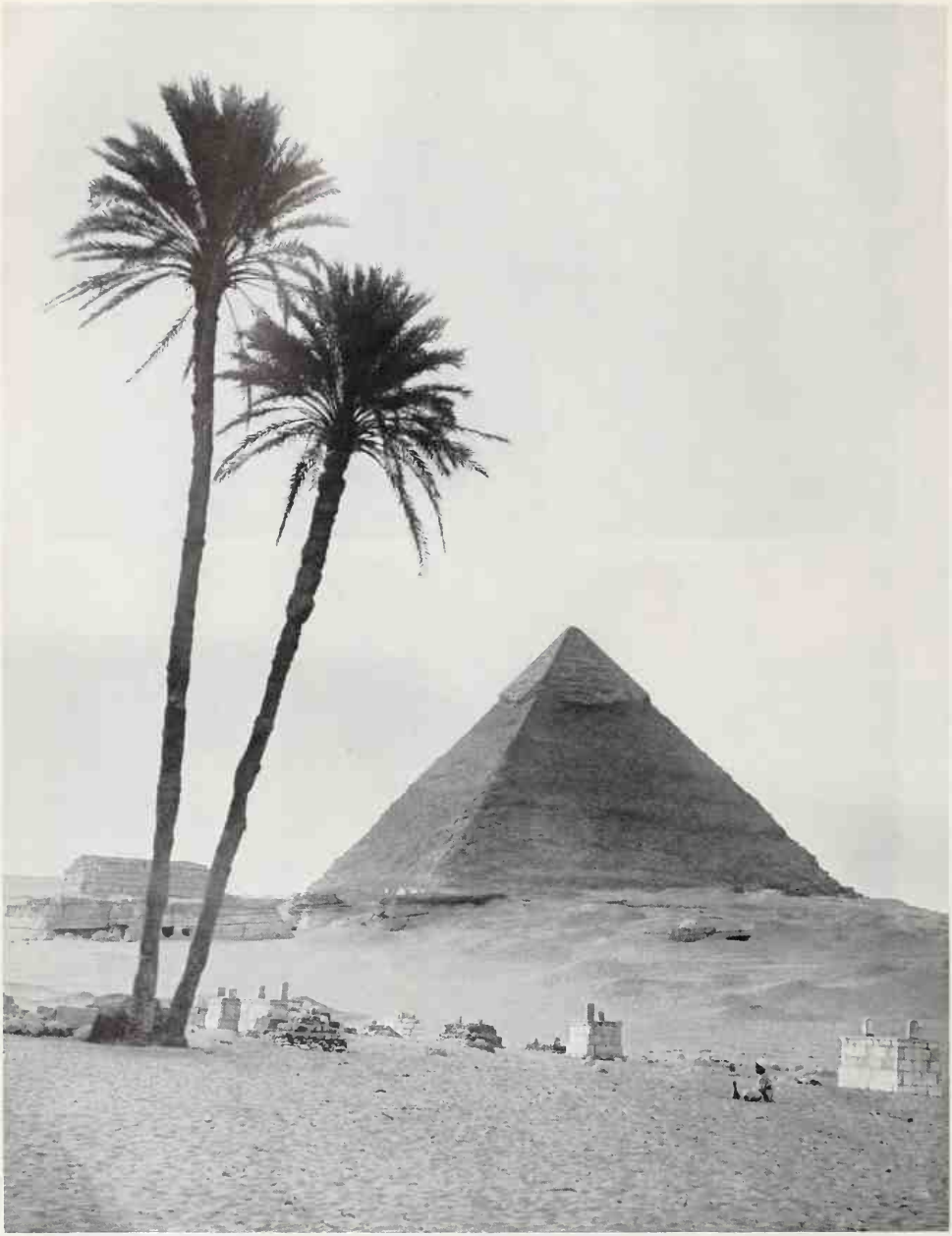
BY COURTESY OF MRS. N. DE GARIS DAVIES AND (1) THE TRUSTEES OF THE BRITISH MUSEUM, (2) THE EGYPT EXPLORATION SOCIETY

WALL PAINTINGS OF THE 18TH DYNASTY

1. Wall painting in the tomb at Thebes of Rekhmire, Vizier of Thutmosis III., depicting negroes bringing tribute of ebony, ivory, pelts, a leopard and a monkey. About 1450 B.C., 18th dynasty

2. Two of the seven daughters of King Ikhnaton (Akhenaton) from a fresco at Tel-el-Amarna, the king's city of palaces. Amarna period, about 1370 B.C., 18th dynasty

EGYPT



PHOTOGRAPH DONALD M. LEISH, COPR

THE PYRAMID OF KHEFREN, GIZEH, EGYPT

The pyramid of Khefren (Kha-f-Ra, IV. dynasty), perpendicular height 447½ ft., is the second largest of the three great pyramids of Gizeh. At the top a fragment remains of the original coating of limestone and unpolished granite slabs

Jabal Musa (Mount of Moses), sometimes identified with the biblical Mt. Sinai in Egypt, with the Greek Orthodox monastery of St. Catherine (A.D. 530) on its northern lower slope

Stanley bay, a popular bathing place at Alexandria, with hundreds of beach cabins

Carpet weavers at their trade in Asyut, Upper Egypt

EGYPTIAN SCENES

A Bedouin woman of Sinai, a barren area inhabited by nomads but militarily important to Egypt because it borders Israel

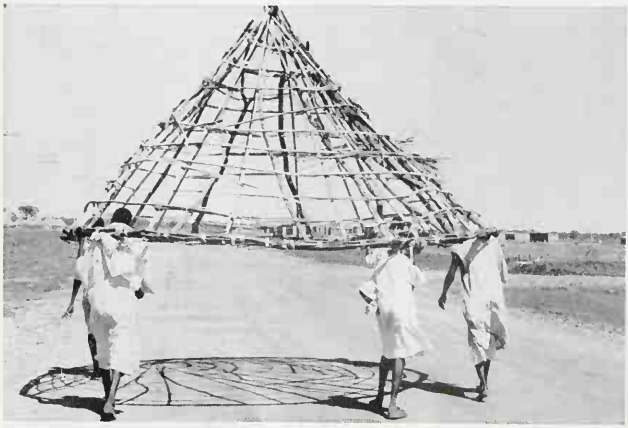
BY COURTESY OF (BOTTOM LEFT) UNITED ARAB REPUBLIC TOURIST OFFICE; PHOTOGRAPHS, (TOP LEFT) ELLIOTT ERWITT—MAGNUM, (TOP RIGHT) J. ALLAN CASH, (BOTTOM RIGHT) WAAGENAAR, CAMERA PRESS-PIX

EGYPT

A view of the mosque of Mohammed Ali in Cairo with the pyramids in the distance

A typical Egyptian plow (*mihrath*) consisting of a wooden pole about 6 ft. long with a metal-tipped stake at the rear which makes a shallow furrow. Drawn by buffaloes, its design is precisely the same as the one used in Pharaonic times

SCENES IN EGYPT



Nubian workers in southern Egypt carrying the skeleton of a roof which is first constructed where the poles are obtained and then carried to a selected location. Most Nubian tribes occupy small villages of round huts with cylindrical walls and conical thatched roofs

PHOTOGRAPHS: (TOP) A. F. KERSTING (CENTRE) ELLIOTT ERWITT—MAGNUM, (BOTTOM) © FREDERICO PATELLANI/CAMERA PRESS-PIX

was not the fault of the rulers, for the earliest caliphs were cultured men. Al-Hakim, of whom the Sunnite authorities record so much evil, founded in Cairo a "Hall of Wisdom" (Dar al-Hikma) for teaching and propaganda, and had among the servants at his court Ali ibn Yunus, most celebrated of Egypt's medieval astronomers. But the fact remained that the dynasty was heretical in the eyes of the great majority of Muslims, and scholars from the far larger Sunnite world shunned these gifted heresiarchs and damned their works. It must also be remembered that, during the long period of Fatimid decline, life and property were insecure; the court libraries of Cairo were several times pillaged by the soldiery.

5. Egypt and the Crusades.—The coming of the first crusade to Palestine coincided with the last enfeebled days of Fatimid rule. The Frankish knights besieged Jerusalem, the immediate goal of their armed pilgrimage, and the unsettled garrison, which had only a short while before captured the city from the Seljuks, failed to offer a determined resistance. On July 15, 1099, the crusaders took the city, while the Fatimid vizier al-Afdal was defeated near Ascalon; whereupon, with the help of the fleets of the Italian maritime states, the crusaders occupied a long stretch of the coastal plain. In 1118 Baldwin I, the Frankish king of Jerusalem, raided Egypt itself and penetrated into the delta. In 1124 the Fatimids lost Tyre (Sur) and their fleet was defeated by the Venetians. In 1153 Ascalon fell to the Franks.

Fatimid military operations were at all points frustrated by internal dissensions among the commanders and troops. Profiting from the poor discipline of the Muslim forces, the crusaders raided the Egyptian coastal lands in 1163. In 1164 under Amalric I, brother and successor to Baldwin III, the crusaders returned to Egypt, this time as allies of the vizier Shawar ibn Mujir in his fight against Shirquh, an officer under the orders of Nureddin, the independent amir of Damascus. In a series of complicated marches and sieges during 1164–68 Shawar and the crusaders were defeated by Shirquh at the battle of Babayn on April 11, 1167, and in 1168 were forced to retire to Palestine. Shirquh, who became vizier to the last Fatimid caliph, al-'Adid, did not live long; on his death in 1169 his nephew succeeded him to the vizierate.

This nephew, Saladin (*q.v.*) (Salah al-Din Yusuf ibn Ayyub), was not a Turk as were so many of his fellow officers, but a Kurd born at Tikrit on the middle reach of the Tigris. His youthful interest was religion rather than fighting, a circumstance that was to influence his career. He began to formulate his life's ambition to champion the cause of Muslim orthodoxy against both the heretics within Islam and the crusaders. His first task was to eradicate heresy from Egypt.

The Fatimid state was in the last stages of decay. Though the first Fatimids had lifted Egypt to the pinnacle of greatness, their Shi'ite ideology had made no permanent impression on the people, who remained devoted to Sunnite orthodoxy. Saladin, though appointed vizier by the Shi'ite sultan al-'Adid, professed to hold his post as the agent of Nureddin of Damascus. The anomaly by which this devout Sunnite professed allegiance to a heretical caliph in Cairo was not lost on Saladin, who, on Sept. 10, 1171, marked the return of Egypt to the orthodox fold by having the name of the caliph of Baghdad mentioned in the Friday prayers. The formal transfer from heresy to orthodoxy was received in perfect calm by the Muslim population. Shi'ism had been tolerated in Egypt only when the Fatimid heresiarchs were strong; now that they were weak their religious deviationism lost its appeal. On the advice of Nureddin, Saladin restored the supremacy of the orthodox Islamic law by discharging the Shi'ite judges and appointing in their places judges of the Sunnite schools.

On Nureddin's death in 1174 Saladin became independent ruler of Egypt (for the history of the dynasty he founded *see* AYYUBIDS). To complete the gesture of restoring Egypt to Islamic orthodoxy he sent an ambassador to the caliph, who, with geographical licence, invested him with the governance of Egypt and its dependencies in North Africa, Syria, Nubia and Arabia. The territory that actually passed to Saladin comprised Egypt as far south as the region of Ibrim together with the control of the desert road to Aydhab and the possession of the Hejaz. His hold over any part of North Africa, in spite of his diploma of investiture, was theoretical.

Serious unrest flared up on the Egyptian southern border, which had become a haven for partisans of the Fatimids. Fatimid refugees, who had fled to Nubia on the downfall of their regime, raided the Arab-Nubian tribe of the Banu Kanz who held Aswan, and Saladin was forced to send two punitive expeditions against them. The first, commanded by the amir Shaja'a al-Din al-Ba'alabaki and his ally the Kanz al-Dawla (the hereditary title of the chief of the Banu Kanz), relieved Aswan in 1172 and drove the rebels southward. A widespread conspiracy to restore the Fatimids, suppressed with much bloodshed in lower Egypt, had repercussions up the Nile when the rebels re-formed, apparently aided by the local Christians. In 1173 Saladin's brother Turanshah, fresh from fighting in Yemen, penetrated Nubia, stormed Qasr Ibrim, pillaged the Christian church and, as an act of piety, slaughtered 700 pigs. To Dongola, capital of the Christian kingdom, he sent an envoy, who reported on the poverty of the country ahead, and Turanshah penetrated no farther into Sudan. In 1174 the Banu Kanz, this time with an Arab and Negro following, invadad Egypt in the Fatimid interest. Saladin sent another of his brothers, al-Adil, against the rebels, who were crushed and their leader, the Kanz al-Dawla, killed.

Having secured his southern border, Saladin began to prepare an offensive against the Frankish territories in Palestine and Syria, but before he could start a war he must first bring order and unity to the Muslim camp. The death of Nureddin in May 1174 had thrown Syria back into its normal condition of fragmentation. Saladin's relations with Nureddin, in spite of his loyalty to this amir, had not been easy, and, when Nureddin's son, al-Malik Salih Ismail, succeeded his father, Saladin's loyalty evaporated. After a tour of inspection of his castles in Sinai and on the Gulf of Aqaba, he occupied Damascus and overpowered an army of the Malik Salih near Hama on April 13, 1175, and dutifully reported the success of his campaign to the caliph at Baghdad. Saladin was now master of most of Muslim Syria. A Frankish embassy visited him in Damascus and negotiated a truce. Malik Salih again took up arms but was worsted in battle on a field between Hama and Aleppo. Saladin then marched against the Assassins, who held a ring of castles in the mountains of northwestern Syria and who had twice attempted to murder him. In 1176 he besieged the stronghold of the Assassins' leader, Rashid al-Din Sinan, the redoubtable Old Man of the Mountain, and extracted from him an undertaking not to attack Saladin again.

Syria was now pacified and Saladin could turn to his main preoccupation—the driving of the crusaders from Muslim territory. Up to this time his encounters with the Franks had been sporadic. In 1171–73 he had twice fought against the principality of Montréal (al-Shawbak, south of the Dead sea) without result. On both occasions the Franks had vigorously counterattacked and driven him back. In 1174 a Sicilian fleet in alliance with Fatimid partisans blockaded Alexandria and besieged the port. Saladin hastened to the coast of Alexandria and drove them back to the sea.

Command of the Red sea was essential to the defense of Egypt and for protecting communications with the holy cities of Arabia. Important pilgrim routes led from the Nile valley in upper Egypt across the desert to Al Qusayr and 'Aydhab. A raid by the crusader Raynald (*q.v.*) of Châtillon, lord of Kerak (Krak or Crac du Désert), a castle to the east of the Dead sea, placed this vital waterway in jeopardy. In 1182–83 Raynald carried out a desperate raid on Muslim shipping. The Bedouin of Sinai helped him to carry ships in sections overland from the Mediterranean coast to Eilat at the head of the Gulf of Aqaba, where they were assembled and launched.

For two years Raynald's fleet was the terror of the Muslims. Sailing over to 'Aydhab they raided the harbour and destroyed a pilgrim convoy, then, riding inland, massacred a pilgrim caravan. Raynald next crossed over to the Hejaz and was on his way to raid Medina itself when his ships were sunk by a Muslim fleet from Al Qulzum. Raynald with some of his men escaped northward to his fief, and the Red sea became safe for Muslim navi-

gation for 300 years, until the arrival of the Portuguese.

It was not until 1187 that Saladin had completed his preparations for a major offensive against the Franks. He now had under his command a united Muslim Syria. Truce-breaking by the Franks provided him with a propaganda weapon for rousing Islam to a total war to eliminate the Latin states that formed a wedge into the heart of the Muslim world. He opened the war by an attack on Kerak, which failed. He then quietly marched north to Tiberias, whose castle fell after six days of siege. A Christian relieving army marched up from the coast and met Saladin's host near Tiberias at a spot called the Horns of Hattin. There, in July 1187, the Christians were utterly defeated. Of the Frankish leaders only Raymond of Tripoli managed to escape to Tyre. Guy de Lusignan, king of Jerusalem, and most of his knights, the grand master of the Templars, and Raynald of Châtillon were taken prisoner. Saladin killed Raynald; the rest were ransomed.

After the victory the Muslim army carried out a systematic reduction of the Frankish fortresses. Acre surrendered on Aug. 9, Jerusalem on Oct. 2. The Frankish realms collapsed, leaving only some isolated inland castles and the coastal towns of Antioch, Tripoli and Tyre with a narrow littoral in Latin hands. Saladin sent to the caliph the crown of the king of Jerusalem and a gilded cross which the crusaders had placed on the mosque of the Dome of the Rock.

The loss of Jerusalem was the immediate cause of the third crusade, which engaged the largest and best-equipped Frankish army that ever took the field in Palestine. The first objective of the combined operation was Acre, which fell on July 12, 1191. Saladin was now experiencing financial difficulties in keeping his troops in the field. The zeal of the amirs was flagging, and they and their troops began to retire from the theatre of war. The great offensive against the Franks was over, peace was signed on Sept. 2, 1192, under which both sides kept their gains, a settlement which left a small Latin kingdom extending from Antioch to Jaffa. Jerusalem remained in Muslim hands but was to be open to Christian pilgrims. Saladin died at Damascus on March 4, 1193. (*See* also CRUSADES.)

Although the result of the war was an armistice, not a victory, Saladin's Syrian policy revealed the hand of a master strategist. After his death Saladin's sultanate was divided among his heirs, of whom his son al-Aziz inherited Egypt. Other sons ruled over districts of Syria. The sons quarreled with one another until their uncle, Saladin's brother, al-Adil I, contrived between 1196 and 1199 to make himself master of Egypt and most of Syria. Al-Adil was known as a man of peace who lived generally on good terms with the crusaders to the advantage of Muslim commerce with the west. He permitted the Venetians to establish a trading station at Alexandria and the Pisans to post consuls there. His sons founded new branches of the Ayyubid family in Syria and Iraq, though the Egyptian branch continued to be regarded as the senior. Al-Adil died in 1218. Shortly before his death a crusader army, transported by the fleets of the Italian states, landed near Damietta (Dumyat) and in 1219 took the town. Al-Adil's son al-Kamil, the fifth Ayyubid, spent nearly two years fighting in the delta before the Franks abandoned Damietta and sailed away in 1221 under an armistice. In 1228 Frederick II, the Holy Roman emperor, arrived in Palestine to conduct the sixth crusade. By diplomacy and little fighting he reconstituted the Latin kingdom. Sultan Kamil used Frederick as an ally against Malik al-Nasir of Damascus and yielded Jerusalem and other strong points as the price of the alliance. The crusaders would have extended their territorial gains but they suffered from a shortage of recruits from Europe and from internecine quarrels. They held Jerusalem until 1244 when Sultan al-Salih (also sometimes known as Ayyub) invited a force of central Asian Turks to recapture it.

The last crusader assault on Egypt occurred in 1249 as al-Salih lay dying. The invasion was led by Louis IX, king of France, and Damietta was again the immediate objective. On June 6, 1249, Louis entered the town without fighting. Then he made the fatal mistake of attempting to advance on Cairo at the season of the Nile flood. Fever racked the crusaders' army as it floundered among the flooded irrigation canals. The Muslim forces cut the Franks' line of communications and utterly routed their army. Louis with most of his knights was held prisoner for a time and then ransomed.

Al-Salih's son Turanshah was called from Iraq to succeed, but this young man soon proved his unfitness to rule by his stupidity and lack of tact. He offended his late father's slave guards, who on April 30, 1250, killed him. An unusual situation arose when Shajar al-Durr, al-Salih's widow, became sultana and on documents and coins had herself styled, among other titles, *Malikat al-Muslimin* ("Queen of the Muslims"). She was recognized in Egypt, but the Syrian amirs refused to pay her homage. The caliph took the side of the Syrians and asked the Egyptian amirs to choose a man in her place. To elude this command the amirs of Egypt appointed 'Izz al-Din Aybak as commander in chief and he at once married Shajar al-Durr. To placate the Syrian Ayyubids, who were still dangerous, they elected Musa, one of the Syrian branch of the family, as co-sultan, and his name appeared on documents and coins. Aybak, however, was the effective ruler. His administration revealed a certain rough vigour, but he lacked the higher qualifications for leadership in the circumstances of Mameluke (*q.v.*) Egypt. He antagonized the amirs; on Sept. 18, 1254, he killed a commander who had successfully suppressed an Arab rebellion in middle Egypt. Many Mamelukes, among them the future sultan Baybars, fled to Syria out of the tyrant's way. Aybak met his death in a palace intrigue when, on April 10, 1257, his consort Shajar al-Durr in a fit of jealousy had him murdered, whereupon the slave women of Aybak's first wife battered her to death with their slippers.

6. Mameluke Rule.—The Mameluke slave-sultans of Egypt are conveniently divided into two series: the Bahri and the Burji.

The Bahri Mamelukes.—The Bahri Mamelukes were so called from the fact that they were drawn from the bodyguard of the sultan Ayyub, who were quartered in barracks on Roda (Ar Rawdah) island in the river (Arabic *bahr*) off Cairo. The 24 Bahri sultans were mostly Turks by descent. On the death of Aybak, al-Muzaffar Sayf al-Din Qutuz, the regent, deposed Aybak's son and seized power. The Mongol *ilkhan* Hulagu, grandson of Genghis Khan, descended upon Baghdad in 1258 and wiped out what little remained of the eastern caliphate. There were no immediate repercussions in Egypt, though in 1260 Hulagu's army penetrated into northern Syria and drove out all but one of the Ayyubid princelings there. Qutuz repelled these amirs, who regarded themselves as the lawful heirs to the Egyptian sultanate. Qutuz also broke the Mongol reputation for invincibility when he had the courage to give battle to an army of Hulagu, commanded by the general Kitbugha, at Ayn Jalut in Palestine on Sept. 3, 1260; the Mameluke army was victorious and so saved Egypt from devastation. Baybars, the Mameluke officer who had fled from the wrath of Aybak in Cairo and had fought for Egypt at Ayn Jalut, quarreled with Qutuz, killed him and took the Egyptian throne (1260), ruling under the title of al-Malik al-Zahir Rukn al-Din Baybars.

Baybars (*q.v.*) was the most able of all the Bahri sultans. His action in establishing a refugee scion of the caliphs of Baghdad as a puppet caliph in Cairo in 1261 was an effort to surround his rule with an aura of legitimacy. The fact that the Abbasid caliphs had long been deprived of any political power was of less importance to Baybars than their symbolic authority as heads of the Islamic community. The Mameluke rulers in Egypt maintained the institution until the end of their government in 1517, when the Ottoman sultan Selim I deposed al-Mutawakkil III, the last of the shadow caliphs of the Abbasid line. Under Baybars' immediate successors the crusaders were finally dislodged from Palestine. Qalawun (1279–90) captured Tripoli in 1289. After Qalawun's death his son al-Ashraf completed the campaign by taking Acre in 1291. The remaining crusader fortresses were all in Muslim hands by the end of the year.

The Burji Mamelukes.—The Burji Mamelukes originated in the regiment created by the Bahri sultan Qalawun. They were quartered in the Cairo citadel (*burj*) and ruled, or rather maintained themselves precariously in power, from 1382 until their displacement by the Ottoman power in 1517. Of the 23 chiefs who hacked

their way to power all but two, Khushqadam and Timurbugha, were Circassians. These able and fearless desperadoes were technically a ruling oligarchy with an elected sultanate, however they were not one united oligarchy but a number of factions, each led by an amir or combination of amirs, with separate bodyguards. Amir strove against amir for the mastery of the state, while each faction intrigued with or fought against the other. But whatever their political deficiencies, there was no doubt of their valour. One of the Burji sultans, Bars Bey, led an Egyptian fleet to Cyprus and conquered it in 1426. The island had been a crusading base, and so long as it was in Christian hands it was a menace to the Mameluke ports of Egypt and Syria. Luckily for the Mamelukes they avoided destruction by Timur, who from his home in central Asia had conquered an empire that stretched from Moscow to India. In 1400 he plundered Aleppo and devastated Damascus. The Mamelukes made peace with him, but it was not until after Timur's death in 1405 that Egypt was delivered from the threat of this Asian scourge.

The Mameluke sultans never had a well-defined frontier on the north. The border states between the Taurus and the Tigris were in constant unrest. The Ramazanoghlu of Adana were their vassals, but farther east Albistan, whose centre was at Marash, was rendered chronically unstable by the presence of Turkmen tribes who were sometimes the allies of the Mamelukes but who prevented Qait Bey from gaining a permanent foothold there. The Mamelukes could never rely on the loyalty of any of the Turkish- and Kurdish-speaking vassal states north of the Syrian plain. In 1481 Qait Bey had protected Jem, a rebel brother of the Ottoman sultan Bayazid II who in 1485 made war against the Mamelukes. After much desultory fighting peace was patched up in 1491 by which Cilicia was placed under Mameluke sway.

The already unfriendly relations between Mamelukes and Ottomans were worsened after the rise of the Safawid dynasty in Persia. The Safawids were Shi'ites whose agents pursued an active campaign of proselytization along the borderlands and northward into central Anatolia where Shi'ism had many adherents. The westward spread of Persian influence represented a major threat to the Ottoman dominions, so that any Mameluke-Persian alliance must be prevented. The Mamelukes were in friendly relations with the Persians. In 1514 Sultan Selim I, who had succeeded Bayazid in 1512, won a decisive victory over Shah Ismail at Chaldiran and occupied his capital, Tabriz. Once the Safawids were weakened Selim felt ready for war against the Mamelukes, but he did not begin hostilities lightly for he had a proper respect for Mameluke military strength.

The Mameluke fighting spirit does not seem to have declined, but their armies were under several disadvantages in a war with the Ottomans. They had failed to keep abreast with the advances in military science and above all in the use of firearms. The subject people in Syria detested them, and they could not trust the security of their communications with Egypt. Moreover, their government was in dire financial straits. Both sides made a show of avoiding war. But when the Mameluke sultan Qansu al-Ghawri asked the Safawid shah for help, war was inevitable. By better generalship and superior firepower, as well as by treachery of a Mameluke commander, Selim heavily defeated the Syro-Egyptian army at Marj Dabiq, between Aleppo and 'Ayntab, on Aug. 24, 1516. Sultan al-Ghawri was killed in the battle. Syria yielded without resistance. Tuman Bey, last of the Burji Mamelukes, energetically prepared Egypt for defense, but he was harassed by a desperate shortage of money and equipment. Selim even now hesitated before assaulting the central bastion of the Mameluke empire, for, although he possessed a superior fighting machine, his own line of communications through Anatolia and Syria was perilously long and exposed. In Jan. 1517 he entered the delta, stormed the Mameluke fortified camp at Ridaniya outside Cairo and, after some days of street fighting, crushed the last resistance in the city. Tuman Bey, fighting to the end, was caught and hanged in April.

If the Mameluke period was one of political and later of economic decay, it was also one of high achievement in the civilized arts. Urban life and artistic creativeness went on their way unconcernedly in the presence of political violence. Among the Bahri rulers were men with a keen artistic sense. Sultans Qalawun, al-Nasir and Hasan built mosques, schools and tombs of elegance and patronized learning; the Burjis Barquq, Qait Bey and al-Ghawri rivaled the Bahris in architectural taste. Scholarship flourished during the last years of Mameluke rule: Ibn Khaldun, the great sociologist; al-Maqrizi, the topographer-historian, and Taghri-Birdi, the author of a compendious history of Muslim Egypt, were all writing under Burji rule. Al-Siyuti, who died in 1505, was pouring forth his encyclopaedic knowledge when Mameluke Egypt was in the final stage of political decline.

7. The Ottoman Period.—The Ottoman occupation of Egypt brought a fundamental change in the external relations of the country. Egypt was now reduced in status from an independent sultanate to a province of an empire. Syria was detached from Egypt and henceforth came within the Ottoman provincial jurisdiction, a reorientation that was to the financial advantage of the Egyptian treasury since the Egyptian revenues were no longer dissipated in the occupation and defense of Syria. Internally Egypt, at least during the first century of Ottoman rule, was spared the ruinous effects of maintaining a chaotic ruling body in continual struggle for power at home. As a province of the empire Egypt paid tribute to Constantinople, but the tribute, originally assessed at about a quarter of the total state taxes, was fixed and comparatively light. Sultan Selim I continued the former Mameluke commercial relations with the European states and confirmed the capitulations previously enjoyed by the French. On Sept. 17, 1517, he concluded an agreement with the Venetians by which they undertook to pay the Ottoman government the annual tribute which they formerly paid the Mamelukes for the possession of Cyprus. In Sept. 1517 he appointed the Mameluke Khayr Bey as *vali* of Egypt and placed Damascus in charge of another Mameluke, Janberdi al-Ghazzali. And though he left in Egypt a corps of six regiments of Bosnian janizaries (increased by his successor to seven), Selim interfered little with the existing administrative organization. Mameluke amirs were retained in the governance of the *sanjaqs* into which Egypt was divided, and were left in possession as fiefholders of most of the cultivated area of Egypt.

Selim's successor, Suleiman I, called by his subjects "the Lawgiver" and by Europeans "the Magnificent," gave close attention to the affairs of Egypt. In 1526 the first Ottoman cadastral survey of Egypt was made, as the former registers, which were the key to the land tax, and hence of the highest importance to the government, are said to have been deliberately destroyed by fire—whether by Egyptians hoping to conceal their tax liability from the conqueror, or by the government desiring to make a new survey unhampered by inconvenient comparisons with past records, is not known. The survey of 1526 was revised in 1550. In 1528 Suleiman renewed the capitulatory privileges accorded to the French in Egypt and extended them to the Catalans. But this was the last capitulation that applied specifically to Egypt. From that year the capitulatory regime in Egypt, as a province of the Ottoman empire, came within the scope of the general concessions accorded by Constantinople to foreign states. The prestige of the Ottoman sultan as guardian (official documents used the pious formulary *khadim*, "servitor") of the two holy places, Mecca and Medina, was assured by a real control of the Red sea, which permitted the safe transit of pilgrims and grain from Egypt to the Hejaz ports.

During the first 50 years of Ottoman rule Egypt was the base for expansion in the Red sea hinterland. In later Mameluke times the Portuguese had been feeling their way round the coast of Africa until in 1497 Vasco da Gama sailed round the Cape of Good Hope and northward into Muslim seas. During the next century the Portuguese possessed themselves of fortified points commanding the entrance to the Persian gulf and the Red sea. In 1513 they bombarded Aden. During the reign of Suleiman I, the governor of Egypt, Khadim Suleiman Pasha, set out from the naval base at Suez and made a voyage to India, and in 1538 defeated a Portuguese fleet. Özdemir Pasha, a Circassian commander, conquered much of Yemen in 1546. Özdemir stayed there as *beylerbey* and, at some period before 1561, crossed the Red sea, occupied Sawakin

(Suakin), Massaua and Zaila, and founded the Ottoman coastal province of Habash. Özdemir's son Osman consolidated his father's gains. In 1541 a Portuguese squadron commanded by Estevam, son of Vasco da Gama, had penetrated the Red sea, sacked Sawakin and threatened Suez. But the Portuguese failed to gain a foothold and the Red sea once again became a Muslim waterway. In Nubia the Ottomans extended their authority beyond the Second cataract and placed janizary garrisons in the three fortresses of Aswan, Ibrim and Sai.

By the beginning of the 17th century the troubles that were besetting the Ottoman power both internally and on its frontiers in Europe and Asia were being felt in Egypt. Under strong sultans such as Selim I and Suleiman I, the Mameluke beys, who had been entrusted with a substantial share in the administration, maintained discipline. They were allowed, however, to maintain private standing armies for which they purchased white slaves, and at the end of the 18th century some of the beys were employing Europeans in their service as gunners and instructors. With the advent of cloistered sultans at Constantinople and the administrative decline of the empire at its centre, the weaknesses of the Ottoman system in Egypt became apparent.

In proportion as the governors were denied effective support from the central government, the beys became less submissive to Ottoman authority. In process of time the janizaries, through intermarriage, became identified with the urban population. Repeated mutinies took place, and the authorities in Constantinople, far from supporting the governor, gave in to the mutineers and confirmed their demands to the extent of originating a custom by which the Ottoman government accepted the Mamelukes' deposition of a governor.

Tension between the governor and the Mamelukes encouraged the rise of a third force, the nomad Arabs, who lived on the fringes of the cultivated lands and who began raiding into the valley. In upper Egypt the Hawwara tribe, under their chief, Human abu Yusuf, usurped the entire government and set up a local state. Though his rule was comparatively unoppressive, Human's presence was a threat to the Ottoman hold on Egypt; he was not unseated until 1769 when a Mameluke expedition led by Mohammed Bey abu Dhahab killed him and broke the Hawwara hold.

In an effort to reduce embezzlement the practice developed whereby the accounts of each outgoing governor were scrutinized and the governor not allowed to leave the country until he had paid all sums he owed the treasury. The laws and regulations that in theory protected the taxpayer from illegal exploitation were administered by officials accustomed to practising corruption, though corruption was not universal and not all Mamelukes were oppressors. Compared with preceding regimes the Ottoman rule bore lightly on the people, and in any case the closely knit Egyptian social structure with its guilds and religious orders had developed through the centuries a toughness, and yet a resilience, which enabled it to accept and survive bad government. All real power was by now wielded by the Mameluke fief-holders represented by their two chief officers of the province: the *shaikh al-balad* (governor of the city), who was the senior of the two, and his junior colleague, who enjoyed the ceremonial title of *amir al-hajj* (leader of the pilgrim caravan). Under the baleful anarchy of the beys Egypt fell back into the state of internal chaos from which Sultan Selim had delivered it.

The dreary story of intrigues, faction fights and assassinations among the ruling beys, which continued into the 19th century, was scarcely interrupted by a Circassian named Ali Bey who gave a fleeting impression of being superior to the usual type of Mameluke. In 1750 Ali Bey became *shaikh al-balad* after his opponent, Khalil Bey, the holder of the office, had been strangled. Once in power Ali killed the murderer of Ibrahim Bey, his former master, an act which so roused the anger of the other beys that he fled to Acre, whose governor obtained his pardon by the Sublime Porte and his reinstatement as *shaikh al-balad*. In 1766, after the death at Constantinople of his patron the grand vizier Raghib Pasha, Ali again fled the country, this time to Yemen, but in the course of the next year his party in Cairo grew strong enough to permit his return. The Porte conferred recognition upon him; then, suspecting his loyalty, ordered the governor to kill him. Ali intercepted the order for his execution, obtained the enthusiastic support of the beys and in 1768 defied the Ottoman power.

In the course of a few months he subjugated the Hejaz and appointed as sharif of Mecca a relation of his, who obligingly bestowed on Ali the title of sultan of Egypt and "lord of the two seas," titles conventionally used by the Ottoman sultans. He thereupon made the gesture of sovereignty by minting coins bearing his monogram and having his name mentioned in public worship in mosques within his jurisdiction. In 1772, after a combination of his former friends had turned against him, he was forced to retire to Syria, where at Acre a Russian war vessel of the Orlov squadron, which was operating in the Mediterranean against the Ottoman navy, supplied him with arms. Then, gaining some successes in Syria, he set out for Egypt, but his army, weakened by desertions, was routed, and he died, perhaps poisoned, in Cairo on May 8, 1773.

Egypt reverted to the former governance of Ottoman *valis* and Mameluke beys, each continually striving to undermine the other. Two colleagues of Ali Bey, who afterward went over to his enemies, Ibrahim Bey and Murad Bey, established joint rule as *shaikh al-balad* and *amir al-hajj*. In 1786 the Porte sent an expedition to Egypt to re-establish Ottoman authority, and Ibrahim and Murad were deposed. Ismail Bey, who succeeded as *shaikh al-balad*, died in a disastrous plague that swept through Egypt in 1791. Constantinople at last recognized the need for competent rulers, so Ibrahim and Murad were recalled to power. They were still in office in 1798 when Bonaparte's expedition arrived in Egypt.

The three centuries of Ottoman rule made little internal impression upon Egypt. The Ottomans, even in the first years of their administration, never established in Egypt a typically Ottoman civilization such as flourished in Constantinople and in Anatolia. In contrast with the Arab conquerors, who imposed their language and religion upon Coptic Egypt, the Turkish-speaking Ottomans, though of the same religion as the Muslim Egyptians, never implanted their own language outside the small circle of the interracial Ottoman minority whom Europeans called Turks and whose social influence persisted until 1952. Even Ottoman political institutions underwent a radical transformation under Egyptian conditions. The Egyptian masses regarded the Ottomans as aliens, while the Ottomans despised the Arabic-speaking subjects as people designed by nature to be ruled over. Only a few Egyptians assimilated themselves to Ottoman ways, and some of these, chiefly men of learning, found employment elsewhere in the imperial service. The later 18th century saw an influx of Syrian Christians, fleeing from Ottoman persecution in their homeland, and of free Sudanese seeking employment in the Mameluke households and in petty trade.

In the arts of civilization the Ottoman contribution was poor, for the Ottoman officers had not the ancient Mameluke appreciation of architecture and fine arts. Their public buildings were generally utilitarian and undistinguished, though in Cairo and Alexandria they built some pleasant private houses which were almost all demolished during the course of the 19th century in the prevailing passion for adopting European, usually Italianate, styles. Two mosques, the Sinaniyya at Bulaq, built in 1571 by Sinan Pasha, *vali* of Egypt, and the small mosque of Mohammed Bey abu Dhahab, built in 1774, are their worthiest monuments. The Mohammed Ali mosque, on the crest of the Cairo citadel, impressive when seen from the city, reveals in its skimped workmanship the less attractive side of the 19th-century Levantine contractor's art.

G. Modern Period

1. French Occupation.—Charles Magallon, French consul in Egypt from 1793 to 1797, had recommended to the Directory in Paris the political and economic advantages to France of the occupation of Egypt. Napoleon Bonaparte, who had access to Magallon's report, became obsessed with the project; the Directory was finally convinced and a French army landed on the western outskirts of Alexandria in July 1798. Accompanying the expe-

dition was a party of 165 French scientists, artists and men of letters. Bonaparte issued an ingenious, if self-contradictory, manifesto announcing that the French had come as friends of the Ottoman empire to restore its authority, to free Egypt from Mameluke tyranny and to uphold Islam. After an easy capture of Alexandria the French marched up the Nile, destroyed a Mameluke army at the battle of the Pyramids on July 21, and entered Cairo. The flight of the Mamelukes had removed the last vestiges of civil order and Bonaparte at once filled the political vacuum by setting up a military government combined with a form of indirect rule. Divans (councils) were created in the larger towns of French-occupied Egypt, with a grand divan in Cairo having bilingual French- and Arabic-speaking secretaries and a French commissary as chairman and adviser. Bonaparte entrusted to the divans much of the local administration such as market control, provisioning and Nile navigation. He confirmed the existing civil rights of Egyptians and made use of the traditional taxing machinery. He failed however to wean the Egyptians from their rooted aversion to accepting responsibility, and was forced against his will to employ Turks in executive posts. He paid deference to Islam and, though he employed Copts and Greeks, he showed them no preference. In Cairo the French savants set up Arabic and French printing presses and, in imitation of the Institut National of Paris, founded the Institut d'Égypte. Its members studied the antiquities, languages, agriculture and medicine of Egypt, surveyed a trace for a future Suez canal and assisted the soldiers in administrative duties. Learned Egyptians visited the meetings of the Institut, inspected the Arabic press (the first in Egypt, which Napoleon had stolen from the Vatican) and watched chemical experiments. The result of these scientific labours eventually appeared in the *Description de l'Égypte*, published in Paris (1809–28). This great work, no less than an encyclopaedia of Egypt, awoke the serious interest of Europe in the splendours of the Egyptian past and inspired the study of Egyptology and of Islam.

The military success of Bonaparte's adventure in Egypt was from the first compromised by the weakness of his position. The sinking of the French fleet by Nelson at Abukir (Abu Qir) on Aug. 1, 1798, placed his only line of communication with France at the mercy of the British navy. This and the failure of Bonaparte's Syrian campaign of 1799 exposed the precarious French hold on Egypt. The Mameluke leaders had fled to upper Egypt but were still dangerous. A column under Gen. L. C. A. Desaix de Veygoux ascended the Nile to Aswan in March 1799 in unsuccessful pursuit of Murad Bey. Several insurrections broke out in Cairo and were suppressed only after desperate street fighting. Napoleon left Egypt on Aug. 22, 1799, and handed over the command to Gen. J. B. Kléber who, on Jan. 24, 1800, signed a convention at Al 'Arish with Rear-Adm. Sir Sidney Smith by which the French were to evacuate Egypt. On the refusal of Adm. Lord Keith to ratify the convention, Kléber reopened hostilities, defeating an Ottoman army at Heliopolis. His assassination on June 14, 1800, left Gen. J. F. (Baron de) Menou in command. As the Ottoman forces were unequal to the task of dislodging the French, they were strengthened by a British force under Gen. Sir R. Abercromby which landed under fire at Abu Qir on March 8, 1801. In a surprise attack by Menou on the British camp on March 21 the French were repulsed and Abercromby was mortally wounded. The Anglo-Turkish allies then took Rosetta (Rashid). On May 31 the British under Gen. Sir J. Hely-Hutchinson (who had assumed the command after Abercromby's death), and the Turks under the sirdar Yusuf Ziya Pasha, attacked Gen. (Count) A. D. Belliard in Cairo. In the Red sea a combined British naval and land force from India put a detachment ashore at Suez, and from Koseir (Al Qusayr) a column led by Gen. Sir D. Baird performed the remarkable feat of marching over the desert to Qina on the Nile in May–June 1801 and from there descending the river in vessels provisioned by their Mameluke allies. But before Baird could join Hely-Hutchinson in Cairo, Belliard had negotiated the withdrawal of his troops to France. Belliard's capitulation undermined the resistance of the French at Alexandria where Menou was compelled to accept similar terms. By Oct. 1801 the last French troops had sailed from Egypt.

2. The Years of Anarchy.—On the departure of the French, Egypt relapsed into anarchy as the Mameluke beys returned from hiding. The Ottoman sultan, Selim III, endeavoured to overthrow the beys by having their leaders massacred at Abu Qir where the Ottoman admiral had invited them to an entertainment. Some, including the Sudanese Mameluke Ibrahim Bey al-Sinnari, were killed; some, like Othman Bey al-Bardisi, were wounded. The survivors were rescued by the British who compelled the Ottoman governor at Cairo to release the beys whom he had arrested there. In March 1803 the British evacuated Alexandria and Mohammed Bey al-Alfi accompanied them to England to discuss means of restoring the Mameluke power. For the next two years Egypt was given up to civil war and plunder. In Cairo the Turkish governor was attacked by his own Albanian troops and fled to the coast. Tahir Pasha took over but was soon killed. One of the commanders of the Albanian formations, Mohammed Ali (*q.v.*), now for the first time came into prominence as an ally of the Mameluke leaders against the Ottoman nominal governor, Khosrau Pasha, whom he took prisoner at Damietta. A new governor from Constantinople, Ali Pasha Jazairli, landed in Egypt but soon, finding himself without support, surrendered to the beys and shortly afterward died either in a fight with his own men or by assassination. The return of Mohammed Bey al-Alfi from England in Feb. 1804 further complicated the situation by splitting the beys into two factions. Months of violent confusion followed during which power shuttled between the beys, the Ottoman governor and the uncontrollable soldiery while the inhabitants of Cairo were subject to oppression. On May 13, 1805, in an attempt to end the impasse, the ulema (Muslim scholars) of the city approached Mohammed Ali and asked him to be their ruler. The Sublime Porte suspected the loyalty of Mohammed Ali and had already designated him as vali (governor general) of Jidda in Arabia in order to remove him from Egypt, but now conceded the wishes of the notables of Cairo and confirmed him in the governorship of Egypt; on July 10 he received a firman of investiture. His authority however was unchallenged only in Cairo; outside the city his title was everywhere disputed by the beys. To eliminate Mameluke opposition he had a number massacred in Aug. 1805. Surrounded by treachery he managed not only to survive but to outwit his enemies one by one. When in March 1807 a British force under Gen. A. M. Fraser landed at Alexandria in support of the Mamelukes, Mohammed Ali deprived them of their ally by an astute promise to the Mamelukes to meet their demands if they would join him in expelling the invaders. A British detachment was ambushed in Rosetta and retreated to Alexandria with heavy loss while a second British attack on Rosetta failed. On Sept. 14 Fraser evacuated Alexandria.

So long as the beys continued to enjoy power Mohammed Ali was not master in his own province. For several years he tried a policy of clemency and concessions with the result that many Mamelukes settled permanently in Cairo. But their inveterate love of violence and aversion to ordered government convinced him that nothing short of their extinction would bring security to Egypt. On March 1, 1811, he invited all the leading Mamelukes then in Cairo to an investiture of his son Tusun with the command of the Egyptian expeditionary force which was about to take the field against the Wahhabi rebels in Arabia. Once inside the citadel the gates were shut and a heavy fire was opened on the Mamelukes from which few, if any, escaped. The massacre in Cairo was the signal for a general slaughter of Mamelukes throughout Egypt. A remnant fled to Nubia whence Mohammed Ali's son Ibrahim dislodged them from Qasr Ibrim. Fleeing southward the survivors set up a short-lived bandit state in Dongola. The Mameluke power in Egypt was at last broken.

Amid all the viciousness of the Egyptian political scene during the last years of Mameluke domination it is pleasing to notice the beginning of a small, apparently spontaneous, revival of letters. Sheikh Hasan al-Jabarti worked in mathematics and astronomy; al-Sayyid Mohammed Murtada al-Zabidi studied theology and grammar; Hasan al-Attar, sheikh of Al Azhar, was one of a literary circle. Best known to latter-day Europe was the chronicler Abd

al-Rahman al-Jabarti (1754–1822), a distinguished member of the Cairo ulema, an intelligent observer of the French occupation (Napoleon nominated him to the grand divan in Cairo) and a fearless critic of Mohammed Ali's earlier regime.

3. The Era of Mohammed Ali.—During his long rule (1805–48), the great vali strived to transform Egypt. At first the pasha gave no indication that he would act differently from any of the other old-time Ottoman governors, but his ascendancy, marked by ruthless despotism and military expansion, was also noted for economic and cultural development.

Arabia and the Sudan.—In 1811, after repeated requests from the Porte, Mohammed Ali dispatched his second son Tusun with an army against the Wahhabis of Arabia, reforming zealots whose iconoclastic zeal had caused them to rise against their sovereign, occupy the holy places, plunder pilgrims and destroy tombs. Tusun, a mere boy of 16, mismanaged the campaign and Mohammed Ali himself went to the Hejaz and took personal command of operations. The campaign ended in 1815 with a treaty with the Wahhabi leader Abdullah ibn Sa'ud. As a secondary consequence of his visit to Mecca, Mohammed Ali was henceforward accorded the pilgrim's title of *al-hajj* (he who has made the pilgrimage). Abdullah failed to honour the treaty to Egyptian satisfaction and in 1816 Mohammed Ali renewed hostilities against the Wahhabis. Ibrahim, his eldest son, took command and the campaign, though arduous, was a complete success. The Egyptian army marched across the Arabian desert from the Red sea to Abdullah's capital at Dar'iyah which Ibrahim took after a long siege in 1818. It is noteworthy, in the light of the subsequent reorganization of the Egyptian armed forces under Mohammed Ali, that the siege artillery was in the charge of a French veteran of Napoleon's Grande Armée, J. M. F. Vaissière. Abdullah was sent a prisoner to Constantinople where, notwithstanding Ibrahim's promise of safety, he was put to death. Mohammed Ali next decided to extend his influence into the northern and central Sudan (*see* SUDAN, REPUBLIC OF THE) which was then in a condition of disorder. After reconnaissances in Nubia by Mohammed Bey Khosrau, the defterdar, in 1818, and by himself in 1819, Mohammed Ali set in motion two invading columns. The first, led by his third son, Ismail, marched up the Nile and entered Sennar, capital of the decaying Funj sultanate, in June 1821; the second, under Mohammed Bey, crossed the Bayuda to Kordofan and occupied El Obeid (Al Ubayyad) in August.

Reform of Armed Forces.—Mohammed Ali had his first experience of active service as a commander of Albanian levies in the Ottoman forces fighting against the French in Egypt. He had observed the decisive superiority of European armament and military discipline when matched against undisciplined Ottoman and Mameluke armies. But military discipline did not come easily to the old-fashioned soldiers. His own sovereign, Sultan Selim III, had lost his throne and his life in a premature attempt to introduce European military methods at Constantinople, and in 1815 Mohammed Ali was faced by a mutiny among his Albanian troops when he tried to do the same. Napoleon during the Egyptian campaign had contemplated forming an army of Negro slaves from Darfur and training them on the European model; Mohammed Ali's invasion of the Sudan undoubtedly had as one of its objects the capture of slaves for replenishing the Egyptian army.

The first problem that faced him was the formation of an officer class distinct from the other ranks of the army, a distinction unknown in the old Ottoman and Mameluke formations. For this purpose young Mamelukes were employed. The instructors were mainly French; a French officer, O. J. A. Sève, another veteran of the Napoleonic Wars, was appointed chief instructor. Training was begun at Cairo but the Muslim recruits objected to receiving instruction from a non-Muslim, and Sève (who became a Muslim) was transferred with his cadets to a secluded camp formed at Isna in 1820. By 1821 the first Negroes captured in the Sudan were being sent down to another camp formed at Aswan to be trained in soldiering by the Mameluke officers from Isna. The supply of Negroes was disappointing. This, coupled with the heavy losses by disease of Turkish and Albanian troops in the Sudan campaigns of 1820–21, led Mohammed Ali to conscript Egyptian peasants who unexpectedly made good infantry. Instructors from the new military school provided the beginnings of a general staff which in its turn formed the nucleus of a war department.

The creation of a navy was the logical corollary to the existence of an army. In 1812 the Egyptian navy in the Mediterranean consisted of a single armed brig. The admiral, Ismail Pasha Jabal Tariq (Ismail Gibraltar to Europeans), was at Stockholm in 1816 learning Swedish naval techniques. The Greek revolt gave Mohammed Ali a pretext for naval expansion and in 1824 Ismail Jabal Tariq commanded the fleet which transported the army of Ibrahim Pasha (*q.v.*) to the Morea (Peloponnese). (*See* GREEK INDEPENDENCE, WAR OF.) Several of his ships' captains were French. A naval school was formed at Alexandria. After the loss of 28 ships at the battle of Navarino (*q.v.*) the Egyptian navy was rebuilt under the direction of the naval engineer L. de Cérisy and Besson Bey, director of the arsenal.

The forces of the *nizam al-jadid* (new model army) were turned to good account in the victories of these European-trained troops against the insurgent Greeks in 1824–27 and against the Ottoman sultan Mahmud II in Syria and Anatolia in 1832–40. In 1821 Mohammed Ali was granted the governorship of Crete which received a small Egyptian garrison. The cession of Syria and Adana by the sultan, as a result of the convention of Kutahya (1833), made Mohammed Ali ruler over a vast region extending from the Euphrates to the Blue Nile.

Structure of the Government.—Mohammed Ali revolutionized the structure of the Egyptian government, giving the various organs of state a recognizably modern form. Although he retained all effective control in his own hands, he developed the use of councils and encouraged deliberation and the taking of responsibility by his subordinates. Of his numerous councils the more important were the divan of home affairs (in part a legacy from Mameluke days), a forerunner of the ministry of the interior; the two judicial divans, and the war divan from which arose the future ministry of war. To these he added technical organs which were novelties in Egypt: the divans of commerce and education, and the directorates of the arsenal, printing works and dockyard at Bulaq. Co-ordinating these departments was the vali's cabinet council, the *ma'iyet-i seniye*. To keep his officials aware of what the various departments were doing, a bilingual official gazette in Turkish and Arabic, *al-Waqa'i'al-Misriya*, was first printed in Dec. 1828. The language of administration at the higher levels was Turkish. An indispensable adjunct of the European type of government, a printing press, was established at Bulaq in Nov. 1821. This renowned government press printed not only official publications, forms and returns but also works of learning in Arabic, Turkish and Persian. Mohammed Ali regarded education primarily as a means of fitting young men for the public service. Those who needed to learn skills which could not be taught in Egypt were sent abroad; students destined to professional careers were usually sent to Paris. In the choice of candidates for posts in the government services Mohammed Ali selected Muslims and Christians indifferently. Medical studies were in the hands of the French physician A. B. Clot (Clot Bey), who came to Egypt in 1825 and founded a medical school and hospital at Abu Za'bal, afterward transferred to Qasr al 'Aini.

Economic Policy.—From 1816 a serious effort was made by means of forced labour to prevent the Nile from overflowing its banks by building protection works. In 1819–20 the Mahmudiyah canal, linking the port of Alexandria with the western arm of the Nile delta, was dug by the same means. Ships of light draft could now reach Cairo from the coast in sheltered water. A department of public works was established in 1835 and placed in charge of L. M. A. Linant de Bellefonds as director. A barrage over the Nile 16 mi. N. of Cairo was begun in 1834 by a party of French engineers, adherents of the school of Claude Henri de Rouvroy de Saint-Simon, but the construction was faulty and the barrage was not put into effective working order until after the British occupation.

A privately owned postal service, the Posta Europea, was founded by an Italian, C. Murati, and a semaphore telegraph,

originally installed by Napoleon, connected Alexandria with the capital.

Mohammed Ali's economic measures were frequently in conflict with the provisions of commercial treaties made by the Powers with the Porte. During the Peninsular War he had sold grain to the British in spite of a prohibition from Constantinople against its export from Egypt. Beginning in 1808 he confiscated a large number of estates, paying inadequate compensation to their owners, and put these lands to profitable use. Later he declared a state purchasing monopoly on cotton, gum arabic and indigo which he imposed on Egypt and the Sudan. The state bought these crops from the cultivators at its own price and sold them at a profit on the Alexandria quayside. Mohammed Ali continually experimented with new crops and sought to improve the quality of those already being grown. He ordered superior cotton seed from the Sudan-Abyssinian border and this, reared by the French cotton expert, L. A. Jumel from 1820, was the ancestor of the famous Egyptian long-staple cotton. Less successful were most of Mohammed Ali's industrial enterprises which languished through the lack of capable managers and operators, and through inadequate planning.

The growing productivity of Egypt had its effect on the growth of the port of Alexandria where 35 foreign commercial houses were established by 1820. Soon after the end of the Napoleonic Wars nearly every state in Europe and the United States had a consul general resident at the port. These, though subordinate to their respective diplomatic representatives at the Porte, were necessarily concerned with diplomatic as well as consular matters; their dispatches showed a lively if critical interest in the new Egypt. About 1835 the European community in Alexandria instituted an improvements committee (*majlis al-tanzim*), the forerunner of the municipal council, called by the Levantines the "Ornato." It was a mixed body of Europeans and Egyptians, the former in the majority. It employed a civil engineer and had power to enforce building and sanitary regulations, and to maintain streets. Sir J. Bowring in his report on Egypt in 1840 commended the work of the "Commission of the Ornament." Suez grew in importance with the increasing Egyptian commitments in the Red sea. From 1835 an overland route between Europe and the east, pioneered by Lieut. T. Waghorn, was operated by means of steamers between Alexandria and Cairo and by carriages from there to Suez. The French established a steamship line between Marseilles and Alexandria in 1835 and shortly afterward the East India company began a service of steamers between Suez and Bombay.

The last years of Mohammed Ali were years of peace. Though shorn of Crete, Syria and Adana by the intervention of the Powers and imperial firmans of Feb. 13 and June 1, 1841, he was accorded the hereditary rule of Egypt and the possession, for his lifetime only, of Darfur (which Mohammed Ali did not possess), Nubia, Kordofan and Sennar. Another firman, dated May 23, extended Mohammed Ali's hereditary rule to the Sudan. In 1846 he secured the lease, renewable annually and for his lifetime, of Sawakin and Massawa. The state trading monopoly was dropped as Egypt was no longer able to defy international trade agreements with the Porte, though traders complained that the monopoly in certain commodities remained in the Sudan. Toward the end of 1847 the old pasha sank into senility and in Sept. 1848 the Porte appointed his son Ibrahim vali in his stead. Ibrahim died in November; his father lingered another eight months, dying in Alexandria on Aug. 2, 1849, aged 80.

4. Abbas I and Said (1848–63).—Under Abbas I (*q.v.*), the son of Tusun, who succeeded his uncle Ibrahim in Nov. 1848, there occurred a general running-down of the machinery of administration in almost all its branches. Within a few weeks of his accession Abbas discontinued the printing of the official gazette on the ground that it was superfluous, and he issued an instruction which defined, and therefore limited, the occasions on which governors and heads of departments need consult him. Abbas was by temperament anti-European but he was in no sense anti-Christian for while he discharged many European experts (including A. B. Clot), he retained Armenian secretaries in high office and continued Mohammed Ali's policy of extending the opportunities of Copts in the civil service. Abbas lived much in seclusion. He was murdered in July 1854 in circumstances which were never satisfactorily explained, and was succeeded by his uncle, Mohammed Said Pasha, a son of Mohammed Ali.

Said, a hearty, lazy, tolerant ruler, who spoke French and cultivated French friends, possessed all Abbas's weakness for delegating responsibility to others, but he signally lacked Abbas's keen sense of the value of money. In 1856 he granted Ferdinand de Lesseps (*q.v.*) a concession for the construction of the Suez canal, a project doggedly opposed by the British government under the influence of Lord Palmerston, whose opposition at the Porte had delayed the ratification of the concession for two years (*see* SUEZ CANAL). Said obtained permission from Constantinople to offer concessions also for the development of telegraphs and for a steamship company, the Medjidieh (Majidiya) company, founded in 1857. In the Sudan, which he visited in the winter of 1856–57, he pursued a policy of decentralization. He made a contribution —more symbolic than real—toward constitutional development by turning the five directors of the interior, marine, education, public works and commerce into ministers, the general direction being given by a privy council (*majlis khususi*), an embryo cabinet. Like Abbas, Said showed loyalty to his sovereign. In 1854 he reinforced the Egyptian contingent which Abbas had dispatched to join the Ottoman army in the Crimean War. But whereas Abbas left the Egyptian treasury practically free of debt, Said died on Jan. 18, 1863, leaving a deficit of more than £3,000,000.

5. Ismail (1863–79).—Said was followed in office by his nephew Ismail Pasha (*q.v.*), a son of Ibrahim Pasha. With Ismail began a period of accelerated economic development which, though it involved Ismail in ruin and led to the British occupation, was not without benefit to Egypt. Ismail was educated in France and was familiar with the exteriors of European culture. Throughout most of his term of office he strove for autonomy, the expansion of Egyptian influence in Africa, and the westernization of Egypt. Yet he remained in some ways an Oriental despot resentful of criticism. In 1866, in return for doubling the Egyptian tribute, he received from the sultan the right of primogeniture in his own family and, in return for liberal gifts in Constantinople, was accorded by the Porte in 1867 the title of khedive (viceroy), a princely honorific of Persian origin which, in spite of its magnificence, conferred no fresh privilege on the holder. In 1873 he was granted virtual autonomy inside Egypt.

Ismail greatly increased the military establishment that had sufficed for Said. He strengthened the field forces by buying Remington breech-loading rifles and Krupp artillery. For the harbour defenses of Alexandria he ordered great muzzle-loading Armstrong cannon. With the object of building up an efficient general staff, he enlisted a group of officers from the United States who had fought in the American Civil War and employed them not only on regular staff duties but also on topographical work and exploration in the Sudan. Brig. Gen. C. P. Stone of Massachusetts was appointed chief of staff. A notable map of Africa, prepared by the general staff and embodying the latest geographical discoveries by khedivial and other explorers, was prepared in 1877.

Ismail's expansionist policy was on the whole unsuccessful. A firman of 1865 ceded to Egypt the ports of Sawakin and Massawa which had reverted to the Ottoman government on the death of Mohammed Ali. Ismail saw clearly the importance of the Suez canal, opened in 1869, which turned the Red sea into a strategic waterway. On his instructions the governor of Massawa laid claim to the hinterland while in 1870 Zaila (Zeila) was occupied by an Egyptian garrison and Ismail finally claimed the Somali coast as far as Cape Guardafui (Ras Assir). In 1875 an Egyptian force occupied Harar. In the region of the equatorial lakes Sir Samuel White Baker was commissioned to annex the territory south of Gondokoro. During 1869–73 he extended the Egyptian sway to the shores of Lake Albert and within 100 mi. of Lake Victoria.

The sultanate of Darfur in the Sudan, until then an independent state, was overrun in 1874 by a private army belonging to Zubayr Pasha Rahma Mansur with the co-operation of Egyptian govern-

ment troops commanded by the governor general of the Sudan, and was annexed to the Egyptian dominions. In 1876 the Equatorial province was in part abandoned by Col. (later Gen.) C. G. Gordon (*q.v.*), Baker's successor. In the previous year, at Gordon's suggestion, Ismail sent a squadron carrying an Egyptian force which landed at Kismayu and claimed the coastal area. The British government protested on behalf of the sultan of Zanzibar who also claimed the territory, and the Egyptian expedition was withdrawn. Ismail's aggressive policy toward Ethiopia (encouraged by the governor of Massawa, J. A. W. Munzinger, a Swiss) resulted in two crushing defeats at the hands of the Ethiopians, at Gundet in 1875 and at Gura in 1876. Ismail's last military intervention was a fiasco. The Egyptian contingent sent to aid the Ottoman defense of the Danube against the Russians in 1877–78 was of poor fighting quality and was ignominiously withdrawn.

Ismail showed a surer hand in planning economic projects than in executing them; he suffered much from his own naïveté and impetuosity and from the incompetence of his officials. Egypt experienced a cotton boom during the American Civil War, but Ismail's grandiose plans for increasing the cotton yield of the Sudan went astray. He was much occupied by railway surveys in the Sudan and by engineering projects to lessen the obstacles to Nile navigation; the only practical outcome was the opening of 30 mi. of railway from Wadi Halfa to Sarras in 1875. On the other hand the railway system in Egypt was greatly extended. In 1864 the Egyptian government bought out the Posta Europea and turned it into a public service which in 1867 issued the first Egyptian postage stamps. The Aziziya Misriya Steamship company, founded in 1863 as a continuation of the former Medjidieh company, was ruined by bad management; in 1870 the government took over the liabilities and operated the service as a state company under the name of the Khedivial Mail line.

Egypt lay outside the range of the Ottoman laws, collectively known as the *Tanzimat*, a series of measures promulgated from 1839 onward with the intention of reforming and modernizing the structure of the Ottoman central and provincial government. Mohammed Ali had already anticipated much of this legislation, particularly in the reorganization of his army. In one detail, however, Ismail's Egypt lagged behind the progress made elsewhere in the Ottoman empire. In 1864 the Porte issued the vilayets decree and in 1877 a law of municipalities, enactments which introduced a radical reform of the provincial government and created provincial and municipal councils whose members were drawn partly from officials and partly from prominent citizens, with representation for the non-Muslim communities. Ismail granted municipal privileges only to Alexandria. In 1868 a preparatory commission under the chairmanship of an Italian resident, A. Colucci Bey, drafted municipal regulations which were approved by the minister of the interior and on June 3, 1869, were issued with government approval.

Ismail left his mark upon urban development. He founded the Ismailia quarter in Cairo and provided it with modern services. Piped water and gas were introduced in Cairo and the larger towns. It is significant that he did not permit municipal experiments in his capital where the technical services continued to be operated by the various ministries of the central government under the executive supervision of the city governor. The technical officers charged with the urban public services were usually French. These radical developments, which converted portions of Cairo, Alexandria, and later Port Said, into European-style quarters, left untouched the problem of the slums in which the great majority of the city dwellers lived. The breakdown of the medieval order and the disappearance of the guilds had left the poor without any form of corporate social protection. Critics of the regime accused the government of skimping on the elementary sanitary services in the poorer quarters where the masses existed in circumstances of incredible squalor. In 1870 the Egyptian government appointed J. M. Cook, son of the founder of the firm of Thomas Cook and Son, as passenger agent on the Nile. Largely by his efforts Egypt was becoming an important tourist resort. In 1872 a modest spa was established at Hulwan (Helouan).

Ismail had no mind for constitutional advance of the kind proposed by liberal thinkers in western Europe. His timid approach toward constitutional government stopped with an organic decree issued at the end of 1866 setting up a chamber of notables (*majlis al-niyabi*), a purely consultative body inaugurated on Jan. 1, 1867, and thereafter known as the legislative council. Even this chamber owed its existence rather to the practical needs of the treasury than to any altruistic design of associating the people with the government, for Egypt was ill-placed for constitutional development on European lines. Public life was monopolized by a small ruling body, Turkish- and French-speaking, superficially Europeanized and having only an economic connection with the illiterate peasant masses who spoke only colloquial Arabic and whose lives were unaffected by modern change. The evolution of Egyptian social life was, however, bringing into existence an urban middle class, to some extent politically conscious but not yet politically articulate. In Ismail's reign the first stirrings of a popular nationalist movement can be observed, a movement which came to a head in the revolt of Arabi Pasha (*q.v.*) in 1882. Egyptian nationalism was not simply the immediate product of economic and social discontent. The cultural revival of a small school of historians and literary men at the end of the 18th century was followed in Mohammed Ali's time by an era of printing of Arabic works and of translating into Arabic from European languages, mostly French, a labour headed by Rifa' Bey al-Rafi'i who with his circle combined a traditional Islamic background with a French education. Under Said and Ismail learned institutions proliferated in Cairo: the Egyptian museum (1857), the Institut d'Égypte (1859), the Egyptian library (Dar al-kutub al-misriya, 1869) and the Khedivial (afterward National) Geographical society (1875).

Egyptian schools were unable to meet the expanding demands of the government, the professions and commerce for educated personnel. As under Mohammed Ali, Egyptian students were sent to universities and technical institutions in Europe. The influx of Syrians, first noticeable under Mohammed Ali, increased; the newly founded colleges in Beirut were then beginning to produce graduates who were to render Egypt a signal service in journalism, publishing and letters. To Said and Ismail the Christian and Jewish communities in Egypt owed their full integration into Egyptian life.

In 1867 Nubar Pasha Nubarian, foreign minister at the time, submitted, with Ismail's support, a proposal to reform the consular courts which, under the regime of the Capitulations, tried cases in which foreigners were involved. In 1876, after a long and costly diplomatic battle with the 14 interested Powers, and with the Porte which objected to Ismail's direct dealings with foreign states, a new code, based on the French Code Napoléon, was introduced. At the same time the consular courts were replaced by mixed courts on which Egyptian judges sat with their foreign colleagues.

Bankruptcy.—The khedive's financial prodigality and general mismanagement had by 1875 so impoverished the country that the ordinary revenues no longer sufficed to repay the mounting interest on the foreign loans. Government credit was so low that no further loans could be raised on the European market. Ismail in desperation sold his Suez canal shares to the British government for nearly £4,000,000. Lobbied by the creditors the Powers compelled Ismail to admit a series of international investigations into his financial affairs. In May 1876 the Egyptian finances were placed under international control through the agency of a Caisse de la Dette Publique (commission of the public debt), set up as a result of a British investigation by S. Cave and Sir J. Stokes, while in November an Anglo-French mission of George (afterward Lord) Goschen and Joubert on behalf of the creditors led to the establishment of dual control whereby the Egyptian revenue was supervised by a British official and the expenditure by a French official. In 1878 another international commission, of which Major Evelyn Baring (afterward Earl Cromer) was a member, extended international control to the khedive's personal estates. Ismail accepted a compromise solution in the form of a constitutional ministry under the presidency of Nubar Pasha, with Sir C. R. Wilson as finance minister and E. de Blignières as

minister of public works. The working of the new control organization provoked unrest in the army, and Ismail, in a gesture of defiance, made his last stand; on April 7, 1879, he dismissed his ministers. Britain and France appealed to the sultan who, on June 26, sent a curt telegram addressed to the former khedive informing him that his son, Mohammed Tewfik Pasha (*q.v.*), was appointed his successor. Within a few days Ismail sailed from Egypt, never to return.

6. The First Nationalist Movement.—In 1872 al-Sayyid Jamal al-Din al-Afghani, a pan-Islamic agitator of Afghan birth, came to Egypt and lectured to large audiences in Cairo calling on Muslims to resist the encroachments of the west. In the struggle for the defense of Islam he advocated violence and even assassination, though his young Egyptian disciple, Sheikh Mohammed 'Abduh, sought the renaissance of Islam through education and persuasion. The government grew alarmed. Yaqub Sanua, an Egyptian Jewish journalist, a satirist both of Ismail and the British, who wrote under the pseudonym Abu Naddara ("The Bespectacled"), was banished in 1878. In that year Jamal al-Din and Mohammed 'Abduh founded al-Hizb al-watani al-hurr (Free National party), which many prominent soldiers, officials and teachers joined. But pan-Islamic propaganda was distasteful to the government and to the governing body of Al Azhar university. Jamal al-Din was expelled from Egypt in 1879; Mohammed 'Abduh followed him to exile in 1882 and the two men moved to Paris and published a journal, *al-'Urwat al-Wuthqa* ("The Indissoluble Bond"), devoted to criticism of British policy in Muslim lands and to the rebirth of Islam. Though banned in Egypt and in India, the journal during its short life had a wide circulation.

When Anglo-French control was re-established shortly after Tewfik's accession, it seemed to many Egyptians that the whole population was being punished for Khedive Ismail's personal shortcomings. Foreign control of the country's finances, however justly administered, was after all being exercised in the interests of the bondholders. The growing resentment, none the less intense for being irrational, was taking three forms: resentment by less-privileged Egyptians, notably junior army officers, against the "Circassians," *i.e.*, the influential Turkish-speaking minority, for blocking their way to promotion; resentment by Muslims against domination by Christian powers, and resentment by all articulate Egyptians against any form of foreign control. All three currents of resentment were present in the rising of 1882. Arabi Pasha (*q.v.*), popular leader of the revolt, was a colonel of peasant origin, who rightly called himself "the Egyptian," while Brig. Gen. Mahmud Sami al-Barudi Pasha was a Circassian of the ruling body, a cultivated man and a poet. Unrest in the army, which first occurred in an officers' mutiny in 1879, spread during 1881. Arabi grew increasingly popular with troops and people as the living symbol of the Egyptian protest against foreign control whether Circassian or European, which the weak government of the khedive had neither the means nor the will to resist. A nationalist ministry headed by Mahmud Sami and with Arabi as minister for war was forced on the khedive early in 1882. An organic decree of Feb. 7 set up a chamber of deputies whose members were to be elected for a term of five years at an annual salary of £E100. The official language was to be Arabic. Tewfik Pasha was unable to contain the movement which now threatened the international control of Egypt. The danger of a serious rising brought the British and French fleets to Alexandria in May. A massacre in the city in June and the defiance of the nationalist commander led to the British bombardment of the forts on July 11. The British government decided to suppress the revolt in the interior and invited the governments of France and Italy to co-operate. Both declined. A British expeditionary force commanded by Sir Garnet (afterward field marshal Lord) Wolseley landed at Al Isma'iliyah (Ismailia) in the canal zone, and at Tall al Kabir (Tel-el-Kebir) on Sept. 13 crushed the main army of Arabi who surrendered shortly afterward in Cairo. The khedive who had taken refuge in Alexandria returned to Cairo and a ministry agreeable to Britain, led by Mohammed Sharif Pasha, was formed. Arabi was tried and was banished to Ceylon with the leaders of the revolt.

7. British Occupation.—Lord Dufferin, British ambassador at the Porte, who had been sent to Egypt as high commissioner, laid down in a report published on March 20, 1883, the general principles by which Egypt was to be administered. Sir Evelyn Baring returned to Egypt in Sept. 1883 as British agent and consul general in succession to Sir E. Malet who had held the post since 1879. To Baring fell the task of carrying Lord Dufferin's principles into administrative practice. A week after Tall al Kabir, the Egyptian army was disbanded and in Dec. 1882 Sir Evelyn Wood, appointed sirdar of the new force, had begun the formation of an army of 4,000 men, a number afterward much increased. The commanding officers were mostly British while the other officers were British and Egyptian, some battalions having British, and some Egyptian, commanders. The men were recruited by ballot and the conditions of service were improved. The new Egyptian army, which attracted an élite of British officers, was smaller and more efficient than the old army. It was to prove its worth in the coming campaigns against the Mahdists and in small wars in the Sudan. The British army of occupation, limited to 14,000 in 1883, was afterward at various times increased and decreased.

An organic law of May 1, 1883, drafted under Lord Dufferin's supervision, created in each province a consultative council having from three to eight members according to the size of the province. The councils were to meet under the chairmanship of the governor who would consult the members on such regional matters as markets, roads and irrigation works. In the capital the same law established two chambers. The first was a legislative council of 30 members of whom the president and 14 members were to be appointed by the government, 14 elected by the provincial councils from among their members, and two by the cities of Cairo and Alexandria. The second was a legislative assembly of 82 members comprising the khedivial ministers, the 30 members of the legislative council and 46 elected members. The statutory function of the legislative council was to examine bills. All laws and ministerial regulations, as well as the annual budgets of the state, had to be submitted to it, though it had no power of veto. No new direct taxes could be sanctioned without its assent. Press and public were excluded from the sessions of both chambers.

The persistent obstruction by the British government to the construction of the Suez canal had given way to an equally tenacious opposition by both government and shipping interests to the predominantly French composition of the Suez canal board. By an agreement signed in London on Nov. 30, 1883, the British representation on the board was increased from 3 to 10 in comparison with 22 French.

Britain and France supported the demands of the creditors as a first charge upon the Egyptian revenues. Baring did his best to maintain the finances on an even keel but his difficulties were increased by the lack of consistent guidance from the British government which was eager to withdraw the British troops as soon as possible. The triumphant Mahdist revolt in the Sudan upset all financial calculations. Mohammed Sharif Pasha's ministry urged the reconquest of the Sudan (*q.v.*), but Baring vetoed the proposal on the ground that Egypt had neither troops nor money to carry it out. The ministers then resigned rather than sign the khedivial decrees for the abandonment of the Sudan. The surviving garrisons were withdrawn through Ethiopia in the course of 1884–85 and the entire Sudan south of Wadi Halfa was evacuated, but Sawakin continued to be held. Nubia, as far south as Wadi Halfa, was placed under a military administration.

As soon as possible Britain put an end to the international regime; the dual control ceased on Jan. 11, 1883. As the Egyptian government was seriously short of funds, it diverted to the treasury certain revenues which should have been paid to the Caisse de la Dette Publique for the benefit of the creditors. The Powers at once protested and the operation was cancelled, but the British government succeeded at last in negotiating the London convention of March 18, 1885, reducing the interest on the debt and empowering the Egyptian government to borrow £9,000,000 on international guarantee. The loan rescued the government from the immediate threat of bankruptcy, and a credit balance of £1,000,000 was profitably invested in improving irrigation. A clause in the

London convention stipulated that, if Egypt found itself insolvent after two years, another international commission would be appointed, an eventuality which Baring bent all his energy to avoid. In Nov. 1885, following an agreement between Sir H. D. Wolff and the Porte, an Ottoman high commissioner for Egypt, Ghazi Ahmad Mukhtar Pasha, was appointed to represent the khedive's suzerain.

Baring directed the re-organization of the Egyptian government in all its branches. The heavy drain on the revenue for the service of the international debt did not prevent his skilful use of available funds for what he considered to be the basic necessity of Egypt at the time: public works, especially irrigation, the foundation of the Egyptian economy. In 1890 the delta barrage, begun in 1834, was at last brought into effective use. The opening of the Aswan dam in 1902 sensibly increased the crop yield of upper Egypt. Baring was compelled to cut down expenditure on other departments of state activity. In 1885 there was a budget deficit of £2,500,000, but by 1887 the danger of bankruptcy, and with it the inconvenience of international control, was past. The budget for 1893 showed a surplus of £4,000,000. The national debt was successfully converted, forced labour was abolished except for Nile protection works when flooding threatened danger to life and crops, and the land tax was reduced in the poorest provinces. The administration of the law was overhauled and a law school established in Cairo. Education, public health and prisons—all were touched by Baring's reforming hand. For the first time the Nile was made the subject of long-term hydrographic studies, accurate mapping and systematic river control. Baring, the principal architect of the new Egypt, was elevated to the peerage as Baron Cromer in 1892. In 1899 he became a viscount and in 1901 an earl (*see* also CROMER, EVELYN BARING, 1st Earl).

The modernization of the larger towns, begun during the reign of Ismail Pasha, was continued under the British occupation. Cairo still lacked an autonomous municipal organization. Some broad avenues were cut through the poorer quarters and an attempt was made to deal with the appalling sanitary problems there, but financial stringency checked more ambitious road-widening schemes. In 1905 Baron Edouard Empain, a Belgian capitalist, laid out a European-type housing estate in the northern suburb of Heliopolis. The first asphalted roads in Cairo were laid in 1906. In Alexandria there was a parallel development. The municipal regime set up in 1869 had become moribund during the revolt of 1882 and the subsequent British occupation. In 1890 a fresh municipality was decreed under which the governor of the town was president of a council of 28 members, half of them elected, with a system of corporate voting for importers and exporters and for property owners.

8. Abbas II (1892–1914).—The khedive Mohammed Tewfik Pasha, who had always maintained co-operative relations with the occupying power, died on Jan. 7, 1892, and was succeeded by his son Abbas Hilmi, a young man who aspired to rid himself of British control (*see* ABBAS II). On Jan. 15, 1893, he dismissed the prime minister, Mustafa Fahmi Pasha, whom he considered excessively pro-British, and appointed Husain Fakhri Pasha, a francophile and a sympathizer with the nationalists, and hence less amenable to British guidance. At Wadi Halfa in Jan. 1894 Abbas made disparaging remarks on the bearing of British officers in the Egyptian army, for which he was forced to make a humiliating retraction. However, a longer acquaintance with Cromer convinced Abbas of the futility of crossing swords with the British representative, and with the easing of tempers Mustafa Fahmi was reinstated on Nov. 12, 1895.

The nationalist movement, quiescent since 1882, now revived. In 1894 Mustafa Kamil, a young lawyer educated in France and influenced by French political ideas, founded the second nationalist party which advocated independence through legal means, and in 1899 started a periodical, *al-Liwa* ("The Standard"), to voice the party's views. Its circulation soon attained 10,000 copies, or almost 20,000 if the French edition, *L'Étendard égyptien*, is included, an enormous circulation in terms of contemporary Egyptian journalism. The new nationalists (who included Copts in their ranks) were less interested in pan-Islam than in secular nationalism. When Arabi Pasha returned from exile in 1899 he found himself almost forgotten by the younger Egyptians.

The Fashoda Crisis and the Sudan Agreement.—By 1896 the British and Egyptian governments were forced to consider reoccupying the Sudan. The European scramble for African territory had resulted in the encirclement of the Sudan by areas occupied by French, Belgians, Italians and British, while the strengthening of Ethiopia by Menelik II, who had destroyed an Italian invading army at Aduwa in March 1896, added a neighbour with territorial claims on Sudanese territory. It was known in Cairo that a French expedition from the direction of the Congo basin would shortly enter the Sudan with the object of occupying the valley of the upper Nile and its tributaries up to the western Ethiopian border. The French held the view that the withdrawal of the Egyptian forces from the Sudan in 1884–85 had constituted a lapse of Egyptian sovereignty, a thesis denied by the Anglo-Egyptian power which maintained that Egyptian sovereignty (more strictly Egyptian sovereignty under the nominal suzerainty of the Ottoman sultan) had never been abandoned and remained intact. In Britain public opinion was deeply moved by the circumstances of the death of Gordon in the defense of Khartoum, an incident which gave rise to an extensive literature advocating the recovery of the Sudan from the Mahdist rebels. Popular emotion was sustained by the appearance of books written by escaped European prisoners from the Mahdist camp (J. Ohrwalder, 1892; P. Rossignoli, 1894; R. C. von Slatin, 1896) and by the advocacy of British officers of the Egyptian army in favour of a forward policy. The decisive factor which determined the issue was not however the strength of British public feeling but the fact that by 1896 the increased stability of Egypt had made the venture practicable.

In 1897–98 a small French column under Comdt. J. B. Marchand occupied Fashoda on the White Nile while a Franco-Ethiopian party marched westward along the Sobat river to attempt a meeting with Marchand. Informed at Omdurman in Sept. 1898 of the arrival of unidentified foreigners at Fashoda, Gen. Sir Herbert (later Lord) Kitchener steamed there with a flotilla flying the Egyptian flag and, in effect, told Marchand that he was trespassing on Egyptian soil. The Fashoda crisis provoked lively feeling in both Britain and France, but Anglo-French moderation prevailed and Marchand retired from Fashoda. By an agreement signed on March 21, 1899, France abandoned its claims in the Nile valley and agreed to the delimitation of a provisional frontier from the Nile-Congo watershed northward to Darfur. The future status of the Sudan was laid down in the Condominium agreement signed by Britain and Egypt on Jan. 19, 1899. By this agreement the British and Egyptian flags were to be flown together, the supreme civil and military command was to be vested in a governor general appointed by the khedive on the recommendation of Britain, and the Sudan was to be excluded from the jurisdiction of consuls and of mixed courts. The Porte protested against the agreement as a violation of the sovereign rights of the sultan.

The long period of tension between the British and French governments ended on April 8, 1904, with a declaration by which Britain and France each recognized the paramountcy of the other in Egypt and Morocco. A khedivial decree issued on Jan. 1, 1905, with the consent of the powers recognized Britain's special position in Egypt without however admitting a formal British protectorate. By this decree Egypt achieved financial independence under British guidance. The power of the Caisse de la Dette Publique was limited to the receipt of revenue from that part of the land tax which was pledged by the decree to the service of debt repayment. Accumulated surpluses to the amount of £10,000,000 in the hands of the Caisse, after meeting the charges of the debt, were remitted to the Egyptian treasury. The decree left untouched the complicated question of the Capitulations in the matter of extraterritoriality in criminal cases enjoyed by Europeans residing within the Ottoman empire, and in respect of legislation whereby no change in any law applicable to certain foreigners could be made without the unanimous consent of no less than 15 powers. This was a state of affairs incompatible with the constitutional po-

sition of Egypt at the time and a flat denial of even an approach to national sovereignty.

The Tabah Incident.—The Egyptian border in the Sinai peninsula had never been fixed. From 1841 to 1892 south Sinai was not administered by anyone. The Porte allowed Egypt to put a police post there to protect Egyptian pilgrims. In 1892 the firman of investiture of the khedive Abbas omitted any reference to Sinai as being within the boundaries of Egypt and, after much correspondence, the Ottoman government accepted the *status quo* as in the time of Tewfik Pasha, that south Sinai was Turkish property administered by Egypt. The Ottoman government then stated that, as Egypt no longer used the desert road but sent its pilgrims by sea, it would resume possession of the region and, in Feb. 1906, a Turkish patrol occupied Tabah, a village on the west coast of the gulf of Aqaba. Cromer appealed to Abbas to claim south Sinai as Egyptian territory which the khedive declined to do as he was convinced that Egypt had no claim. Strong British representations were then made to the Porte which gave way, withdrew its troops from Tabah, and agreed to the delimitation of an administrative boundary running in a straight line from Rafah to Aqaba. In an agreement signed by the Egyptian government and the Porte on Oct. 1, 1906, Ottoman sovereignty over the disputed region was recognized.

The Cultural Awakening.—The creation by Mohammed Ali, and especially by Ismail, of schools, printing presses, libraries, newspapers and state translation bureaus, made possible a revival of Arabic letters. The subsequent British occupation stimulated the patriotism of Egyptians, for instead of seeking merely to reproduce an imitation of European culture, they developed a pride in their own national heritage. A remarkable feature of the new movement was the part played by young Syrian immigrants in both literature and publishing. The newspapers *al-Ahram* (1876) and *al-Muqattam* (1889) were founded by Syrian Christians, as were the periodicals *al-Muqtataf* (transferred from Beirut to Cairo in 1884) and *al-Hilal*, the last by Jurji Zaidan, one of the pioneers of the Arabic historical novel. Poetry was represented by Ahmad Shauqi (*q.v.*), Hafiz Ibrahim and Khalil Mutran; the essay by the stylist Mustafa Lutfi al-Manfaluti, the novel by Jurji Zaidan. Mohammed 'Abduh, a moderate in his political views, had returned to Egypt in 1888. As mufti of Egypt from 1899 he set about introducing reforms in the Muslim academy of Al Azhar against massive opposition from the conservative professors who, much as they disliked the British occupation, disliked the spirit of western civilization still more. Mohammed 'Abduh and the more forward-looking intellectuals both religious and lay founded the *salafiya* movement which worked for a return to the purity of primitive Islam and a restatement of Islamic teaching in terms of modern science which, they held, was not an enemy but an ally of religion. The organ of the movement, *al-Manar* ("The Lighthouse"), started in 1896 by Sheikh Mohammed Rashid Rida, a disciple of Mohammed 'Abduh, had a powerful influence in Egypt and throughout the Muslim world. A thinker much influenced by Mohammed 'Abduh was the lawyer, Qasim Amin, whose book *al-Mar'a al-Jadida* ("Modern Woman"), published in Cairo in 1900, caused a sensation by its advocacy of the emancipation of women. An authoress, Malak Hifni Nasif, wrote articles on marriage, polygamy and divorce.

End of the Cromer Regime.—It was inevitable that Cromer's long domination of the Egyptian political scene should have made him the object of fierce criticism by the nationalists who regarded the British occupation as the main hindrance to the achievement of independence. They accused Cromer of failing to carry out the liberal intentions of the Dufferin report of 1883. They pointed to his meagre contribution to Egyptian education (a modest non-government university was founded by Prince Ahmad Fu'ad in Cairo in 1908) and they attacked him for increasing the number of British officials in the government services and illegally changing their role from advisers to administrators. Even Cromer's wisdom and certitude were hindrances to his popularity in Egypt; he invariably explained his policies with arguments which ignored his nationalist critics. He looked upon the nationalists as out of touch with the material needs of the peasants, and in his report for 1907 wrote of nationalism as detrimental to the welfare of Egypt. In June 1906 an affray between villagers at Dinshaway in the delta and a party of British officers on a pigeon-shooting trip was followed by judicial sentences of such severity that the incident provoked indignation in Egypt and questions in the British parliament. In Egypt it served to unite the peasants and the nationalist middle class against the British occupation and in particular against Cromer, who in fact had nothing to do with the sentences

Anglo-Egyptian Relations, 1907–14.—Cromer's resignation in April 1907 and the appointment of Sir J. Eldon Gorst in his place marked the transition from an authoritarian to a more liberal control. Gorst had been adviser to two key ministries, interior and finance, and was experienced in Egyptian public affairs. He had learned Arabic and was on friendly terms with Egyptian officials. The khedive, who was ill at ease in Cromer's powerful presence, was attracted to this milder, more sensitive, man who however lacked Cromer's dignity. In carrying out the British government's intention to liberalize its Egyptian policy Gorst encountered much difficulty. The nationalists who had been temporarily divided after the death of their leader, Mustafa Kamil, on Feb. 10, 1908 (he died a pasha at the early age of 34), again presented a united front and, instead of winning over Gorst as an ally, rejected his advances. Gorst also failed to win the confidence of his more intransigent British colleagues in Egypt. Such measures as were passed during his term of office were useful rather than spectacular, such as enlarging the powers of provincial councils. Between 1907 and 1914 the first peasants' co-operatives were founded by the initiative of Husain Kamil Pasha (the future sultan) and Omar Lutfi Bey with encouragement from the Khedivial Agricultural society. But the growth of a social conscience in Egypt was slow, the legislative assembly was lukewarm and the movement died in the economic upheaval of World War I. In Nov. 1908 Mustafa Pasha Fahmi resigned the premiership, and the appointment of Butrus Pasha Ghali, a Copt, as his successor, was unwise. By a fateful misfortune Butrus Pasha introduced a bill in the legislative assembly for the extension of the Suez Canal company's concession, a measure which that body rejected. Two days later, on Feb. 10, 1910, the prime minister was assassinated, an act which temporarily alienated the Coptic community from the nationalist movement. Gorst died on July 12, 1911. In his report for 1910 he had admitted the failure of his plans for moderate constitutional advance.

The British cabinet decided that a forceful personality was required to deal with the situation in Egypt. Kitchener received the appointment and took over the duties of British agent and consul general in Sept. 1911. On Sept. 29 Italy declared war on the Ottoman empire and Italian forces occupied coastal towns in Tripolitania, on Egypt's western frontier. Egyptian Muslim enthusiasm for the defense of the empire rose high, but the use of Egypt or of Egyptian troops by the sultan in the war against Italy was quietly prevented. In Dec. 1911 Egyptian forces occupied the Gulf of Sollum (Khalij as Sallum) on the evacuation of the Ottoman troops from Cyrenaica, and Sollum was formally incorporated in Egypt on April 9, 1912, thus ending a long-standing frontier dispute with the Ottoman government. During Kitchener's tranquil, and on the whole popular, proconsulship, nationalism returned to quiescence. The pan-Islamic fervour in the early years of Abbas's reign, and again in 1911 during the Italo-Ottoman war, never took firm hold of public imagination, and with the Ottoman military collapse in Tripolitania and the Balkans had become an anachronism. The Anglo-French settlement of 1904 which denied the nationalists further French support, and the Young Turk revolution of 1908 which shook the foundations of the Ottoman caliphate, convinced the Egyptian nationalists that they must seek their independence alone.

Kitchener continued and extended Gorst's program of reform. In 1912 a homesteads exemption law, commonly known as the Five Feddans law, was passed. This measure protected small peasant holdings from seizure for debt. An organic law of July 21, 1913, which replaced the enactment of 1883, constituted a single legislative assembly in place of the two former chambers. The new body

comprised 66 members elected by indirect vote, and 17 members appointed by the government to represent minority interests. While it was still confined to consultative and deliberative functions, it had power to veto any increase in direct taxation, to delay bills and initiate bills of its own, and to record its opinion on administrative business. The president and one of the two vice-presidents were appointed by the government; the other vice-president was elected by the chamber which chose Zaghlul Pasha Saad (*q.v.*), the future Wafdist prime minister, for the post. Khedive Abbas intervened less in public affairs. From the beginning of his term of office Kitchener had treated him with undisguised scorn. Abbas had first favoured, and then quarreled with, the nationalists, and now he fell foul of the legislative assembly. He dismissed the prime minister, Mohammed Said Pasha, and appointed Husain Rushdi Pasha, who held office throughout World War I. (R. L. Hl.)

9. The Protectorate.—When World War I broke out, Kitchener was on leave in England, and Abbas was on a visit to Constantinople. The former became secretary of state for war and never returned to Egypt. The latter was frankly involved with the enemies of England, and was forbidden to return. The Egyptian government immediately on the declaration of war threw in its lot with the Allies. The legislative assembly was suspended and did not meet again. Ministers took action against German ships in Egyptian ports, and by the proclamation of martial law on Nov. 2 it became possible both to enforce administrative measures without reference to the assembly and to take action affecting foreign subjects without first obtaining the assent of the powers. The entry of Turkey into the war rendered the juridical position of Egypt, from the British point of view, an impossible anomaly; and accordingly a proclamation was issued on Dec. 18, 1914, by the British foreign office, with the following announcement:

In view of the state of war arising out of the action of Turkey, Egypt is placed under the protection of His Majesty and will henceforth constitute a British protectorate. The suzerainty of Turkey over Egypt is thus terminated and His Majesty's Government will adopt all measures necessary for the defence of Egypt and protect its inhabitants and interests.

A further proclamation issued the following day announced that the khedive, Abbas Hilmi, who was in Constantinople, was deposed on the grounds of adherence to the king's enemies and that his uncle, Prince Husain Kamil, had accepted the succession and would bear the title of sultan of Egypt. The new status was introduced without disturbance if without enthusiasm. A new high commissioner, Sir Henry McMahon, took over the control of Egyptian foreign affairs and Gen. Sir John Grenfell Maxwell was sent to direct the British war effort. After the abortive Turkish attack on the Suez canal, the prestige of the sultan increased, but the failure in the Dardanelles produced some reaction and in 1915 two attempts were made on his life. The Egyptian people, however, bore with patience the disabilities and unwelcome requisitions which the war entailed, and thus laid Great Britain under obligations, both moral and financial, for which too little credit was given.

On the other hand the withdrawal of British officials for service elsewhere opened a wider field for nationalist propaganda, which found ready listeners among the younger men who had not known the pre-occupation conditions. A dissatisfied civil service regarded the growing number of British officials as a bar to promotion, and the nationalist ranks were reinforced by the students, whose prospect of obtaining state employment was diminished by foreign competition. A professed intention to train Egyptians to manage their own affairs seemed inconsistent with the constant increase of British officials and their steady encroachment on administrative functions. Moreover, as their number grew, they became a separate community living aloof from Egyptians, and with loss of contact their moral influence diminished.

At the end of 1916 Gen. Sir Reginald Wingate, who had been serving as sirdar and governor general of the Sudan since Dec. 1899, replaced Sir Henry McMahon as high commissioner. The declining health of Sultan Husain made a settlement of the succession urgent. His only son, Prince Kemal al-Din, declined the position of heir apparent, which was then offered to the sixth son of Ismail, Prince Fuad, who had been educated at the military school at Turin. Sultan Husain died in Oct. 1917. The principles formulated by the president of the United States toward the close of World War I had a far-reaching effect on educated opinion in Egypt. Numerous disclaimers by British statesmen of any intention to occupy the country permanently were insistently recalled, and the acceptance of the principle of self-determination was represented as having given international sanction to the aspirations of Egyptians to govern themselves. When, in Nov. 1918, an Anglo-French declaration announced that the Allies contemplated the enfranchisement of the peoples oppressed by Turkish rule, Egyptians regarded their title to manage their own affairs to be even stronger than that of Syria, Mesopotamia or Arabia, where an independent kingdom had already been established. No sooner was the Armistice signed than Zaghlul, as the head of a strong nationalist committee, asked permission to go to London and submit Egypt's case for independence. The British foreign office rejected the proposal; and the reception in London of Egyptian ministers, though earnestly supported by Wingate, was dismissed as equally impossible. Ministers thereupon resigned; Wingate was summoned to London to make a personal report, and there was a steady drift of moderates into the nationalist camp. A Wafd or nationalist delegation to proceed to London and Paris was constituted on a democratic basis. The prime minister resigned in final protest against his country's being unrepresented at the peace conference; and the Wafd movement became so serious that martial law was invoked and Zaghlul and three of his most important adherents were arrested on March 8, 1919, and deported to Malta.

Outbreak of March 1919.—These measures brought immediate results which revealed the gravity of the situation. There were anti-British and anti-foreign demonstrations in Cairo and the provinces, railway and telegraph lines were cut and, in upper Egypt, eight British subjects were pulled from a train and murdered by a fanatical crowd. The Egyptian police and, except for a few units in the Sudan, the Egyptian army were unaffected. But these were exceptions. The agitation was supported by every class, including the Copts, and if the movement was only partial among the peaceful, unchanging fellahin, it was quite apparent that even among them the experiences of the war had aroused a spirit of discontent. Thus, even the class which had benefited most from the British occupation was now seriously disaffected and order was only restored, at the end of March, after the vigorous use of British military forces. When this was done, however, Field Marshal Lord Allenby, who had been appointed special high commissioner in the absence of Wingate, adopted a policy of conciliation. His removal of the embargo on the free movement of Egyptians permitted the liberation (April 1919) of Zaghlul and his associates. They left Malta for Paris, where their hopes of obtaining a hearing at the peace conference were disappointed.

Milner Mission.—The British government now decided to send a mission of inquiry to Egypt, under the chairmanship of Lord Milner. Its wide terms of reference showed some understanding of the reality of Egyptian feeling, but its departure was delayed. Lord Milner and his colleagues did not reach Cairo until Dec. 1919 and by that time effective arrangements for boycotting it had been leisurely elaborated. The sheikhs and students of Al Azhar university identified themselves with the nationalists, there were repeated assaults on British soldiers and three attempts to assassinate Egyptian ministers. The mission was undeterred, however, and, although the boycott was never abandoned, an increasing number of Egyptians took part in informal discussions. The mission found no difficulty in estimating the feeling in the country. The working of every department was thoroughly investigated and before the mission returned to England, in March 1920, a large volume of material had been assembled and certain proposals, unanimously adopted, had been provisionally drafted.

It was now apparent that the Egyptian delegation, then established in Paris, represented a majority of Egyptian opinion. Zaghlul was invited to London, and in August a memorandum, which came to be known as the Milner-Zaghlul agreement, was drafted. Its terms coincided closely with the proposals of the Milner mission. There was to be a treaty of alliance in which Egypt contracted definite obligations in return for the recognition

of its independence. Reference to the Sudan, however, was specifically excluded. Four members of the Zaghlul delegation proceeded to Cairo to consult opinion in Egypt and on their return to London, in Oct. 1920, reported that the proposed settlement had been well received. The Egyptian delegates then left England and the mission reported on Dec. 29, 1920. The preconditions for a satisfactory settlement seemed to have been arrived at but the British government delayed acting upon the mission's recommendations.

In London there was influential resistance to the mission's proposals and it was believed that Zaghlul was losing his hold over the Egyptian people. The British government invited the sultan of Egypt to send an official delegation to London to discuss substituting for the protectorate "a relationship which would, while securing the special interests of Great Britain and enabling her to offer adequate guarantees to foreign powers, meet the legitimate aspirations of Egypt and the Egyptian people." The sultan accepted and Adly Yeghen Pasha formed a ministry. But Zaghlul at once attacked the new government and there were outbreaks of mob violence at Tanta, in Cairo and in Alexandria, where the principal sufferers were the large Greek community. A military court of inquiry attributed responsibility to Zaghlul's party and 16 of the rioters were executed.

It was not until six months had elapsed after the presentation of the Milner report that the delegation, under the presidency of Adly Yeghen Pasha, arrived in London. One stumbling block was avoided by the omission of all reference to the Sudan in the ensuing conversations. But the negotiations broke down completely on the issue of military occupation. The British government refused to accept the Egyptian view that, while a permanent British force should be stationed in Egypt to ensure imperial communications, it ought, in peace time, to be confined to the area of the Suez canal. Adly resigned, a successor could not be found and disorder again broke out in Egypt. Once more Zaghlul with five others was deported by the British authorities. The disturbances were vigorously suppressed and by the end of 1921 a settlement again seemed possible. Lord Allenby, who still favoured conciliation, proceeded to London with his proposals and his resignation in his pocket. He was soon back in Cairo and, on Feb. 28, 1922, the day of his return, a declaration was published which stated that the British government desired to recognize Egypt as an independent sovereign state. Certain subjects, however, were reserved absolutely to the discretion of the British government: the security of British imperial communications; the defense of Egypt; the protection of foreign interests and of national minorities; and the Sudan.

10. Independence.—The sultan then assumed the title of king as King Fuad I (*q.v.*) and proclaimed Egypt a monarchy. Sarwat Pasha, while carefully protecting himself from the appearance of accepting all the terms of this declaration, set to work to form a ministry, with a program which contemplated a democratic constitution, ministerial responsibility and the revocation of martial law, which had existed since Nov. 1914. Meanwhile the British government informed the Powers (March 15, 1922) that, although Egypt was independent, Great Britain was not prepared to admit any question or discussion of the relations between the two countries, stating further that any interference in the affairs of Egypt by a third power would be regarded as an unfriendly act.

The constitution, drawn up by a representative commission, was completed by Oct. 1922. The British government took exception to some clauses referring to the Sudan and to the description of King Fuad as "king of Egypt and the Sudan." Acute differences also arose between the ministry and the king, whose views regarding the suitability of democratic institutions for Egypt seemed hardly consistent with his public utterances. Sarwat resigned in November and Tewfik Nessim Pasha became prime minister for the second time. There ensued a period of renewed tension with outrages against British subjects, which prevented the abrogation of martial law. Tewfik Pasha resigned and after an interval Yehia Ibrahim Pasha formed an administration which took office in March 1923. Parliamentary government was formally established in Egypt by the constitution, which was promulgated on April 19, 1923. There were two representative chambers: a senate, one-half of which was renewable every five years; and a chamber of deputies. Two-fifths of the senators were appointed by the king, and the remainder elected on a basis of one senator for each 180,000 inhabitants. Deputies were elected for five years by indirect universal suffrage, one deputy for each 60,000 inhabitants. In July 1923 martial law was at last terminated by a proclamation of the commander in chief after an exchange of notes between the British and the Egyptian governments. Yehia also carried through a law governing the retirement and compensation of foreign officials in the Egyptian service, and met with some success in his attempts to curb political violence.

11. Zaghlul's Ministry.—Zaghlul Pasha Saad then returned from exile to meet an enthusiastic reception, which was quickly translated into political terms. After the elections of Jan. 1924, he dominated both the chamber of deputies and the senate. Yehia resigned and Zaghlul formed a ministry. The almost simultaneous advent to power of a Labour government in Great Britain revived Egyptian hopes of further concessions. But Zaghlul, who was then completely master of the situation in Cairo, had repeatedly affirmed his claim for the incorporation of the Sudan with Egypt. The British prime minister, J. Ramsay MacDonald, declared that his government had no intention of evacuating the Sudan, evoking strong protests from Cairo. Meanwhile, nationalist propaganda had transferred its activity to the Sudan and disturbances broke out (Aug. 9–11) at Khartoum, the Atbara post and Port Sudan. A proposal from Cairo for an Egyptian-Sudanese commission of inquiry elicited the reply from the British government that the British governor general of the Sudan was alone responsible for order there.

A meeting between the British and Egyptian prime ministers eventually took place in London (Sept. 25–Oct. 3, 1924), but the points of view of the two governments were irreconcilable. Zaghlul returned to Cairo, replaced men of moderate opinions in the public service with uncompromising nationalists, and engaged in an attack on the palace. He tendered a resignation which King Fuad did not dare to accept. Events were marching steadily to a crisis.

On Nov. 19, 1924, Sir Lee Stack, the sirdar and governor general of the Sudan, was shot and killed while driving through the streets of Cairo. The assassination roused a storm of indignation in Great Britain. Rigorous measures were called for, and the demands put forward by Allenby, on Nov. 22, were tantamount to an ultimatum. The Egyptian government accepted demands for rigorous action against the perpetrators of the crime, for the prohibition of all popular political demonstrations and for the immediate payment of £500,000. But it rejected two conditions which affected the Sudan. These were the withdrawal, within 24 hours, of all Egyptian officers and military units in that country, and the insistence that Egypt should acquiesce in the intention henceforth to allow unlimited irrigation in the Gezira, where irrigation had hitherto been limited in the interests of Egyptian users of Nile water.

The British government took immediate action to implement these proposals and at the same time occupied the customs at Alexandria.

12. Ziwar.—These measures led inevitably to the resignation of Zaghlul, who was replaced by Ahmad Ziwar Pasha, the president of the senate, with a ministry of moderate views. They lost no time in negotiating a settlement with the British government, which secured the evacuation of the customs, after which Ismail Sidky Pasha joined the cabinet. The Egyptian parliament was dissolved at the end of 1924. An appeal for intervention, which it had addressed to the League of Nations and to the parliaments of other countries, remained without response.

A new Union party (Ittihad) was formed under the inspiration of the palace in the hope of drawing away from Zaghlul's party, the Wafd, some of those to whom his intransigence had become distasteful. But when the new parliament assembled, on March 23, 1925, it became evident that his influence was still supreme. Accordingly, the king again dissolved parliament, and 11 months elapsed before a new one was elected. In the interval a commission, under the chairmanship of Ismail Sidky, revised the electoral law so as to restrict the suffrage. The opposition appeared for the

moment disconcerted and disarmed; and in May a number of persons were brought to trial for complicity in the murder of Sir Lee Stack. Seven of them were executed.

Lord Allenby resigned the high commissionership in May 1925. He was succeeded in October by Lord (formerly Sir George) Lloyd, who found the autocratic power of the palace to be reviving in the absence, for health reasons, of the prime minister in Europe and in default of a parliament. Efforts to extend the royal prerogative drove many moderates into the arms of the Wafd and incurred the displeasure of Lord Lloyd. When a new parliament was elected, in May 1926, it showed an overwhelming majority for the Wafd, which was returned with about 150 candidates against about 65 from other parties.

The efforts of the police authorities and the support they had received from Ziwar had succeeded by the beginning of 1926 in bringing before the Egyptian courts the remainder of those responsible for the political crimes which had occurred in Egypt since 1920. Among them were two members of Zaghlul's government, Ali Maher, a former minister of education, and Mahmud Nokrashi, undersecretary of the interior, both subsequently to become prime ministers. To the consternation of some Egyptians and almost all foreigners, their trial ended, in May 1926, in their acquittal and involved the resignation of Judge John Felix Kershaw, the British member of the tribunal, on the ground that his two Egyptian colleagues had acted in defiance of the evidence. The British government declared that it could not accept the verdict as establishing the innocence of the persons acquitted and reserved the right to ensure the safety of foreigners in Egypt.

13. Sarwat.—Ziwar had resigned just before the new parliament met but Zaghlul declined to form a government himself, and Adly Pasha again became prime minister with a coalition cabinet. Constantly obstructed by the Wafdists, his position became impossible and in April he resigned and was succeeded by Abd al-Khalik Sarwat, who seized the opportunity to try and bring about more friendly relations following King Fuad's visit to England (July 4–26, 1927) and his friendly reception there. Informal discussions with the British foreign secretary (Sir Austen Chamberlain) resulted in a draft treaty of alliance between the two countries. Sarwat accepted the draft during a second visit to London, but before it could be forwarded officially to Cairo, Zaghlul died (Aug. 23, 1927) at the age of 74. The greatest national leader since Arabi, he had fought an unflinching battle for the independence of his country, and his death was the signal for a striking outburst of national grief. In his stead, the leadership of the Wafd was conferred on Mustafa al-Nahas Pasha and the Wafd and Liberals combined in a declaration of their determination to remain united in following the principles of the dead leader.

The draft, which reached Cairo in Nov. 1927, provided for an alliance between Great Britain and Egypt. Once more the problem of the Sudan was left for future settlement. Britain was to propose Egypt for admission to the League of Nations, and there were other concessions to Egyptian demands. But the paramountcy of the Wafd was once more apparent and Nahas made no secret of his hostility to the proposed treaty. As it did not provide for the complete evacuation of Egypt by the British army it was not, he alleged, worth a moment's consideration. On March 4, 1928, Sarwat informed Lord Lloyd that the cabinet had refused to discuss the draft and he was accordingly resigning. Nahas became prime minister and formed a Wafdist cabinet (March 16).

14. King Fuad's Coup.—With the renewed crisis in Anglo-Egyptian relations however, a crisis of an entirely different character ensued. It represented a new phase in the perpetual struggle between the palace and the Wafd. On June 24 two Cairo newspapers published photographs of documents which, if authentic, would seriously have compromised Nahas and other prominent Wafdists. Nahas denied their genuineness and the intrigue which they purported to disclose, but the king dismissed him, and Mohammed Mahmud Pasha, one of the two Liberal members of Nahas' original cabinet, accepted office. On July 19, 1928, King Fuad suspended the constitution and dissolved parliament, which did not assemble again for three years; he determined to carry on the government by royal decree.

Once more, in 1929, the advent to power of a Labour government in Great Britain led Egyptian nationalists to hope for further concessions. They reiterated their familiar demands and Lord Lloyd was recalled and replaced by Sir Percy Lorraine. Mohammed Mahmud Pasha, who enjoyed the king's confidence, visited London for negotiations and a draft treaty was drawn up on Aug. 3, 1929. Great Britain now conceded, in principle, the restriction of British forces to the canal zone and agreed to replace the high commissioner by an ambassador. But a treaty could only be ratified by an Egyptian parliament and elections, on Dec. 21, 1929, resulted in a victory for the Wafd. Further negotiations in London, in April and May 1930, were jeopardized when the Egyptians once more raised the issue of the Sudan. But in other matters, the British government was conciliatory and the conclusion of a treaty seemed likely when negotiations were interrupted by developments in Egypt. Pursuing their long-standing feud, King Fuad dismissed Nahas Pasha on June 17, 1930. The new prime minister, Ismail Sidky Pasha, dissolved parliament, instituted a censorship of the press and, in 1930, put through a number of alterations in the constitution, which considerably increased the powers of the king. The Wafd boycotted the subsequent elections and there followed a period of authoritarian government, first under Ismail Sidky and then, from Sept. 27, 1933, under Abdul-Fattah Yehya Pasha, which continued until the illness of the king, in 1934, brought a relaxation. Tewfik Nessim formed a nonparty cabinet (Nov. 1934) which eventually (Dec. 12, 1935) restored the constitution of 1923. Meanwhile the outbreak of the Italo-Ethiopian war and the subsequent deterioration of Anglo-Italian relations sharpened British apprehensions about the security of the middle east, and negotiations for a settlement of Anglo-Egyptian differences were once more begun. Tewfik Nessim resigned, in Jan. 1936, making way for a caretaker government, presided over by Ali Maher Pasha, and elections in May 1936 resulted in another triumph for the Wafd and the return of Nahas Pasha to power.

Meanwhile King Fuad died (April 28, 1936) and was succeeded by his son Faruk I, then 17 years old.

15. Treaty of 1936.—Nahas Pasha led a delegation to London and an Anglo-Egyptian treaty was signed on Aug. 26, 1936. This provided for close co-operation in time of war and for a gradual modification of the British occupation, in step with an increase in Egyptian powers of self-defense, which would lead eventually to the British forces being limited in number and confined to the Suez canal zone. The treaty was to last for 20 years and made no significant provision for the problem of the Sudan. The two countries then exchanged ambassadors.

In May 1937 a special session of the League of Nations, called at the request of the United Kingdom, unanimously elected Egypt to membership. In the same month a conference at Montreux, Switz., attended by representatives of Egypt and all the capitulatory powers, agreed upon a new convention, which was ratified by the Egyptian chamber on July 19, 1937, and came into force on Oct. 15. The system of capitulations was abolished and Egypt obtained full rights of jurisdiction and taxation over all residents. The mixed courts for non-Egyptians were thenceforth to apply Egyptian law only and they were to be abolished after 12 years.

For the next ten years the British and Egyptian governments were drawn together by a common concern for the security of the middle east. But while Egyptian governments, of whatever complexion, were influenced by this consideration, their opponents customarily took the opportunity to attack them on the ground of their subservience to Great Britain.

During the early years of the kingdom, the rivalry between the palace and the Wafd had centred around a strong personal hostility between King Fuad and Zaghlul. A similar rivalry now showed itself between their two successors. King Faruk came of age on July 29, 1937, and in October of that year he made a show of his political independence by appointing Ali Maher Pasha as his chief political adviser with the title of chief of the royal cabinet. Nahas Pasha objected strongly to this appointment but the popularity of the Wafd was declining, and there was an attempt on the life of the prime minister. On Dec. 30, 1937, the king dismissed Nahas Pasha and appointed Mohammed Mahmud Pasha as prime minis-

ter, with a cabinet in which there were no Wafdists. Well organized and with great resources of emotional appeal, the Wafd was still exceedingly powerful in the country, but its ability to retain the support of the mass of the people was offset by the discontent of many of its most prominent members. In Aug. 1937 Nokrashi Pasha, who had generally been regarded as Nahas Pasha's chief lieutenant, had quarreled with his leader and had gone formally into opposition. In 1939 other prominent Wafdists broke away from Nahas and formed a party, named the Saadists, which cooperated with the government. Mohammed Mahmud Pasha remained in power until Aug. 1939, in the teeth of an increasingly fierce opposition, of which the effectiveness was limited by dissension among the leaders of the Wafd. On Aug. 18, 1939, Ali Maher Pasha formed a cabinet with the support of the Saadists.

16. World War II.—Within a few years Wafd opposition once more took a markedly anti-British form. But there was a general fear of Italian ambitions regarding Egypt and the new government was, in Sept. 1939, able to implement the Anglo-Egyptian treaty by breaking off relations with Germany and by declaring a state of siege, which amounted to the imposition of martial law. The intervention of Italy, on June 11, 1940, brought the war to the soil of Egypt. British forces in the country were at once engaged with the Italian armies in Libya, while those in the Sudan clashed with the Italian garrisons in Ethiopia. Italian aircraft dropped bombs on Egypt and Egyptian territory was invaded. There was conflict within the cabinet and, on Sept. 22, 1940, four Saadist ministers resigned in protest against the refusal of their colleagues to declare war on Germany and Italy. But the British government did not press for this step since Egyptian nonbelligerency enabled the Allies to profit from the acceptance by their enemies of the status of Cairo as an open city. Also, there was no immediate prospect of being able to adequately equip the Egyptian armed forces. Ali Maher Pasha fell and, after the Wafdists had refused to co-operate in the formation of a government of national union, Hussein Sirry Pasha formed a government which pursued the policy of nonbelligerency. In order to meet the economic crisis resulting from the loss of the European market for Egyptian cotton, Great Britain purchased the entire year's crop in 1940.

The latent political crisis came to a head early in 1942. The relations between the king and the government were impaired by a dispute over a minor constitutional issue. The Wafd had been gaining steadily in popular favour from some time and commanded the support of the great majority of the people. The British pressed for establishing a more representative government, but the Wafd refused to enter a coalition and the king was unwilling to call upon Nahas Pasha to form a ministry.

At this point the British ambassador intervened with a show of military force. He confronted the king with the alternatives of surrender and abdication. Faruk agreed to summon Nahas Pasha, who became prime minister with a cabinet composed entirely of Wafdists, in Feb. 1942. Because of the Wafd's long history of opposition to British influence in Egypt, the significance of this change was not at first fully understood, and the German radio continued for more than a month to treat the change as a serious setback for Great Britain. But Nahas Pasha had declared his sympathy for the Allies. He regarded the treaty of 1936 as his own work and the Wafd had always claimed that it stood for a policy both of independence and alliance, so far as Great Britain was concerned. Before long his resolution was put to a severe test.

In the summer of 1942 the Italo-German forces came near to the conquest of Egypt. They advanced to within 40 mi. of Alexandria. For several weeks the fate of Egypt was in the balance. Nahas Pasha and his government maintained a steady and determined policy of co-operation with the British, and Egyptian public opinion remained calm. This attitude was justified by the British victory of El Alamein (*see* EL ALAMEIN, BATTLE OF) and the eventual expulsion of the Italo-German forces from northeastern Africa. Thus the Wafd made a real contribution to Allied victory but did irreparable damage to its own prestige in Egypt.

After Nahas Pasha became prime minister parliament was dissolved. New elections, in March 1942, were boycotted by the anti-Wafd parties and 231 Wafdists were returned to a house of 264 members. But the party's characteristic failure to hold its leaders together soon showed itself. As early as May 1942 Makram Ebeid Pasha, minister of finance and the second man in the party, had resigned to form a new group.

With the passing of the German invasion crisis, Egyptian interest in external politics became fixed upon the affairs of the Arabic-speaking world. The idea of a union of the Arab peoples of Asia—Egypt's eastern neighbours—was not new, but in its early days Egypt had played no part in the Pan-Arab movement. The prospects of some sort of political union between the Arab states, publicly encouraged by the British government, were already being seriously discussed when, in 1943, French intervention in the Lebanon provoked a strong outburst of pro-Arab sentiment among all parties in Egypt. King and government joined in expressions of sympathy with the Lebanese. During 1944 the growing expression of Zionist claims upon Palestine further encouraged Arab sentiment, and a conference of leaders of the Arab states, presided over by Nahas Pasha, met at Alexandria in September. When it was dissolved, on Oct. 7, it had laid the foundations of the Arab league (*q.v.*). Next day Egyptian internal politics reasserted themselves, with the abrupt dismissal of Nahas Pasha by the king.

The new government, led by Ahmed Maher Pasha, a Saadist and once a close associate of Nahas Pasha, included Makram Ebeid Pasha and the leaders of almost all groups except the Wafd. Elections, in Jan. 1945, were this time boycotted by the Wafd and resulted in a large majority for government candidates. Meanwhile the Allied leaders in conference at Yalta, U.S.S.R., had decided that attendance at the peace conference would be confined to representatives of those governments which had actually declared war upon their enemies. Accordingly the Egyptian government declared war on Germany and Japan (Feb. 24, 1945). On the same day, the prime minister was assassinated by a young Egyptian recently released from internment and suspected of Fascist sympathies. Ahmed Maher Pasha was succeeded by the foreign minister, Nokrashi Pasha, and the cabinet remained otherwise unchanged.

The Wafd, in opposition, reverted to the policy of demanding further concessions from Great Britain and of accusing the government of neglecting Egyptian interests. It called for the complete evacuation of British troops and for "the unity of the Nile valley," by which was meant the absorption of the Anglo-Egyptian Sudan into the kingdom of Egypt. Popular support for these claims was encouraged by circumstances. The war in the far east was expected to last for perhaps two years after the defeat of Germany and Egypt became an important base in the preparation of the projected campaign against Japan. Thus the end of the European war seemed to Egyptians to be followed by an increase rather than by the expected lessening of the number of foreign troops in the country. At the same time, British plans for the development of Sudanese self-government led to a fear that that country would be severed from Egypt.

17. Anti-British Agitation.—When Japan surrendered on Aug. 14, 1945, censorship in Egypt was abolished in accordance with the terms of the treaty of 1936. The political temperature at once rose sharply. Each party accused the others of lack of patriotism. The British intervention of 1942 was then described and discussed in the press and popular feeling against Great Britain rose swiftly. In Dec. 1945 Egypt asked for a revision of the treaty. The Wafd continued its agitation and was joined in this, although in nothing else, by the new and powerful Muslim Brotherhood. Disorders broke out and, on Feb. 15, 1946, Nokrashi Pasha resigned. He was succeeded by Ismail Sidky Pasha, who at once pressed Egyptian claims against Great Britain and kept a firm hand on the internal order of the country. On May 7, 1946, the British prime minister, Clement Attlee, announced his government's intention to withdraw British forces from Egypt, according to a timetable to be jointly agreed upon, and to negotiate a treaty of mutual assistance. A year earlier such a treaty might have been as popular in Egypt as its predecessor once had been. But the delay had exacerbated feelings and parties hostile to the government were able to find support for increasingly extreme demands. Sidky Pasha visited London and made an optimistic statement, but in

Egypt his claim that "Britain has accepted the unity of Egypt with the Sudan under the Egyptian Crown" was held to conflict with the British prime minister's assurance that "no change in the existing status and administration of the Sudan is contemplated." Thus the Sudan became once more the chief stumbling block to Anglo-Egyptian agreement. Sidky Pasha, who had been ill for some time, resigned on Dec. 8, 1946, and was succeeded by Nokrashi Pasha.

Under the pressure of the war years, industrialization had begun to be important in Egypt. For example, nearly 30,000 workers were then employed in the manufacture of textiles at Al Mahallah al Kubra, while a sharp rise in population began to cast ominous shadows over the already exiguous living standards of the fellahin. It seemed likely that grave social problems would command the attention of Egyptians when the absorbing tension with Great Britain was relaxed.

But this relaxation was not forthcoming. The theme of Anglo-Egyptian conflict ran strongly through the history of Egypt for the next ten years. British troops were evacuated to the canal area and Alexandria was no longer a British naval base, but the issue of the Sudan seemed insoluble. Egyptian feeling was very strong and equally in Great Britain there was a powerful demand that the Sudanese be allowed to decide their own future. The Arab league took sides with Egypt in this matter, and on Jan. 5, 1947, announced its support for Egyptian demands on the Sudan. But when the Egyptian government pressed the matter upon the attention of the United Nations Security council as a danger to peace, it was discussed without result. On June 19, 1948, the British-controlled government in Khartoum announced provisions for a substantial measure of Sudanese self-government. And on July 7, Egypt formally protested to Great Britain against this intention.

18. Palestinian War.—The affairs of Palestine began to dominate the Egyptian scene. Along with other Arab states, Egypt intervened when the British withdrew from their mandate for Palestine on May 14, 1948. At midnight on May 14–15, Nokrashi Pasha broadcast the announcement the Egyptian forces had entered Palestine to restore order and to put an end to "massacres perpetrated by terrorist Zionist gangs." Within a few days Egyptian troops had occupied Gaza and Beersheba and were reported within four miles of Bethlehem. Egyptian aircraft had bombed Tel Aviv-Jaffa and an attack, apparently in error, on an airfield still held by British forces, resulted in losses on both sides. Meanwhile the United Nations' mediator, Count Folke Bernadotte, visited Cairo and announced a truce. But his proposal for its renewal four weeks later was rejected by both sides and fighting began again with the Israeli forces generally on the offensive. On July 15 a single aircraft bombed Cairo and Egyptian towns were subjected to a partial blackout. In Oct. 1948, the Egyptian line, which had hitherto held firm, was attacked by the greater part of the Israeli forces and by the end of the month Beersheba in the east and Ashdod in the west were in Israeli hands.

Egyptian opinion was ready to lay the blame for military defeat upon the corruption of its rulers and this situation, along with the victory of the Independence Front at the first Sudanese election, brought increased political support for the Muslim Brotherhood. This organization, founded about 20 years earlier for the propagation of religious and social ideals, and led by Sheikh Hasan al-Banna, had, during World War II, developed a nationalist and terrorist wing, which was tolerated by Egyptian governments perhaps as a convenient counterweight to the illegal Communist party. On Dec. 8, 1948, however, the chief of the Cairo city police, Selim Zaki Pasha, was assassinated, supposedly by a member of the Brotherhood. The prime minister, Nokrashi Pasha, thereupon ordered its dissolution and was assassinated on Dec. 28. King Faruk appointed Ibrahim Abd al-Hadi Pasha, the head of his royal cabinet, as prime minister. On Feb. 13, 1949, Sheikh Hasan al-Banna was himself murdered in the streets.

The new prime minister, an independent in politics, added several independents to the existing cabinet of Liberals and Saadists. The Egyptian military and naval forces were still in contact with the Israelis and on Jan. 5, 1949, the latter achieved several penetrations into Egyptian territory, which drew a prompt expression of concern from the British government. Two days later, Ralph J. Bunche, the acting United Nations' mediator, began negotiations which were ended by an agreement signed on Feb. 24. This provided for the evacuation by Egyptian forces of the besieged town of Faluja and of Bir 'Asluj (Halutsa) by Israeli forces, and for a demarcation between the two armies to begin about 20 mi. S. of Tel Aviv-Jaffa. A coastal strip including the town of Gaza was to be left in Egyptian and Beersheba in Israeli hands.

During the Arab-Israeli war popular attention had been diverted from Anglo-Egyptian relations and there was less public recrimination than for several years. Echoes of old conflicts could still be generally heard amid the clamour of Egyptian politics, but it was noteworthy that the British government had seemed chiefly concerned to prevent the invasion of Egyptian territory by Israeli forces. Along with other maritime powers, however, Great Britain protested when, after the armistice, Egypt retained contraband control measures in the Suez canal designed to prevent the shipping of oil to Palestine. Meanwhile, on May 7, 1948, a series of long negotiations had been brought to an end with the conclusion of a new agreement between Egypt and the Suez Canal company. By this Egypt was to receive a greater share in the profits of the company and to acquire, by 1964, five additional seats on the board of directors, bringing its total representation up to seven. More Egyptians were to be employed by the company, and large constructional works were announced which would allow the number of ships passing daily through the canal to be increased from 35 to 60. The agreement was to hold good until the expiry of the original concession in Nov. 1968. The concession, the Egyptian government stated, would not then be renewed.

In June 1948 Egypt celebrated the closing of the mixed courts, in accordance with the terms of the Montreux convention. Foreign consular courts were closed down on Oct. 14.

19. Wafdist Government, 1950.—On Jan. 3, 1950, for the first time in ten years, an Egyptian general election was contested by all parties. The results showed a big swing toward the Wafd. Nahas Pasha became prime minister on Jan. 12. Press censorship and martial law were ended and the "national aspirations" were voiced by the king in his speech from the throne. The same speech gave some indication of the growing urgency of social problems. There were to be measures of social insurance, and primary, secondary and technical education was to be made free. For a short while there were indications of an unusual alliance of the Wafd and the palace, but the past continued to throw long shadows. Once more there was conflict among the leaders of the Wafd and a manifesto, signed by leaders of other political groups, accused the government of attempting to set up a virtual dictatorship.

Memories of the treaty of 1936 and of wartime co-operation had led to the hope that the new government might begin a period of better relations with the United Kingdom. But as the interior situation deteriorated the chances of an agreed settlement of Anglo-Egyptian difficulties appeared to grow more remote. Great Britain, the United States, France and Turkey presented a plan for a combined middle-eastern command, which Egypt was to be invited to join, early in Oct. 1951. Nahas Pasha found himself in growing political difficulty and was perhaps not unwilling to profit from the distractions to be afforded by a renewal of Anglo-Egyptian conflict. He anticipated the presentation of the new proposals by requesting parliament to abrogate the treaty, to end the condominium and to proclaim Faruk king of the Sudan. British rejection of this unilateral action was supported by France and by the United States, and the British forces in the Canal zone, now substantially reinforced, began to be involved in clashes which increased in number as the situation grew worse.

In spite of the British government's protective gestures during the Palestinian war, agreement with Britain had grown more difficult to achieve. The Sudan was being prepared for independence and the worsening international situation had made Great Britain less willing to evacuate the canal zone. Widespread guerrilla activities, often supported by the Egyptian police, were directed against the British troops and on Jan. 25, 1952, British forces sent to occupy the police barracks in Al Isma'iliyah (Ismailia) were met with strong and, to the Egyptian defenders, costly resistance.

The consequences of this episode were to alter the entire face of Egyptian political life. The next day, serious riots broke out in Cairo. While the police, embittered by the events of the day before, stood aside, about 400 buildings associated with foreigners were destroyed, thousands of families made homeless, and a number of British residents murdered. It was not until the evening that the king sent the army to restore order and dismissed Nahas Pasha. But the disaster, which discredited his old enemy Nahas, brought no advantage to the king. No less than that of the Wafd, his authority was now destroyed. For six months two prime ministers, Ali Maher Pasha and then Ibrahim al-Hilaly Pasha (a dissident Wafdist), struggled with a difficult situation in what amounted to a political vacuum, which was not filled until the night of July 22–23, 1952, when a group of officers, led by Lieut. Col. Gamal Abd-al-Nasser, seized control of Cairo and soon afterward of the whole country, without bloodshed. They set up a government, headed by Gen. Mohammed Naguib, an Egyptian hero of the Palestinian war. It was a turning point in the history of the entire middle east.

20. Mohammed Naguib.—The new regime made a clean sweep of the symbols of the past. King Faruk was obliged to abdicate and left the country unharmed. He was succeeded by his infant son as King Ahmed Fuad II, with a council of regency. But on Feb. 10, 1953, Naguib assumed "supreme powers," and on June 18 Egypt was declared a republic. The titles of pasha and bey were abolished. Meanwhile the constitution had been abrogated on Dec. 10, 1952, and on Jan. 16, 1953, all political parties were dissolved.

Beneath the surface there was ample evidence that the regime had to contend with many serious social and economic difficulties. The demographic problem grew ever more threatening as the population increased at a pace perhaps unequaled anywhere in the world. In the long run these would be the principal concerns of any Egyptian government. But the leaders of the new regime had inherited and personally shared the strong Egyptian feeling for both the reality and all the appearances of complete independence. The two great difficulties of Anglo-Egyptian relations—the future of the Suez canal zone and of the Sudan—remained unsolved. Thus, while the new Egyptian government pursued, and with considerable success, wide policies of social reform and economic development, which might ultimately have a profound effect upon the country's future, domestic affairs attracted much less attention, both within and without Egypt, than did the new republic's always important and sometimes dramatic role in international affairs.

New Anglo-Egyptian negotiations were begun in May 1954 and there seemed no doubt that both governments were anxious to reach a satisfactory and final settlement. But in Egypt national feeling was strong and might again become dangerous, while the British Conservative government was obliged to give weight to the anxieties, vociferously expressed, of a small group of its supporters, who undoubtedly had the silent sympathy of many more.

21. Gamal Abd-al-Nasser.—Early in 1954 serious difficulties were revealed among the group of military officers who ruled Egypt. On Feb. 25 a communiqué which accused him of aiming at a personal dictatorship announced the resignation of Naguib. The announcement evoked a sharp popular reaction which indicated the extent of Naguib's popularity and he was restored to the presidency on Feb. 27, but on April 18 Nasser became prime minister and during that summer Naguib's opponents gained strength. On Nov. 14 he was again dismissed on the ground that his "jealousy" of Nasser had led him to become a "tool" of the Communists and of the Muslim Brotherhood. He was placed under house arrest and from that time his influence, which within the inner circle of the ruling group had never been comparable to that of Nasser, was replaced on a nationwide scale by the authority of the real architect of the revolution. Meanwhile, during the unseen conflict between these two, the Muslim Brotherhood, the only substantial force in Egypt outside the ruling group, had been crushed or driven underground. All these events were intimately connected with attempts to solve outstanding Anglo-Egyptian problems. Determined to clean out the corrupt regime, as they saw it, of palace and politicians which had despoiled Egypt for so long and to improve the standards of living of the mass of Egyptians, the peasants and small townsmen, from whom most of them were sprung, the revolutionary officers nevertheless regarded the evacuation of British troops from Egypt as their primary task. The circumstances were not unfavourable. Great Britain, by declaring for Sudanese independence, had avoided one familiar stumbling block without, it should be added, giving any satisfaction to Egypt. Expert opinion was increasingly skeptical as to the military value of the British garrison in Egypt and after hard bargaining a draft agreement was signed, in July 1954, which provided for a phased evacuation by the British over a period of 20 months, for the maintenance of the British base in the canal zone by civilian contractors and for its "reactivation" in certain specified circumstances of apparent danger to the security of the middle east.

The British government honoured its agreement and the evacuation was duly completed. Yet even this success was incomplete in the eyes of the purists among Egyptian nationalists. Their demand had always been for "unconditional evacuation." In this agreement there were conditions and, although practical men might think them small ones, Nasser and his colleagues soon found themselves, like every Egyptian government since 1932, under attack for subservience to the British. After it had fomented disturbances among students, the Muslim Brotherhood was outlawed in Jan. 1954, but it continued to exist and fiercely attacked the Anglo-Egyptian agreement "which sells Egypt to the west." An unsuccessful attempt to assassinate Nasser, in Alexandria on Oct. 26, 1954, led to the suppression of the organization, the trial of its leaders before a military tribunal and the execution of six of them.

The evacuation of the canal zone was carried out without incident and Anglo-Egyptian relations were more cordial than at any time since immediately after the signing of the treaty of 1936. The cordiality did not last long. Hopes that Egypt would eventually join an alliance of western and middle-eastern powers for the defense of the middle east were rudely shattered by the opposition of Nasser to the Baghdad pact (*q.v.*). Egyptian relations with the principal architects of the pact—Turkey and Iraq—were already bad. Its reception was highly unfavourable throughout the Arab world, except in Iraq, where popular opinion was not free to express itself. The pact was roundly denounced by the French government and initially was quietly deplored by the United States, but Great Britain adhered to it. In Egypt the pact was regarded as an affront which the government felt obliged to denounce and resist. Since Arab opinion generally was hostile to the pact, Egypt by its resistance to it became for a time the leader of the Arab world. Closer relations with India (distrustful of Pakistan's subsequent adherence to the Baghdad pact) and with Yugoslavia gave the Egyptians a diplomatic expression of their new-found sense of independence. This was much enhanced by Nasser's personal and successful attendance at the Bandung conference (*q.v.*) in April 1955. Egypt became increasingly identified with the resurgent orient. It was attracted like India and Burma to a middle, neutralist course and, after China agreed to buy large quantities of Egyptian cotton, was alive to the possibilities of trade with the Soviet bloc.

Early in 1955 came a worsening in Egypto-Israeli relations. In common with other Arab nations, Egypt had never formally recognized the Israeli state and kept the Gulf of Aqaba and the Suez canal closed to Israeli shipping. The Egyptians were, therefore, the cornerstone of the Arab boycott of Israel. Yet in spite of hostile references in public speeches and on the Cairo radio, the Egyptian government gave so little sign of active hostility to the Jewish state that U.S. diplomacy was, in Feb. 1955, seeking Egyptian support for a Palestine settlement. On Feb. 28, 1955, however, Israeli forces attacked the Egyptian army headquarters in Gaza and the Egyptian-Israeli border became an evident threat to peace. Anxiously interpreting this incident as the beginning of a phase of aggressive frontier policy, which they connected with the return to politics of David Ben-Gurion in Israel, the Egyptians replied by employing a guerrilla force—the *fedayeen*—for limited counteraction and pressed for arms from the western powers. Relations between Egypt and the west then deteriorated swiftly. The United States, Britain and France were all seen as calling for co-operation

when they really sought subservience. A fall in British cotton purchases from Egypt and the entry of the United States into the international cotton market were both interpreted as signs of hostility, as was the cancellation by France of an agreement to supply arms, six months after it had been signed and while it was still unfulfilled.

Meanwhile Egypt's links with the Communist countries grew apace. Between July 1955 and July 1956 China became the largest buyer of Egyptian cotton. There were industrial exhibitions in Egypt from the German Democratic Republic, Rumania, Hungary and China, followed by substantial purchases of capital goods, bridges from Hungary, cranes from Bulgaria, rolling-stock from Poland and oil from the U.S.S.R. By mid-1956 trade with the Communist countries represented 35% of Egyptian foreign commerce. A Communist cultural invasion of Cairo was even more apparent. Cheap, well-produced books from China and the U.S.S.R., many of them in English, flooded the Cairo bookshops. There were exhibitions of films, paintings and ballet from the east. But there was no complementary development in Egyptian politics. Egyptian Communists had been subjected to persistent attack by the government, both through public trials and imprisonment without trial, since the summer of 1953. This pressure was not relaxed as official relations with the Communist powers grew closer. In spite of this, however, and of persistent internal divisions, the Egyptian Communist party maintained an effective underground existence and even shared in the general tendency to Egyptianization. Whereas in the past most of its leaders had been Armenians, Jews and Greeks of Egyptian nationality, they were now native Egyptians of Muslim origin.

Anxiety persisted about Israeli intentions and, in Sept. 1955, it was announced that Egypt had signed a contract to obtain arms from Czechoslovakia. The causes for this move, which made a dramatic impact upon world opinion, were two. The growing urgency of the search for heavy arms, arising from the steadily increasing anxiety about the intentions of Israel and the refusal of Egypt to accept arms from the United States on terms—including a military pact and the acceptance of a U.S. military mission —which had been acceptable to some other middle-eastern countries. In spite of this however, in Feb. 1956 an agreement was reached in principle with the International Bank for Reconstruction and Development whereby Egypt would receive a loan of $200,000,000, for the construction of the cherished project of a high dam above Aswan. On June 23, 1956, Nasser, the only candidate, was elected president of the Egyptian republic by 5,496,965 votes to 2,857, and at the same time the electors approved a new constitution by a comparable majority. The constitution, which was to be implemented in the future, provided for a national assembly of 350 members and made the president both chief of state and head of the government with a six-year term of office. Future presidents were to be nominated by the national assembly and confirmed by a popular referendum. *See* also *Government* below.

22. The Suez Crisis.—Nasser's campaign against Nuri as-Said, the Iraqi premier, and against the Baghdad pact, was pursued with vigour and his prestige grew steadily throughout the Arab world. In July there was an abrupt and dramatic change of scene from internal Egyptian and Arab affairs to an international crisis of the first magnitude. Nasser announced Egypt's readiness to accept help from the United Kingdom, the United States and the International bank for the execution of the Aswan dam scheme. But on July 18 the U.S. secretary of state, John Foster Dulles, abruptly announced that the U.S. government was no longer willing to give the help it had offered. The declared reason for this rebuff was its doubts regarding the economic soundness of the scheme. World opinion generally, however, more readily accepted the view that the new U.S. policy was the result principally of the growing accord between Egypt and the states of the Communist bloc and, to a smaller extent, of the pressure of the Israeli government and its supporters within the United States. This announcement was followed by similar decisions on the part of the United Kingdom and of the International Bank for Reconstruction and Development. To this rebuff from the west Nasser replied by announcing on July 26, in a speech at Alexandria, that the Egyptian government had nationalized the Suez Canal company and would employ the revenue from the canal in the construction of the Aswan dam. For reasons of historical association and of prestige perhaps, as much as anything else, the reaction of both the British and French governments to this announcement was very sharp. In August, Nasser refused to take part in a conference of the maritime powers, called to meet in London to discuss the status of the canal, and in the following month Robert G. Menzies, the prime minister of Australia, who had been appointed as their representative by the conference, left Cairo without having obtained any satisfaction. Negotiations for a settlement, however, went forward, both through diplomatic channels and at the United Nations, while the British and French governments built up their armed forces.

Other factors were contributing to make the situation even more critical. Anti-western nationalism had been making rapid progress both in Syria, where the Baath socialists entered the government in June 1956, and in Jordan, where the rejection of the Baghdad pact and the dismissal of Lieut. Gen. J. B. Glubb was followed in October by a general election which left as the strongest party the National Socialists, led by Suleiman Nabulsi, who became prime minister on Oct. 29. This marked growth of militant nationalism in neighbouring countries, along with the prospect of Soviet assistance to Egypt, was alarming to the government of Israel. These developments appeared in London as a serious threat to Britain's remaining influence in the middle east, especially as Nasser's influence seemed to be progressively outstripping that of Britain's principal remaining Arab ally, Nuri as-Said, the prime minister of Iraq. Equally disturbed was the French government, which believed Cairo to be the mainspring of the Algerian revolution. Thus, at a point when international negotiations on the future of the Suez canal presented a favourable appearance, Nasser's three principal international antagonists —Israel, France and Great Britain—suddenly joined, whether by collusion or not, in a single dramatic act of war against him. While international attention was focused upon events in Hungary, the Israeli army attacked Egypt on Oct. 29 and immediately obtained brilliant successes.

When the Israeli mobilization became known, the British government warned Israel that an attack on Jordanian territory would automatically invoke the Anglo-Jordanian treaty. It made no reference however to any attack on Egypt which should have involved British, French and U.S. action in accordance with the terms of a tripartite declaration made in 1950. When the United States referred the Israeli invasion to the UN Security council, Selwyn Lloyd, the British foreign secretary, objected to Israel being stigmatized as "the aggressor." On Oct. 30 Egypt and Israel received a joint Anglo-French ultimatum which instructed them to order an immediate cease-fire and withdraw ten miles from the canal, in default of which British and French forces "would intervene with whatever strength may be necessary to secure compliance." The ultimatum also required that Egypt allow British and French troops to move temporarily into key positions at Port Said, Ismailia and Suez in order to guarantee freedom of transit through the canal. The Israeli government, which had no substantial force of troops within ten miles of the canal, accepted the ultimatum. Nasser rejected it.

In the Security council, Great Britain used the veto for the first time against the U.S. resolution calling on all members to refrain from the use of force, and British and French aircraft began the systematic bombardment of Egyptian airfields and other targets. The Egyptian government armed civilians in the canal cities and blocked the waterway. On Nov. 5 British and French forces landed in the region of Port Said, in defiance of a resolution of the UN general assembly, which, on Nov. 2, had adopted a U.S. resolution calling upon Israel to withdraw behind the armistice line and other states to refrain from military action. Only Australia and New Zealand voted with Britain, France and Israel against this resolution, which was one of the most solid condemnations ever issued by the assembly on a major matter. On Nov. 4 the assembly passed without dissent a motion inspired by Canada,

to secure the establishment of an emergency international force. Its military objectives then achieved with the surrender or withdrawal of Egyptian forces, Israel accepted the cease-fire. On Nov. 6 the British government, under great internal and external pressure, agreed to suspend hostilities at midnight on the understanding that Israel and Egypt accepted the cease-fire and that the international force would take over the Anglo-French positions.

Egypt had suffered heavily in both casualties and material damage. The native area of Port Said was destroyed and the Suez canal rendered temporarily unusable. But the episode which came in Egypt to be known as "the Anglo-French aggression" had failed either to bring down the Egyptian government or to secure international control of the Suez canal. Within a few days units of the UN emergency force began to arrive by air and the Anglo-French withdrawal was completed on Dec. 22.

British and French subjects resident in Egypt had been remarkably free from molestation during the invasion, but their goods, and those of Anglo-French businesses, were now sequestrated and both countries lost their hitherto pre-eminent position in the economic, social and cultural life of Egypt. The British base in the canal zone was seized and the Anglo-Egyptian agreement terminated. Many other European nationals departed and for the first time in modern history the foreigner no longer enjoyed a privileged position in Egypt. The United Nations deprived the Israelis of the fruits of victory by obliging their troops to withdraw behind their original frontier and assisted Egypt, during 1957, by clearing and reopening the Suez canal at the cost of £10,000,000 to the maritime nations. The U.S.S.R. quickly replaced the many arms lost in the fighting and the only concession forced upon Egypt was that of the passage of ships to Israel through the Gulf of Aqaba. For France and Great Britain the balance sheet was a total loss. *(See* also SUEZ CANAL.)

23. The United Arab Republic.—By defying their ultimatum and leaning upon international pressure Nasser had frustrated the purposes of two major powers. This fact entirely overshadowed his initial defeat at the hands of the Israeli armed forces, and his prestige in the Arab world rose to new heights. It was this, along with a growing Arab determination not to let Soviet domination replace that of the west, which led, early in 1958, to the sudden and dramatic union of Egypt with Syria in the United Arab Republic. During 1957 the Communist party in Syria, led by Khalid Bikdash, the most prominent Communist in the Arab world, grew in influence. Pro-Russian officers obtained important positions in the Syrian army and the Baath leaders determined to act. Shukri al-Kuwatli, the Syrian president, made several visits to Cairo and in effect confronted the reluctant Egyptian government with the alternatives of accepting responsibility for Syria, or being made to appear publicly as the obstacle to a major step in Arab unity. On Feb. 1, 1958, came a joint announcement from Cairo, signed by the two presidents, declaring that Egypt and Syria had decided to merge in a united Arab republic. This united state came formally into existence after a plebiscite, on Feb. 21, had shown a nearly unanimous vote in its favour in both Egypt and Syria. In accordance with the Egyptian pattern Syrian political parties, including the Communist party, were then dissolved and Khalid Bikdash went into exile.

This union was a rebuff for Soviet diplomacy in Syria but Nasser continued to pursue a policy of economic involvement with the Soviet bloc alongside one of hostility to international Communism, and of political detachment. In March 1958 the kingdom of Yemen combined with the United Arab Republic to form the United Arab States. This federation was for purposes of a common foreign policy and defense and not for domestic affairs although there was planned co-ordination for education and culture. However, after the secession of Syria from the United Arab Republic *(see* below), Egypt formally ended the federation with Yemen in Dec. 1961. Throughout 1959 the Egyptian government was much preoccupied with the possibility of communization in Iraq, while trade with the Communist states went on apace and Egypt accepted offers of Soviet aid, both technical and financial, for the construction of the first stage of the Aswan dam.

While so much fierce light beat upon the problems of Egypt's international relations, there had been striking developments and to all appearances substantial progress in the country's domestic economy. The military leaders, and in particular Nasser, never ceased to express a continuing interest in the condition of the people, although a great part of the practical administrative work was carried out by men who had not been members of the original group of revolutionary officers. Notable among them were Abd-al-Moneim Kaissouny, at various times minister of finance and of commerce and national economy, and Sayed Marei, minister of state for agrarian reform. Some officers however turned their attention to civil administration, notably Mahmud Yunus, who successfully managed the nationalized Suez canal.

The union quickly provoked another among its Arab neighbours, the Hashemite kings of Iraq and Jordan. This new union was ill received in Cairo and was at once subjected to the hostility of the radio and other propaganda weapons which were controlled from that city. Accordingly the Baghdad uprising of 1958, which brought down the Iraqi monarchy and resulted in its replacement by a military regime, was thought, at first, by some to have been the handiwork of the Egyptian government. There can be no doubt that some of the popular enthusiasm which it evoked was attributable to the voice and example that came from Cairo, but it was quickly apparent that the timing at least of the coup had taken the Egyptians by surprise. Before long Egypt and Iraq had resumed their familiar role of rival contenders for the leadership of the Arab world. In particular, the Egyptian government was concerned at the apparent increase of Communist influence in Iraq. Opposition toward Communist activities, which had played so important a part in bringing about the union of Egypt and Syria, was intensified, and there were arrests of domestic Communists in Egypt and in Syria throughout 1959. Meanwhile the belief that the Soviet government was supporting the Communists and encouraging the military regime in its hostility toward the pro-Egyptian Arab Nationalist party in Iraq, brought a deterioration in relations between the United Arab Republic and the U.S.S.R. After the abortive rising of Iraqi Nationalists in Mosul in March 1959, Nasser accused the Communists—along with Britain—of responsibility for its failure and the Communists, alone, of assiduously reviving old plans to bring the whole Fertile Crescent under the rule of Baghdad by incorporating Syria with Iraq. Relations between the U.S.S.R. and the United Arab Republic then deteriorated. They were patched up in the summer of 1959 but they never regained their earlier cordiality. In September of the same year the Syrian Communist leader, Khalid Bikdash, emerged from obscurity to make an attack upon the United Arab Republic at the tenth anniversary celebrations of the People's Republic of China. Nasser temporarily withdrew his chargé d'affaires from Peking in protest.

Within the Arab world the familiar pattern of common aspiration and sectional rivalry remained singularly little changed by the upheavals of 1958. The Lebanese revolt of that year had brought U.S. intervention alongside the return of British troops to Jordan. But each occupation proved to be temporary and the emergence of an Egyptophile prime minister at Beirut did not alter the essential independence of the Lebanese republic. Relations with Jordan remained far from cordial, although diplomatic representatives were again exchanged in Aug. 1959. Next year there were renewed fierce denunciations of the Hashemite kingdom not only from Cairo radio but from Nasser himself. Relations with Pres. Habib Bourguiba's Tunisian government were strained but there was a *rapprochement* with Saudi Arabia. King Saud's seeming complicity in a plot against the life of Nasser at the time of the union with Syria was officially forgotten when the Arab king visited Cairo in Sept. 1959. But the bitter propaganda conflict with Gen. Abd-al-Karim Kassem's Iraqi government persisted with little respite.

In 1960 the Egyptian position grew stronger. The estrangements which had arisen from the Suez affair became less important. Financial and cultural relations between France and the United Arab Republic were restored in 1958, although the Algerian war and the presence in Cairo of a provisional government of Algerian

leaders delayed the establishment of full diplomatic relations. Thanks to the mediation of Eugene R. Black, president of the International bank, an agreement was signed in 1959, which covered the financial claims of Egypt and Britain arising from the Suez war. Normal relations were slow to follow, partly because the Egyptians still believed that such weight as Britain possessed was thrown against them in the internecine conflict of the Arab states and especially because it was supposed that Britain supported the new regime in Iraq much as it had supported the old. But there was a marked increase in Anglo-Egyptian trade and Egypt bought Comet aircraft from Britain and released from sequestration four of the affiliated companies of the Royal Dutch-Shell group. In Sept. 1959 James Swinhurn, who had been arrested for espionage in Aug. 1956 and imprisoned along with another British subject, James Zarb, was released.

On Dec. 1, 1959, Egypt re-established diplomatic relations with the United Kingdom and Nasser declared in a public speech that Egypt was "ready to forget the past and start a new era."

Cooler relations with the Communist powers did not bring an end to their economic co-operation with Egypt, but Nasser's determination to be dependent on neither of the two great world groupings was apparent in the strengthening of Egypt's economic links with the Federal Republic of Germany and the United States, as well as in the restoration of trade with France and the United Kingdom. The year 1959 was indeed notable in the Egyptian government's pursuit of economic independence, and the grant of a loan by the International bank toward the cost of extending the facilities of the Suez canal set the seal of international recognition upon the nationalization of this great waterway.

The U.S.S.R.'s undertaking to finance the first stage of the construction of the Aswan high dam held good and in Nov. 1959 an agreement between Egypt and the Sudan removed the principal remaining obstacle to the implementation of that key project in Egypt's economic policy. Early in Jan. 1960, when Nasser switched on a turbine at a British-designed power station at Aswan, the waters of the Nile were, for the first time, used to generate electric power in Egypt. Meanwhile, in Dec. 1959 the United Nations Educational, Scientific and Cultural organization had appealed for international co-operation to protect the valuable Egyptian antiquities, including the temple of Abu Simbel (Abu Sunbul), which would be submerged when the scheme was completed.

The process of bringing effective unity to the two sections of the United Arab Republic—Egypt and Syria—officially known as the Southern and Northern regions, had presented many difficulties, although not perhaps so many as had been anticipated. The enthusiasm with which Nasser had been greeted on his first visit to Damascus, immediately after the union, had surprised even the most optimistic of the Egyptians who accompanied him. It may be attributed to two principal factors. All Syrians, without significant exception, were devoted to the concept of the greater unity of the Arabic-speaking peoples. Many, too, were seriously concerned about the more immediate political and economic problems which faced their country. The prospect of solving these immediate problems by means of a dramatic step in the direction of Arab unity, and of calling in Nasser who, so soon after the Suez affair, was the popular hero of the Arab world, was exhilarating, almost intoxicating. Thus the welcome which the Syrian people gave to the union in 1958 was genuine. But the fact of the separate Syrian identity remained. In the years which followed it presented serious problems to the Egyptian government, problems which were a good deal aggravated by the persistent hostility between Egypt and Iraq and the apparent ambitions of certain elements in Iraq to fulfill the ancient dream of uniting the two great elements of the Fertile Crescent in a single state.

Bad harvests in 1957 and 1958, caused in part by drought and locusts, had cut Syria's agricultural income by 40% and left the country with a substantial trade deficit and a good deal of popular discontent. When Nasser visited the Northern region in March 1959, his reception was by earlier standards cordial rather than enthusiastic. In October of that year Nasser delegated full powers as governor of Syria to Field Marshal Abd-al-Hakim Amer, an Egyptian, who opened a vice-presidential office in Damascus for the direction of the affairs of the region. The new governor toured the country extensively, opened an office for the receipt of the complaints of citizens and took a number of steps designed to ameliorate the economic situation, notably the two strikingly dissimilar measures of reducing import duty upon luxury goods and upon capital goods required for the region's substantial development program. At the end of 1959 too, the last remaining political party within the United Arab Republic ceased to be an effective force. The Baath party, a Syrian-Arab Socialist organization, had led the movement for the union of the two countries. One of its leaders, Akram al-Hourani, had become the first vice-president of the new union. During the autumn of 1959 all Baathist ministers resigned or were dismissed, supposedly because they shared the passion, always widespread in Syria, for political activity and discussion and therefore declined to accommodate themselves to the Egyptian principle of a nonparty state.

Citizens of the United Arab Republic were, however, not entirely debarred from political activity. Nasser had declared his intention of discovering a new democratic model more suitable to the Egyptian and Arab peoples than the western models which had hitherto been followed. When the details of his scheme became known, it appeared to indicate the intention to combine the decentralization of administrative functions in some important matters with a pyramid of councils which would provide scope for discussion and for the sounding of public opinion in an effectively one-party state. On July 8, 1959, the election of about 40,000 members to local councils of the National Union, the government-controlled political organization, took place in both regions. These, it was announced, would elect, in their turn, councils for each of the provinces of the state. These provincial councils were to have important administrative functions and were also to act as electoral colleges, whose function was to choose a national council, from which the president would personally select a national assembly. It was during these elections that the decline of the Syrian Baath party first became apparent. The party won no more than 250 out of a total of 9,445 local council seats in the Northern region. In May 1960 Nasser issued a presidential decree placing all newspapers and the main publishing houses under the control of the National Union.

In the autumn of 1960, Nasser, along with many more of the world's political leaders, attended the meeting of the UN general assembly in New York. His role there emphasized the extent to which he had succeeded in his declared policy of establishing his country in a neutral role uncommitted to either of the major power blocs. Such "neutralism" had gained both in strength and in respectability during the years of his ascendancy in Cairo and the importance of his own role among the leaders of uncommitted states was emphasized when he, along with the prime ministers of India, Indonesia and Ghana, and President Tito of Yugoslavia, tabled a motion which called upon the heads of the Soviet and the U.S. governments to resume direct negotiations. This determination to remain uncommitted in the major power issue of the day showed itself equally in the economic policies of the Egyptian government. In the domestic economy of the United Arab Republic there was much control and much capitalism and the external proceedings of this deliberately mixed economy were equally and deliberately varied. Thus, while Soviet credits had been accepted in 1958 for the financing of the first stage of the high dam, western contractors had been awarded much of the work on that project and it was made clear that western suggestions for the financing of later stages would be favourably considered. Economic links with the Communist group remained important and the U.S.S.R. and Czechoslovakia, in particular, were a major source of the capital goods upon which the government's development plans depended. On the other hand, and with the deliberate intention of redressing the balance, trade with the western powers was officially encouraged. West German interests were particularly active in Egypt and substantial assistance was being received from the United States and from western financial organizations like the International bank, which, in Dec. 1959, had agreed to grant Egypt a loan of $56,500,000 at 6% for 15 years to meet half

the estimated cost of widening and deepening the Suez canal. In 1959, too, the Egyptian government took effective measures to check the growing dependence of Egyptian cotton exports on the Communist market. Throughout the years after the Suez war there was no change in the uniformly bad and officially nonexistent relations between Egypt and Israel.

Union Ended.—On Sept. 28, 1961, following a military coup, Syria declared itself independent of Egypt, thus disrupting the United Arab Republic. However, Egypt alone continued to be officially known as the United Arab Republic. Nasser made little resistance and a government was established in Syria with Mahmoun al-Kuzbari as prime minister. (H. S. D.)

Nasser blamed reactionary Arab forces for Syria's withdrawal and intensified his program of social revolution. He continued to confiscate property from the wealthy and in 1962 took steps to limit individuals to a yearly income of £5,000, a shareholding of £15,000, and ownership of 100 ac. In 1965, running unopposed, he was elected to another six-year term. (X.)

VI. POPULATION

The population of Egypt at the census of 1960 was 26,085,326, more than double that of 1897. This represents an increase of more than 20% subsequent to 1947 and shows that the startling rise in the annual rate after 1943 was maintained. Egypt has a major population problem which has deep repercussions upon its economy. The dense overcrowding in the Nile valley and delta, which composes about 4% of the entire country, is in sharp contrast to the sparsely populated areas of the region as a whole. If the total habitable area is taken at 13,835 sq.mi., the density of population was 1,883 per square mile in 1960 (1,408 in 1947). The majority of the people live by agriculture and agrarian population densities of this order are found nowhere else in the world except in the great river deltas of monsoon Asia. Except for the scattered oases in the Western desert and the Red sea littoral practically the whole population of Egypt is to be found in the valley and delta of the Nile; of these about 63% (16,500,000) are in the delta (lower Egypt) and the canal zone.

The densest populations are in upper Egypt, where none of the valley governorates has a density of less than 1,130.1 per square mile; Sawhaj, with 2,653.5 persons per square mile, is the most thickly populated entirely agricultural governorate of Egypt. The apparently lesser densities of lower Egypt, however, reflect the presence of large tracts of waste and marshland not yet cultivable; it is unlikely that the population density of cultivable land varies much between valley and delta. The marked increase in population after World War II was because of a remarkable drop in the death rate (particularly infant mortality) rather than any notable rise in the birth rate.

Egypt, 1960 Census

Governorates	Area (sq.mi.)	Population	Density (per sq.mi.)
Lower Egypt			
Ad Daqahliyah	1,337	2,014,833	1,507.0
Al Buhayrah	1,773	1,685,679	950.7
Alexandria (Al Iskandariyah)*	112	1,516,234	13,537.8
Al Gharbiyah	770	1,715,212	2,227.5
Al Minufiyah	585	1,347,953	2,304.2
Al Qalyubiyah	364	988,055	2,714.4
Ash Sharqiyah	1,815	1,819,798	1,002.6
Cairo (Al Qahirah)*	83	3,348,779	40,034.7
Damietta (Dumyat)	231	387,962	1,679.5
Ismailia (Al Isma'iliyah)*	320	284,115	887.9
Kafr ash Shaykh	1,348	973,019	721.8
Port Said (Bur Sa'id)*	153	245,318	1,603.4
Suez (As Suways)*	118	203,610	1,725.5
Upper Egypt			
Al Fayyum	692	839,163	1,212.7
Al Minya	878	1,560,311	1,777.1
Aswan	341	385,350	1,130.1
Asyut	600	1,329,588	2,216.0
Bani Suwayf	507	859,832	1,695.9
Giza (Al Jizah)	416	1,336,419	3,212.5
Qina	699	1,351,358	1,933.3
Sawhaj	595	1,578,858	2,653.5
Desert governorates			
Al Bahr al Ahmar (Red Sea)	—	25,452	—
As Sahra' al Gharbiyah (Western Desert)	—	103,453	—
As Sahra' al Janubiyah (Southern Desert)	—	33,932	—
Sina	—	49,769	—
Total	13,853†	26,085,326‡	1,883.0

*Urban governorate. †Habitable land only; total area 386,100 sq.mi. ‡Includes nomads.

High fertility coupled with the previous relatively short expectation of life has produced a young population. Although by the 1960s expectation of life had improved, the number of older folk remained relatively small and about two-thirds of the people were under 30. Most inhabitants lived in hamlets and villages but the proportion of urban dwellers was increasing; about 30% of Egyptians lived in towns of more than 25,000. The two major cities of Cairo and Alexandria (*qq.v.*) overshadowed all others and accounted for about 19% of the total population. The foreign community (*see* above, *The People*) is chiefly engaged in business and in the professions.

VII. ADMINISTRATION AND SOCIAL CONDITIONS

1. Government.—The dynasty of Mohammed Ali virtually ended with the abdication of his great-great grandson King Faruk in July 1952, although his infant son nominally succeeded him as King Ahmed Fuad II until the country was declared a republic on June 18, 1953. Gen. Mohammed Naguib thereafter held the offices of president and prime minister until deposed by the military junta which he nominally led. Col. Gamal Abd-al-Nasser, the strong man of the Revolutionary Command council, was elected president in June 1956 when after three years of military government a new constitution was proclaimed. By this, Egypt was to be an Islamic Arab country under a republican and democratic form of government with a single-chamber national assembly. This constitution was barely in force when the union with Syria was announced. All political parties were suspended and in 1957 a single political body, the National Union, was formed with Nasser as its leader. This organization, conceived as a popular base for representative government, maintains committees which start at village level, are reproduced in the governorate councils and extend to a supreme committee at Cairo under a controller who is the minister of local government.

In the United Arab Republic, which was proclaimed in Cairo as a union of Egypt and Syria by President Nasser of Egypt and Pres. Shukri al-Kuwatli of Syria on Feb. 1, 1958, Egypt became the Southern region of the republic. Three weeks later the union was confirmed, Nasser was elected as the first president by plebiscites which were virtually 100% in favour, and Cairo became the capital. The provisional constitution of 1958 vested executive power in the president who also had the power to appoint members of a central government and two regional governments, one for Egypt and one for Syria. This he did in Oct. 1958, forming a central government of 22 members (15 Egyptian and 7 Syrian). Syria withdrew from the U.A.R. in Sept. 1961.

Egypt is divided into 24 governorates or *muhafazat* (*see* Table). The *muhafiz* at the head of each governorate is appointed by presidential decree on the advice of the minister of local government; his office is a political one equivalent to vice-minister. The administrative head is the mudir, who presides over the governorate council of members elected from each district (markaz). Councilors are elected from a list, approved by the minister and the *muhafiz*, in which priority is given to members of the National Union. Under the mudir are the district officers (*ma'mur markaz*) and police commandants (*hukumdar*). Each village (*nahiya*) is administered by an elected committee whose functions include the maintenance of law and order by about 20 *ghaffirs* (irregular police).

Governorate councils are responsible for normal administrative functions and for public health, agriculture and education at all levels. Governorates control their own budgets, development programs and other local projects, including the *ezbas* (hamlets), which constitute the administrative centres of the former great estates. No buildings may be erected on cultivated lands outside the limits of a village.

The regular police are under the orders of the ministry of the interior but the mudirs and *ma'murs* have direct control of those in the governorates. The Cairo city police come directly under the ministry of the interior. Until 1950 Alexandria was the only city with a fully constituted municipality governed by an elected municipal council. In that year a municipality for Cairo was inaugurated and later one for Port Said. These three munici-

palities are managed by elected municipal councils.

2. Living Conditions.—The upheavals of the 1950s were political rather than social; there remained a small stratum of very wealthy landowners, merchants and industrialists, with most of the population living in extreme poverty. The emergence and expansion of a middle class was related to increasing urbanization, the expansion of industry and the ousting of foreigners. The fellahin of the countryside continued a way of life that had varied little over the centuries. Their arduous, small-scale agriculture, relying upon traditional hand implements, suggests gardening rather than farming. Many of the fellahin owned plots of land too small for subsistence, thus they must also work for larger landowners (*see* below, *The Economy: Land Ownership*). Their remuneration was very small, a reflection of the pressure of population whereby many men are in eager competition for work. A minimum rural daily wage of 18 piastres (about $0.50 or 4*s.* British) was enacted but was not enforced because apart from certain peak seasons it was unusual for fellahin to find employment for more than a few days each week. The townsman was better paid but had to meet a higher cost of living; a minimum daily wage of 20–25 piastres was laid down with a working week of 5½ days. Housing conditions by western standards were very poor in both villages and working-class urban quarters. The traditional gray mud-brick, one- or two-room houses of the fellahin, with flat roofs of date-palm leaves or durra stalks, tightly huddled into hamlet or village, had their counterparts in the slums of the growing industrial towns. By the 1960s, however, attempts were made to improve housing by the erection of workers' flats in the suburbs of Cairo and Alexandria.

3. Public Health and Social Welfare.—Despite progress made in the 20th century in public health, especially after World War II, there was much still to be done, particularly among the rural population. By the early 1960s there were more than 260 government hospitals under the ministry of public health (formed from the health department in 1936) in addition to several hundred private hospitals. Much of this effort, however, went to the towns; the services in rural areas lagged. The increasing establishment of Rural Combined centres (numbering more than 250 in the early 1960s) was an attempt to make up for this deficiency. Each catering for 15,000–20,000 of the rural population, these centres co-ordinate medical, educational, social and agricultural services through village councils. Great progress was made in stamping out diseases that were formerly endemic; cholera, plague, smallpox and malaria were successfully controlled. However, ophthalmia (Egyptian eye disease), bilharziasis and ancylostomiasis (hookworm) were virtually occupational diseases of the fellahin and affected large numbers in the country districts.

Cairo, Alexandria and Port Said have water supplies purified to western standards and pure water supplies are available to towns of 10,000 persons and over. Many smaller centres have boreholes or receive water from large filter plants set up to supply groups of villages. Despite these measures a very high proportion of the rural population still depends upon untreated supplies. The ministry has health laboratories with sections for chemistry, bacteriology and pathology (dealing also with the preparation of vaccines and serums). There is a Pasteur institute in Cairo with a centre at Luxor for preparation of vaccine to combat the large incidence of rabies in upper Egypt.

In 1933 legislation was passed prohibiting the employment of children under 12 years of age except in agriculture and the textile industry, limiting the working hours of women and children and forbidding their employment in certain dangerous industries. In 1936 employers became liable for accidents to their employees and in 1942 insurance against such accidents became compulsory. A ministry of social affairs was established in 1939, responsible for labour matters, the prisons and co-operative societies. After 1952 the ministry became the co-ordinating authority for the work of the ministries of health, education, irrigation and agriculture with co-operatives, social assistance, etc. The local representative of the ministry is the head of a provincial social unit. A trade-union congress was formed in 1955. By the early 1960s there were more than 1,300 trade unions in Egypt with an estimated membership of 440,000, and about 2,000 agricultural co-operatives.

4. Justice.—Egypt was the first Arab country to abolish (1956) the *mahkama shari'a* courts (courts of Islamic law). The cases previously handled by such courts are now heard in the civil courts by former *shari'i* judges. The religious courts maintained by non-Muslim minorities were also abolished. The civil and penal codes are based upon French law and procedure also follows the French model. There are more than 160 summary tribunals over which a single judge presides; 18 central tribunals (courts of first instance) each consisting of three judges who hear appeals from the lower courts and deal with the larger civil claims; courts of appeal at Cairo, Alexandria, Tanta, Al Mansurah and Asyut; and a court of cassation above the courts of appeal, established in 1931 and consisting of 31 judges divided into four chambers for the various kinds of cases. There is a central court in Cairo and local courts in each province and district. The independence of the Egyptian judicature was guaranteed in 1952 by removing from the minister of justice the power of appointments, promotions and transfers of judges and granting those powers to a professional judicial council. The favoured treatment of foreigners arising from the capitulations of the Ottoman empire and the mixed courts which adjudicated where the parties were of different nationalities were abolished; the capitulations in 1937 and the mixed courts in 1949.

5. Education.—After the 1952 revolution Egypt's educational system was radically reorganized. There are four stages of schools: primary, post-primary or preparatory, secondary and technical. Attendance at primary school is compulsory for all children between 6 and 12 years of age, but this could not be rigidly enforced because of lack of schools and teachers. Post-primary schools with emphasis upon practical training (three-year courses in either agriculture, commercial subjects or domestic subjects) exist for those who have not qualified for preparatory school. The preparatory stage of secondary education is for 12-year-olds who have passed the requisite examination from the primary school. After successfully completing four years at preparatory school, pupils proceed to secondary schools. There, three years' study and the necessary certificate qualify for the university; pupils who fail to reach this standard can enter technical schools. These cover a three-year course and specialize in agriculture, commerce or industry. Government control over schools is rigorous up to the university level. After 1950 education, with textbooks and midday meals, was free at all levels. Teachers are civil servants; the curricula are laid down by the ministry of education and are standardized throughout the country.

There are four modern universities in Egypt as well as the old seat of Koranic learning at Al Azhar, Cairo. The oldest of the modern establishments is Cairo university at Giza, founded in 1908 and taken over by the government in 1925 as the Egyptian university; it became Fuad I university in 1940 and Cairo university in 1954. It consists of ten faculties and eight institutes and has a student body of about 30,000. Alexandria university, founded as Farouk university in 1942, has eight faculties and four affiliated institutes. Ain Shams (formerly Ibrahim Pasha) university was founded in 1950 at Heliopolis, a northeastern suburb of Cairo. Asyut university in upper Egypt opened in 1957 with two faculties (science and engineering). Al Azhar university, centred on the mosque of that name in the medieval quarter of Cairo, was founded in 970; it has been recognized for centuries as the world's chief centre of Islamic and Arabic learning and attracts students from all parts of the Muslim world. Its library of Arabic literature is unrivaled. Although it still specializes in theological studies, Islamic jurisprudence and the Arabic language, recent reorganization includes studies in social sciences. There are 14 principal institutes affiliated with Al Azhar and the total number of students exceed 30,000.

Girls constitute 35% of the enrollment in primary schools, 15% in the secondary schools and 10% in the universities and higher institutes. There is no coeducational schooling below university level. Women are now admitted to many of the professions.

The British occupation assisted the development of education along western lines, but mostly at primary and secondary levels and the budget allocations were meagre. Little attention was paid to the kuttabs, or Koranic elementary schools, the only ones available to the great majority of the people. The budgetary allocation for education increased rapidly after Egypt was declared independent in 1922. By the early 1960s education accounted for about 15% of the total budget expenditure. Until the 1940s an excessive proportion of the educational budget was devoted to higher education, and basic elementary education remained neglected; thus several thousand young men graduated each year, most of them with little prospect of finding employment, while about 77% of the total population was illiterate. A series of *ad hoc* attempts to overcome this widespread illiteracy were begun with the institution of evening classes for illiterates between the ages of 12 and 45 years. The response was such that it proved impossible to deal with the numbers involved; a modified plan (mainly for younger people) was substituted. However the degree of illiteracy remained very high, probably exceeding 50%.

6. Defense.—Before 1936, when the Anglo-Egyptian treaty was signed, the strength of the army was about 12,000 and a small air force was being built up. A British military mission followed to supervise the training and expansion of the independent Egyptian forces. In 1939 the army strength was 1,000 officers and 25,000 men and that of the air force about 80 officers and 1,000 men. In World War II the Egyptian army gave valuable service in the defense of Marsa Matruh and Siwah, in manning antiaircraft batteries and as lines of communication troops. The Palestinian war (1948–49) revealed defects in organization and equipment, and the military *coup d'état* of 1952 was to some extent a reaction against the neglect of politicians and higher command. The subsequent reorganizations were incomplete when, on Oct. 29, 1956, Israeli forces invaded Sinai, routing the Egyptian army east of the Suez canal. Subsequently Britain and France occupied Port Said and practically wiped out the Egyptian air force. By the early 1960s the strength of the armed forces was about 100,000 and they had largely been re-equipped with Soviet and Czechoslovak armaments. There is compulsory national service (but with many exemptions) for men at the age of 18. The army consists of approximately seven infantry brigades and some armoured formations. There is also a national guard. The air force has about 30 squadrons of fighters and bombers and several transport squadrons. The principal naval units comprise a few destroyers, frigates and submarines, coastal force craft and landing craft.

VIII. THE ECONOMY

The limited cultivable area in Egypt and the increasing millions who struggle to live off it are potent factors in shaping the economy. Industrialization after 1945 owed much to population pressure; the import of food grains and the export of manufactured goods were further manifestations of it. Nevertheless, Egypt is still essentially an agricultural country since 70% of its employed population is engaged in agriculture which supplies over 90% of all exports, accounts for 40% of the national income and absorbs about two-thirds of the national capital. That such a large population can live on so small an area is due primarily to the inherent richness of the alluvial soils of the Nile valley and delta (now heavily supplemented with artificial fertilizers); to the sunny climate; to most elaborate and efficient irrigation systems; and to the unremitting toil of a population skilled in extracting the utmost return from the soil by gardening rather than farming methods.

1. Irrigation.—Rainfall is negligible and practically all agriculture in Egypt depends upon irrigation from the Nile. Agricultural land is therefore dependent particularly upon the elevation of the land with regard to the Nile and the availability of Nile water. The great fluctuations in the regime of the Nile throughout the year and also from year to year have necessitated control and conservation, although only during the 20th century has irrigation engineering been able to rise to the challenge. Nile water is now conserved and so distributed that the greater part of Egypt's farmland receives adequate water during the early summer when the Nile is at its lowest and the land otherwise would be parched. Irrigation in Egypt is of two types: basin and perennial.

Basin Irrigation.—The area under basin irrigation is now small: about 940,000 feddans (1 feddan = 1.038 ac.), or rather less than one-sixth of Egypt's farmland. The system derives from the pharaonic form of agriculture where seed was sown in the mud left after the annual floods had subsided. Subsequently, for greater control, the land was divided by earth banks into basins in which the silt-laden water remained for about a month before being run off. Modern basin irrigation resembles that of the past except that little of the water now flows direct from the Nile but is led to the basins by canals with masonry regulators to control admission. This is done in mid-August, when the Nile is in flood, the basins being inundated to a depth of four to six feet. After 40 to 60 days (when the river is subsiding) this water is led off, leaving a layer of fertile alluvium. Crops are then grown without further watering and are harvested in April and May, when the land remains fallow and parched until the next August floods. Only one crop a year is thus obtainable, except near the Nile or where well water is available to supply a summer crop. Basin-irrigated land, confined principally to Qina and Sawhaj governorates of upper Egypt, is gradually being converted to perennial irrigation as more Nile water becomes available and as it becomes increasingly imperative to crop farmland to the maximum.

Perennial Irrigation.—This is much more complicated, involving great storage reservoirs in both Egypt and Sudan; a system of barrages across the river to raise the level of the water to feed an intricate system of canals, each with its own regulator; and a comprehensive drainage system to carry the used water to the sea.

Perennial irrigation began in Egypt about 1820 when Mohammed Ali sought to extend the area for summer crops (and particularly cotton) by creating a canal system in the delta to distribute Nile water. However, the Nile being very low in summer, the canals had to be dug several metres below ground level, necessitating laborious lifting of the water onto the fields. Further, the canals silted rapidly, requiring much labour to maintain them. In 1834 it was decided to build a barrage across the Nile above its bifurcation at the head of the delta, thus raising the level to supply canals taking off from above the barrage. Attempts to construct the barrage had limited success until after 1890, when British engineers completed a major work of reconstruction, enabling the delta barrage to function as Mohammed Ali had envisaged and rendering perennial irrigation possible throughout the delta. The occupation of Egypt by the British after 1882, and their desire to create prosperity, inspired a vigorous irrigation program. This involved not only the construction of other barrages across the Nile but also of the vast dams at Aswan (completed 1902 and heightened 1912 and 1934) and Jabal Aulia, Sudan (1937), where part of the Nile floodwater is retained for release during early summer. Thus, effective perennial irrigation in Egypt was developed only during the 20th century. The floodwaters are allowed to pour straight through the open sluices of the great dams, cleansing them of silt. The dams are filled November–December as the flood is subsiding, and the stored water is released to maintain the level in the irrigation canals (via the series of barrages or weirs) in the February–July low-water period. Major canals take off from above each barrage and in turn supply branch and distributary canals from which the fellah receives his allotted water. This may be by a small sluice raised for a permitted period or by lifting the water onto the land. By these means about 4,900,000 feddans are under perennial irrigation and can grow crops during the summer. Construction of the Aswan high dam, sited four miles above the Aswan dam, began in Jan. 1960, with Soviet financial and technical assistance. It is designed to store 130,000,000,000 cu.m. (as against 5,500,000,000 cu.m. for the Aswan dam) and will permit the expansion of Egypt's irrigated area by about 2,000,000 feddans. Perennial irrigation has made the agricultural season virtually yearlong, permitting two or more crops to be taken from the same land each year. It obtains the

maximum value from the water resources and has achieved the reclamation of barren and saline lands in the delta. The provision of water in summer permits the widespread growth of cotton, the cash crop upon which Egypt's economy largely depends. On the debit side, the fields no longer receive an annual layer of Nile silt; instead much artificial fertilizer is required to maintain fertility, while the yearlong presence of water facilitates the spread of the debilitating intestinal worm diseases which now affect most of the fellahin. Little of the resultant increased prosperity seemed to have reached the mass of the people, partly because the pattern of land ownership confined the majority of the land (and therefore wealth) to a small select class, and partly because the fecundity of the peasantry caused the population to increase faster than the national income.

2. Methods of Agriculture.—The tremendous pressure of population on the land has served to fossilize farm practice and efficiency. Agriculture is labour-intensive and the tiny plots are worked with tools differing little from those of pharaonic times. Much irrigation water is lifted onto the fields by the age-old shadoof (a hand-operated pivoted lever with a weight at one end counterbalancing the bucket at the other). Other irrigation appliances are the sakieh, a wheel with buckets around the periphery actuated by a buffalo or other animal, and the *tambur* or Archimedean screw; both introduced by the Greeks. The plow (*mihrath*) is a thick stake tipped with metal serving as both share and handle, with a pole attached at an angle above the horizontal to which the draft animals are attached (two buffaloes, a buffalo and a donkey, or two camels). The *mihrath* makes a shallow furrow but does not turn the soil and is regarded as ideal for local conditions. For leveling and smoothing the *nattata* (a balk of timber on which the farmer may stand to give greater weight) pulled by draft animals is used. The *fass*, a hoe with a heavy triangular blade, serves also as spade and pickax; reaping is done with a sickle (*mingal*); for threshing the *norag* is generally used—a sled mounted on sharp metal discs which breaks up the ears and cuts the straw. The plenitude of labour keeps its cost down and perpetuates the use of such implements.

3. Land Ownership.—Universal land ownership has been possible in Egypt since 1881 when previous obstructions to ownership and inheritance were rescinded. Land development companies, speculators and capitalists bought large holdings when the success of perennial irrigation became evident. Moreover, the smaller farmers sought to possess their land and between 1895 and 1914 the owners of plots of five feddans and less almost doubled. But after 1900 the area available increased by only 15%, while Muslim laws of inheritance led to the fragmentation of plots; so that the number of owners trebled but the average holding shrunk by 60% and land prices rose steeply. Until 1952 a mere 6% of landowners shared between them two-thirds of Egypt's farmland in large estates, the remaining one-third being shared between about 2,500,000 peasants, roughly 2,000,000 of whom held an average of two-fifths of an acre each. A series of reforms of land ownership after 1952 reduced agricultural holdings and in July 1961 these were restricted by decree to 100 feddans per individual. The land is distributed by the Higher committee in charge of the Agrarian Reform program. Agricultural Co-operative societies, in each of which there is a representative of the Higher committee, assist landowners in various ways such as renting agricultural implements and buying fertilizer and seeds.

4. Crop Production.—Perennial irrigation makes it possible to grow up to three crops a year on the same land, but a two-year rotation is preferred to avoid soil exhaustion. Even so, land is fallowed for barely 4 months in 24. The usual rotation in the first year is: November–May, cereals (wheat or barley), berseem (Egyptian clover), beans, onions; June–July, berseem or fallow; February–November, cotton. In the north of the delta rice is grown as a summer crop and sugar cane may replace cotton in upper Egypt. The cultivation of berseem helps to rejuvenate the soil and provides fodder for animals. The agricultural year is divided into the *shitwi* or winter season (cereals, beans, onions, berseem); the *seifi* or summer season (long-staple cotton, rice, millet); and the *nili* or autumn season (mainly maize; *i.e.*, corn).

This agricultural sequence makes heavy demands upon the soil, which is poor in organic matter but rich in lime and potash. As dung is used mostly for fuel by the fellahin, artificial fertilizer is now widely used for crops. Under basin irrigation winter crops were the main products and exports were limited to wheat, barley, beans and onions. Perennial irrigation encouraged the cultivation of cotton as a summer cash crop at the expense of grain crops. Cotton exports rose from 47 tons in 1821 to 28,180 in 1861, 273,675 in 1901 and by the early 1960s exceeded 396,200. After 1945 the rapid increase of population outpaced the production of food grains and by the early 1960s the imports of wheat were increasing. The area of wheat increased at the expense of barley. The acreage under rice each year varies with the level of the Nile flood and much is exported. During World War II the shortage of artificial fertilizers restricted cotton acreage in favour of grain crops. In nearly every crop the yields are among the highest in the world. The only other export crop of note is onions, grown chiefly in Sawhaj and Al Minya governorates where the average yield is 15,000 lb. per acre. There is a dehydrated-onion industry and exports of both raw and dehydrated onions average 160,000 tons annually, the Netherlands and the Federal Republic of Germany being the chief importers.

The area under fruit trees expanded from 34,254 ac. in 1930 to 56,052 in 1939, 87,192 in 1949 and to more than 104,000 in the early 1960s. The increase (in response to a prohibitive tariff on imported fruit) was mainly in citrus fruits, bananas, grapes and figs. Date palms (estimated at about 6,000,000) occupied 7,785 ac. Viticulture occupied about 20,760 ac., and there was some wine production by Greeks in Alexandria. Sweet melons, watermelons, peaches and apricots are also grown.

5. Livestock.—The pressure on the land for human food precluded much cultivation of fodder crops. The apparently high cattle and buffalo population was kept for draft purposes rather than for milk or meat; in the early 1960s there were about 1,000,000 donkeys and an even higher number of cows, buffaloes and sheep. These animals are fed on berseem, chopped straw and maize fodder. Sheep and goats provided a secondary cash income in addition to wool and hair for weaving; poultry was kept throughout the country and on a commercial scale in Al Fayyum governorate. About 65% of milk production was obtained from buffaloes, 28% from cows and the remainder from sheep and goats. Only about 10% was consumed as milk, 60% was churned into butter and 30% into cheese.

6. Fisheries.—The marine catch, which is about one-third that from inland waters, is provided chiefly by the important sardine fisheries of the Mediterranean coast and at the mouths of the Rosetta and Damietta branches of the Nile. The delta lakes and Lake Qarun near Al Fayyum supplied most of the remainder. The Bardawil lagoon, east of the Suez canal, has a considerable fishery for gray mullet. Other species from the lakes are bolti, two kinds of mullet, eels and soles. The average annual catch usually exceeded 100,000 metric tons. The annual consumption of fish in Egypt is about 7 lb. per person (*cf.*, about 32 lb. in the United Kingdom). There is a valuable sponge-fishing industry on the Mediterranean coast.

7. Mining.—Petroleum, phosphate, manganese, natron (natural soda), salt and iron ore are the only minerals extensively worked. The production of the older oilfields at Al Ghurdaqah on the Red sea coast and Ras Gharib on the Gulf of Suez is declining, but new fields in Sinai (Sudr, 'Asal, Wadi Fayran and Balaeim) had increased average annual production to more than 3,000,000 metric tons by the early 1960s. Oil deposits were also discovered at Bakr and Karim on the Red sea coast and production was commenced. Phosphate of lime is also mined on the Red sea coast and manganese at Umm Bugma in Sinai, 15 mi. inland from the ore port of Abu Zanimah. Iron ore from just east of Aswan is sent down the Nile to a small iron and steel works at Hulwan, near Cairo. Common salt is mined on the Red sea coast and is evaporated from sea water at Port Said; much is exported. Natron from the soda lakes in the Wadi an Natrun, about 60 mi. W.N.W. of Cairo, is used largely for caustic soda manufacture. Gypsum is found in many localities. Deposits of talc, lead, wolframite,

asbestos and chrome are also mined in small quantities. Gold, copper and precious stones (emeralds and turquoises) were mined in ancient times but are no longer of importance. Building materials (limestone, alabaster, marble, granite, basalt) are widely available.

8. Industries.—The middle years of the 20th century saw marked progress in the size and scope of manufacturing industry in Egypt motivated by the need to increase the national income, to diversify the economy and to satisfy the aspirations of nationalism. Industrial development showed little vitality before the imposition in 1930 of a customs tariff. Capitalists preferred to invest in land, the home market for manufactured articles was small, the range of raw materials and fuel available was limited and there was a shortage of technicians and skilled labour. Tariffs imposed on imported manufactured goods changed the situation. The government made industrial loans available and the all-Egyptian Banque Misr, founded in 1920 to foster Egyptian industry, established a succession of companies engaged in printing, cotton ginning, transport, spinning and weaving of linen, silk and cotton, shipping, insurance, vegetable-oil extraction and the manufacture of pharmaceuticals and rayon.

World War II, with Egypt a major Allied base but cut off from European industry, gave further impetus to Egyptian industrial development. By the early 1960s the best developed industries were cotton spinning and weaving. The Egyptian motion-picture industry, unique in the Arab world, produced films mainly in Giza, a suburb of Cairo, for distribution throughout Arab countries. Egypt, with its many archaeological sites, also derived a considerable revenue from the tourist trade. Cairo and Alexandria contain two-thirds of all the factories, employing half the total industrial labour force. Most businesses in Egypt were nationalized in the years 1957–61. Egypt's industry derives its power mainly from oil, usually in the form of electricity.

Plans for economic development were launched in 1957 with industrialization programs which in 1960 were superseded by a five-year plan for industry, aimed at increasing national income and saving foreign currency. Under the plan, state investment in industry was scheduled to achieve a value increase of two-thirds in gross national production; made up by 89% increase from petroleum, mining and processing industries, 24% increase from consumer goods industries and 92% in electricity output.

Textiles.—This industry, involving the spinning, weaving, dyeing and printing of cottons, wool, silk and rayon, is the most important in Egypt. In the early 1960s there were more than 1,000,000 cotton spindles and 160,000 cotton looms. Local mills take about one-fifth of the raw cotton and annual production averaged 85,000 tons of yarn and 500,000,000 sq.m. of cotton fabrics. Fabrics produced include poplins, shirtings, furnishings and toweling; only a few of the higher grades were imported. Cotton yarn and cotton piece goods were exported, mainly to China and to the U.S.S.R. and its satellites. Woolen fabrics and blankets are also produced. Hosiery and knitwear manufacture is increasing but carpetmaking is still confined to about 1,000 small workshops. The rayon industry has usurped the position and plant of the old established silk industry, which has been relegated mostly to small hand-loom workshops. Rayon yarns, staple fibre and cellulose paper are produced at Kafr Ad Dawal, near Alexandria. Local flax and jute and sisal from India and Pakistan are manufactured into twine, rope, twill and hessian. The two major textile towns of Egypt are Al Mahallah Al Kubra (65 mi. N. of Cairo) and Kafr Ad Dawal, where workers' towns have arisen around gigantic modern mills accommodating about three-fifths of Egypt's installed spindles.

Food Industries.—These are second in importance. There are five sugar mills in upper Egypt and a refinery near Cairo; production exceeds 300,000 metric tons annually and is more than sufficient for national consumption. The canning and dehydration of fruit and vegetables and production of tomato paste is on a modest scale; annual exports averaged 4,000 tons. The milling of rice in the north delta doubled after 1945; about 1,600,000 tons of paddy are produced annually and yield about 1,100,000 tons of decorticated rice. The tobacco industry has long been established but is a small employer of labour. Until the 1950s most leaf was imported from Greece and Turkey but the growing preference for American-type cigarettes has caused tobacco from the United States and Rhodesia to predominate. The home market is supplied but there are only limited exports. There are breweries in Giza and Alexandria and cola-type drinks, made by Egyptian concessionaires, captured the soft-drink market after 1946. Alcohol is distilled from by-product molasses from sugar refining and some is exported. Vegetable-oil extraction produces about 110,000 tons of oil annually, 90% of it from cotton seed. Much of this oil is used in other industries, such as soap manufacture.

Chemicals.—Industrial chemicals (sulfuric acid, caustic soda, sodium carbonate, ammonia, glycerin and bleaching powder) are produced from indigenous raw materials; soap manufacture is well developed. Egypt is self-supporting in superphosphates but the Suez factory (using gas from the oil refinery) produces only two-thirds of the nitrate requirements. Fertilizer imports declined in the early 1960s after an Egyptian company with a productive capacity of 480,000 tons a year was established. There is a growing manufacture of pharmaceuticals, including penicillin.

Petroleum Products.—Two oil refineries at Suez, with a joint capacity of about 4,500,000 tons annually, are linked by pipeline with Cairo. There are smaller refineries at Musturud, near Cairo, and Alexandria. Local oil is of a heavy base and supplies requirements of furnace fuel and gasoline but only about 60% of diesel oil and 40% of kerosene requirements, which are made up by imports from Saudi Arabia and the U.S.S.R.

Cement.—Egypt is self-supporting in cement manufacture and produces more than 1,500,000 tons annually from two factories near Cairo and one at Alexandria. About one-third is exported.

Metals.—There are iron and steel foundries using scrap metal (output about 50,000 tons annually) and a small iron and steel works at Hulwan with an annual capacity of 300,000 tons. Together they supply three-quarters of the home market. The blast furnaces use imported coke, Aswan iron ore and local limestone. Two small copper refineries smelt imported ore and produce about 2,000 tons of refined copper a year. A variety of articles in copper and other nonferrous metals is made in small workshops.

Other Industries.—Three-quarters of the leather produced is used by the boot and shoe industry which meets home requirements and has small exports; factories at Cairo, Alexandria and Damietta produce about 5,000,000 pairs annually. There is a growing rubber industry with a wide range of manufactures including tires. Glass manufacture has improved and expanded and includes plate and window glass. Egyptian factories meet most of the demand for matches. All wood for building and furniture is imported but there are no large furniture factories. By the early 1960s there was a small but growing production of railway cars, motor vehicles, bicycles, electric cables, radios and household equipment such as refrigerators and sewing machines. Although Egyptian industry has made great progress, it enjoys tariff protection and frequently its goods are priced above world prices.

9. Trade and Finance.—At the beginning of the 19th century Egypt was an agricultural country with a self-sufficient economy. The economy changed when cotton became an export crop and foreign trade developed, regulated principally by the value of cotton exports. Until 1930 imports were mainly of manufactured and luxury goods and fuel, but in that year the uniform custom duty of 8% ad valorem was replaced by graded tariffs, designed as both protectionist and fiscal measures and generally shutting out foodstuffs and manufactured goods while facilitating imports of raw materials and semifinished products. Behind this tariff wall Egyptian industry developed slowly, but exports of manufactured goods remain small and are limited chiefly to textiles and footwear. The traditional reluctance of the Egyptian to engage in commerce was virtually overcome by the sequestration of foreign businesses after the Suez conflict in 1956 and the expulsion of French and British businessmen.

Political differences with the west during the 1950s, particularly over financing the Aswan high dam project and the seizure

of the Suez canal, served to deflect trade to the Communist countries, which by the early 1960s were supplying more than 30% of imports, the share of the U.S. and western Europe being about 48%. Principal imports (by value) were machinery, grain and flour, fuels and oils, iron and steel, chemicals, electric goods and fertilizers. Exports include raw cotton and its by-products (usually 75%–80% of total exports by value), cotton piece goods, rice, onions and other vegetables, phosphates, manganese ore, cement, mineral oils, crude petroleum and footwear. More than 40% of Egyptian exports went to the Communist countries and about 20% to western Europe and the U.S.

Banking.—Banking activities in Egypt date from the foundation of the Anglo-Egyptian bank in 1864. The Crédit Lyonnais began operations in Egypt in 1866, the Ottoman bank in 1867, the Crédit Foncier in 1880. Italian and Greek banking houses followed. The National Bank of Egypt and the Agriculture Bank of Egypt were founded with British capital in 1898 and 1902. The first purely Egyptian bank was the Banque Misr, established in 1920. The Agriculture bank was designed to help the small holders, but the "five-feddan" law of 1931, whereby holdings smaller than five feddans could not be seized for debt, caused a gradual reduction in the bank's business, and it closed in 1937. In 1931 the Banque du Crédit Agricole, to which this law did not apply, was founded with government support. Its statutes were amended in 1948 to enable it to function as a co-operative as well as an agricultural bank.

The National Bank of Egypt assumed the main functions of a central bank, being responsible for the note issue, acting as the government's banker and operating the exchange control on its behalf. The legislation for its transformation into a central bank was passed in 1951. Supervision was provided by a committee consisting of three government and three bank members and presided over by the minister of finance. The bank was fully nationalized in 1960. The Banque Misr, also nationalized in 1960, controls a large group of industrial companies in addition to conducting ordinary banking business. In 1957 all English and French banks and insurance companies were nationalized, being taken over by various Egyptian joint-stock companies, and laws were passed requiring that all shareholders, directors and managers of banks and insurance companies must be Egyptian nationals. In July 1961 nationalization of all financial institutions was effected by presidential decree.

The rise in the cost of living, inflation and other factors caused considerable increases in bank deposits and savings after 1939. A post-office savings bank was inaugurated in 1901.

Currency.—In 1885 the unit of currency was declared to be the Egyptian gold pound, silver being used only for token coinage, while the English sovereign, the napoleon and the Turkish pound were admitted at specified rates. The Egyptian pound is divided into 100 piastres and 1,000 millièmes. The National bank was authorized in 1898 to issue bank notes with 50% gold cover, though these notes were not legal tender until 1914. In 1916 the bank was authorized to use British treasury bonds and bills as backing, and the note issue rose from £E2,700,000 in Dec. 1913 to £64,000,000 at the end of 1919. Government currency notes and silver coins were issued to the value of more than £E3,000,000 during World War I. In 1938 the note issue was £E22,000,000. It rose again considerably during and after World War II, reaching £E148,000,000 in 1945, £E195,000,000 in 1954 and exceeded £E220,000,000 in the early 1960s. In 1948 and 1950 further important changes were made in the provisions governing the note issue, designed to make Egyptian currency independent of sterling. At the end of 1958 the note issue of £E217,000,000 was covered by gold to £E60,600,000; by Egyptian government treasury bills, £E139,400,000; by Egyptian government securities, £E7,800,000; and by foreign government securities, £E9,200,000.

The Egyptian pound (in Arabic, *geneh*) is approximately equivalent to £1 0s.6½*d.* ($2.87). Terms commonly used in Egypt for coins in circulation are: ½ piastre = *ta 'rifa;* 1 piastre = *kersh sagh;* 2 piastres = *nusf frank;* 20 piastres = *riyal.* The *talari* ("dollar") of 20 piastres is used in cotton quotations only.

Public Finance.—Aspects of Egypt's early financial difficulties are discussed in *History* above. From 1904 until the outbreak of World War I there was a continuous increase in the flow of foreign capital into Egypt; so much so that the situation began to cause some alarm. The finances of the government were, however on a firm basis, and revenue had increased from £E10,000,000 in 1884 to nearly £E18,000,000 in 1914.

The year 1914 was one of financial stringency; cotton prices fell by one-third and trade was at a standstill. In 1915 there was a marked recovery, and the five years following were of exceptional prosperity because of the enormous increases in the price of cotton and the heavy expenditure by the British army in Egypt. The fall of cotton prices to their prewar level in 1921 resulted in a temporary depression. The following three years were comparatively prosperous, but a second fall in cotton prices in 1926 occasioned another slump and the government attempted to keep up the price by purchases from its reserve fund. Between 1928 and 1931 a further fall to a price below that ruling before World War I caused considerable financial distress. Among the agricultural population this situation was reflected in arrears and debts due to the mortgage companies. The government in 1933 arranged for the interest to be consolidated into a new loan, of which two-thirds were taken over by the state. In spite of these vicissitudes, between 1914 and 1933 Egypt reduced its external debt by nearly £E73,000,000. Between World Wars I and II it was estimated that there was a slight apparent deficit in the balance of payments. The great reduction in external debt was made possible by the large surpluses in the external trade balance between 1915 and 1919, from which £E62,000,000 of capital was repatriated.

Until 1939 interest on the public debt amounted to 10% of government expenditure, but during World War II this was reduced to about 3%. In the years immediately preceding it there were small deficits, but both during and after the war substantial surpluses were shown, which enabled the government to increase the reserve fund; this in 1950 stood at about £E80,000,000, of which £E31,000,000 was "blocked" in long-term agricultural loans. Expenditure on the Palestinian war (1948–49) necessitated the withdrawal of approximately £E43,000,000 from the general reserve and much of the remainder was reduced by armament purchases and the Suez conflict.

By the end of World War II Egypt's sterling balances were £405,000,000 and they stood at £356,000,000 in 1947 when Egypt left the sterling area. These balances were blocked and released only gradually, mainly to prevent excessive conversion of pounds to dollars. The Sterling Releases agreement (1951) arranged for the balances to be withdrawn over 13 years, and allowed £14,000,000 to be converted immediately to dollars ($39,200,000). Subsequently, trade increased with east European countries which assisted with credits, armaments and technical help, while political and monetary difficulties with the west culminated in the withdrawal of the International bank, U.S. and British offers to help finance the Aswan high dam, the seizure of the Suez canal and the intervention of Britain and France in 1956. The U.S.S.R. granted Egypt a long-term loan of 700,000,000 rubles in 1958. Final agreement with Britain in 1959 settled financial differences resulting over the Suez conflict and released the outstanding sterling balances of about £90,000,000 which were earmarked for development projects. After 1946 money was set aside for development plans and after 1952 special development and social-service budgets were prepared and development loans raised. In Jan. 1957 the Economic Development organization was set up to supervise, coordinate and promote public enterprises in accordance with the national interest. By the early 1960s the total investments of the E.D.O. had increased fourfold.

The unified or central budget, which in the early 1960s was a balanced one, only embraced expenditures of the presidency, the ministry of foreign affairs, the national assembly and the armed services. A development budget was established in 1953 to cover drainage, irrigation, land reclamation, communications and other projects. Budget expenditures are covered by Suez canal dues, surpluses from other budgets and internal and external loans.

Under new laws instituted in July 1961 a steeply graded surtax system was enforced: the top net income allowed is £E5,000 a year. No man may be a director of more than two companies at

the same time or hold shares valued at more than £E10,000. A tax on revenue from real property varying from 10%–40% was levied after Jan. 1, 1962.

10. Transport and Communications.—The enclosed nature of habitable Egypt meant that many of its external contacts have been by sea. After the Suez canal was cut in 1869 and the Mediterranean and Red seas were linked, one of the world's most important commercial routes no longer passed through the Nile delta, Egypt, for long the meeting point of trade routes between Europe and the orient, relinquished its hold on this transit trade until the canal was nationalized in 1956.

Internal traffic is carried on the Nile, a supplementary network of canals, and road and rail systems. The peculiar shape of habitable Egypt, where road, rail, river and canal usually run parallel, makes a co-ordinated transport policy essential. This has become apparent as development plans have progressed and attention has been paid to the whole communications network.

Roads.—About 65% of Egypt's 12,600 mi. of roads are simply packed-earth roads suitable only for farm animals and carts. The modern highways date mostly from 1936 and were built mainly for strategic needs; these roads are constructed of concrete, of concrete base with asphalt top, or of limestone base with asphalt top. Unpaved roads, usually alongside canal banks, are of silt, always dusty and rendered muddy and nearly impassable by a little rain. The tardy development of a modern road system also reflects the strong competition from state-owned railways able to back economic measures with legislative and administrative action. The main roads are Cairo to Alexandria (desert road); Cairo to Alexandria via Banha, Tanta and Damanhur; Cairo to Suez; Cairo to Ismailia via At Tall Al Kabir (Tel el Kebir); Port Said to Suez following the canal on the western bank via Ismailia; Tanta to Ismailia via Az Zaqaziq (Zagazig); the Western desert road from Alexandria to As Sallum (Sollum) via Fukah and Matruh; the Sinai road from Ismailia, where the Suez canal is crossed by ferry, to Al 'Awja on the Israeli frontier with a branch to Al 'Arish, Rafah and Gaza; Cairo to Al Fayyum; and Cairo up the Nile valley to Asyut. There is no direct road connection between Alexandria and Port Said.

After 1952 road development received priority over railways and a number of improvement plans were begun, including the Cairo-Alexandria expressway completed in 1960. The number of motor vehicles in 1960 exceeded 100,000 (33,682 in 1939) and it was then estimated that half the total volume of the country's freight (26,000,000 tons) used the roads.

Railways.—The first railway in Africa was completed in 1855, connecting Alexandria and Cairo via Damanhur, Tanta and Banha. Other lines are: Port Said to Cairo; Al Qantarah to Rafah, crossing the northern edge of the Sinai desert to the Israeli frontier; Tanta to Damietta; Cairo to Suez; Suez to Ismailia, following the western bank of the Suez canal; Cairo to Ash Shallal (Shellal) (three miles above Aswan), following the left bank of the Nile to Naj' Hammadi, then crossing to the right; and from Cairo to As Sallum, via Matruh. The line from Al Qantarah to Rafah was built by the British army during World War I and purchased by Egypt in 1948, together with the World War II line between Matruh and As Sallum. The latter was abandoned, as also was a wartime line constructed from Qina to Al Qusayr to permit oil imports should the Suez area have become untenable. All these lines are of standard gauge; there are narrow-gauge light railways in the delta and Al Fayyum governorate. The main lines are double tracked. There is no rail link between Port Said and Alexandria nor with Sudan railways in the south. In the 1950s a number of locomotives were converted from coal to oil burning and a program of dieselization was begun. In 1958 the Cairo–Hulwan (15 mi.) line was electrified.

Nile Waterways.—The Nile, its branches and major navigable canals are much used for the transport of bulky goods. Most are navigable all the year but the low-water level in winter then prevents craft from being loaded to their full capacity. In all, there are about 2,000 mi. of navigable waterway (951 mi. of the Nile), the majority of the navigable canals being in the delta. The bulk of the inland fleet is estimated to be 10,000 flat-bottomed feluccas but there are also about 400 steam or motor vessels, nearly 300 tugs and nearly 800 towed barges.

Shipping and Ports.—Most passengers and goods entering and leaving Egypt do so by sea and via the port of Alexandria (*q.v.*), which ranks third among Mediterranean ports after Marseilles and Genoa and ships 90% by value of Egypt's exports. New wharves for coke and grain imports were constructed in 1959–60. Port Said (*q.v.*) developed rapidly after the completion of the Suez canal and with Port Fuad (Bur Fu'ad) on the eastern bank is the third town in Egypt; it has some entrepôt trade but most of its business is with vessels passing through the canal. Suez and Port Taufiq (Bur Tawfiq) have grown similarly at the southern terminus of the canal and, with two refineries, have a growing trade in oil. Damietta (Dumyat) and Rosetta (Rashid) at the mouths of the Nile distributaries have declined with the increasing size of ships and are left with only local trade by sailing vessels. The Egyptian shipping industry carries about 10% of the country's trade. The Egyptian shipping lines—the Misr Navigation company, the Alexandria Navigation company, the Khedivial Mail line and the Egyptian Shipping company—were amalgamated and nationalized in 1961 under the title of the Société Générale pour la Navigation Maritime.

Air Transport.—International and internal services are provided by the United Arab Airlines, with its base at Almaza airport at Heliopolis (Misr al Jadidah) on the outskirts of Cairo, and by Misrair, based on Cairo airport and owned by the government. There are airfields at Alexandria, Port Said, Luxor, Aswan, Asyut and Matruh.

Postal Service and Telecommunications.—The government administers all postal, telegraph and telephone services. Post and telegraph offices are to be found in practically every town and village and the demand for telephones far exceeds the supply. Automatic exchanges operate in Cairo, Alexandria, Port Said and five governorate capitals in the delta. The Eastern Telegraph and Marconi Telegraph companies operated all wireless communications from Egypt until they were nationalized in 1957. Radio telephone service is available to many countries. The U.A.R. Broadcasting service and the Television administration are in Cairo.

See also references under "Egypt" in the Index volume.

(A. B. M.)

BIBLIOGRAPHY.—*General Works:* H. Ziock, *Guide to Egypt* (1956); E. W. Lane, *An Account of the Manners and Customs of the Moslem Egyptians*, 2 vol. (1836); Sir A. Colvin, *The Making of Modern Egypt* (1906); E. Baring, Lord Cromer, *Modern Egypt*, 2 vol. (1908); W. S. Blackman, *The Fellahin of Upper Egypt* (1927); Baron G. A. Lloyd, *Egypt Since Cromer*, 2 vol. (1933–34); R. A. Bagnold, *Libyan Sands* (1935); C. S. Jarvis, *Desert and Delta* (1938); J. Ball, *Contributions to the Geography of Egypt* (1939); W. H. Worrell, *A Short Account of the Copts* (1945); E. B. Worthington, *Middle East Science* (1946); A. A. I. al-Gritly, "The Structure of Modern Industry in Egypt," *L'Egypte Contemporaire* (1947); H. H. Ayrout, *Fellaheen*, Eng. trans. by H. Wayment (1948); D. Warriner, *Land and Poverty in the Middle East* (1948), *Land Reform and Development in the Middle East* (1957); E. Garzouzi, *Old Ills and New Remedies in Egypt* (1958); Institut National de la Statistique, etc., *Memento Economique: L'Egypte* (1950); H.M.S.O., *Overseas Economic Survey: Egypt* (1947, 1951); H. E. Hurst, *The Nile* (1952); A. B. Mountjoy, "Egypt's Population Problem," *Transactions of the Institute of British Geographers* (1952), "The Development of Industry in Egypt," *Econ. Geogr.*, vol. xxviii (1952); C. P. 'Isawi, *Egypt at Mid-Century* (1954); J. and S. Lacouture, *Egypt in Transition* (1958); K. Wheelock, *Nasser's New Egypt* (1960). (A. B. M.)

Archaeology: H. Hartleben, *Champollion, Sein Leben und sein Werk*, 2 vol. (1906); E. A. Wallis Budge, *The Rosetta Stone* (1929); W. M. F. Petrie, *Seventy Years in Archaeology* (1931); S. R. K. Glanville, *The Growth and Nature of Egyptology* (1947); W. R. Dawson, *Who Was Who in Egyptology* (1951); Pierre Montet, *Isis* (1956); V. Gordon Childe, *Piecing Together the Past* (1956).

General Egyptological Works: Hermann Kees, "Aegypten," *Handbuch der Altertumswissenschaft*, vol. iii, pt. i (1933); J. Vandier, *Manuel d'archéologie égyptienne*, vol. i and ii (1952–55); R. Engelbach, *Introduction to Egyptian Archaeology* (1946); L. Klebs, *Die Reliefs des alten Reiches, Die Reliefs und Malereien des mittleren Reiches, Die Reliefs und Malereien des neuen Reiches*, pt. i: *Szenen aus dem Leben des Volkes* (*Abhandlungen der Heidelberger Akademie der Wissenschaften*) *philosophisch-historische Klasse*, 1915, 1922, 1934); W. C. Hayes, *The Scepter of Egypt: a Background for the Study of the Egyptian Antiquities in the Metropolitan Museum of Art*, pt. i and ii (1953–59); J. A. Wilson, *The Burden of Egypt: an Interpretation of Egyptian Cul-*

ture (1951); A. H. Gardiner, *Egypt of the Pharaohs* (1961).

History: Ancient Period: Eduard Meyer, *Geschichte des Altertums,* vol. i–iii (1925–37); J. H. Breasted, *A History of Egypt From the Earliest Times to the Persian Conquest,* 2nd ed. rev. (1950); W. M. F. Petrie, *A History of Egypt,* 6 vol. (1894–1925); H. R. H. Hall, *The Ancient History of the Near East,* 8th ed. (1932); *Cambridge Ancient History,* rev. ed. (1961 *et seq.*); E. Drioton and J. Vandier, *Les Peuples de l'Orient méditerranéen,* vol. ii, *L'Égypte,* 3rd ed. (1952); A. Scharff and A. Moortgat, *Ägypten und Vorderasien im Altertum* (1950); Sir A. H. Gardiner, *Egypt of the Pharaohs* (1961); H. Kees, *Ancient Egypt: a Cultural Topography,* in Eng. trans. (1961).

Historical texts in translation include: J. H. Breasted, *Ancient Records of Egypt,* 5 vol. (1906–07); J. B. Pritchard, *Ancient Near Eastern Texts Relating to the Old Testament,* 2nd ed. (1955); J. A. Knudtzon and O. Weber, *Die El-Amarna Tafeln,* 3 vol. (1915). Studies of special periods: W. B. Emery, *Archaic Egypt* (1962); I. E. S. Edwards. *The Pyramids of Egypt* (1961); G. A. Reisner, *A History of the Giza Necropolis* (1946–55) and *Mycerinus* (1931); H. E. Winlock, *The Rise and Fall of the Middle Kingdom in Thebes* (1947); G. Posener, *Littérature et politique dans l'Égypte de la XII^e dynastie* (1956); G. Steindorff and K. Seele, *When Egypt Ruled the East* (1942); H. von Zeissl, *Äthiopen und Assyrer in Ägypten* (1944); F. K. Kienitz, *Die politische Geschichte Ägyptens vom. 7. bis zum 4. Jahrhundert vor der Zeitwende* (1953).

Ancient Civilization and Culture: The Land and People: P. Montet, *Géographie de l'Égypte ancienne* (1957); H. Frankfort, *Birth of Civilization in the Near East* (1951); H. Gauthier, *Dictionnaire des noms géographiques* (1925–31); J. Baikie, *Egyptian Antiquities in the Nile Valley* (1932); F. Hartmann, *L'Agriculture dans l'ancienne Egypte* (1923).

Social Life: A. Erman and H. Ranke, *Aegypten und aegyptisches Leben im Altertum* (1923); H. E. Winlock, *Models of Daily Life in Ancient Egypt* (1955); P. Montet, *Les Scènes de la vie privée dans les tombeaux égyptiens de l'ancien empire* (1925) and *Everyday Life in Egypt in the Days of Ramesses the Great,* Eng. trans. by A. R. M. Hyslop and M. S. Drower (1958); W. C. Hayes, *The Scepter of Egypt,* vol. i and ii (1953–59); N. de G. Davies, "The Town House in Ancient Egypt," *Metropolitan Museum of Art Studies,* vol. i (1928–29); H. Frankfort, *The Mural Painting of el Amarneh* (1929) and (with others) *The City of Akhenaten* vol. i–iii (1923–51); W. M. F. Petrie, *Illahun, Kahun and Gurob* (1891); H. Bonnet, *Die aegyptische Tracht* (1917); B. M. Cartland "The Dress of the Ancient Egyptians," *Bulletin of the Metropolitan Museum of Art,* vol. xi (1916); A. Lucas, "Ancient Egyptian Wigs," *Annales du service des antiquités de l'Egypte,* vol. xxx (1930); H. Hickmann, "Instruments de musique" *Catalogue général des antiquités égyptiennes du musée du Caire* (1949); H. G. Farmer, "The Music of Ancient Egypt," *The New Oxford History of Music,* vol. i (1957); I. Lexova, *Ancient Egyptian Dances* (1935); E. Brunner-Traut, *Der Tanz im alten Aegypten* (1938).

Law: E. Seidl, *Einführung in die ägyptische Rechtsgeschichte* (1951); S. R. K. Glanville (ed.), *The Legacy of Egypt* (1942); P. M. Meyer, *Juristische Papyri* (1917); J. G. Winter, *Life and Letters in the Papyri* (1933); R. Taubenschlag, *The Law of Greco-Roman Egypt in the Light of the Papyri* (1955).

Warfare: W. Wolf, *Die Bewaffnung des altaegyptischen Heeres* (1926); W. M. F. Petrie, *Tools and Weapons* (1917); H. E. Winlock, *The Slain Soldiers of Neb-hepet-Re' Mentu-Hotpe* (1945); Somers Clarke, "Ancient Egyptian Frontier Fortresses," *Journal of Egyptian Archaeology,* vol. iii (1916); R. O. Faulkner, "Egyptian Military Organization," *Journal of Egyptian Archaeology,* vol. xxxix (1953).

Transport and Trade: C. Boreux, *Études de nautique égyptienne,* Mémoires publiés par les membres de l'institut français d'archéologie orientale (1925); T. Säve-Söderbergh, *The Navy of the Eighteenth Egyptian Dynasty* (1946); W. F. Edgerton, "Ancient Egyptian Ships and Shipping," *American Journal of Semitic Languages and Literatures,* vol. xxxix (1922–23); J. Hornell, *Water Transport* (1946).

Arts and Crafts: C. Aldred, *The Development of Ancient Egyptian Art,* 3 vol. (1949–51); K. Lange and M. Hirmer, *Egypt. Architecture, Sculpture and Painting in Three Thousand Years,* Eng. trans. by R. H. Boothroyd (1956); H. Ranke, *Masterpieces of Egyptian Art* (1952); W. Stevenson Smith, *Art and Architecture of Ancient Egypt* (1958); Nina M. Davies, *Ancient Egyptian Paintings,* 3 vol. (1936); A. Mekhitarian, *Egyptian Painting,* Eng. trans. by S. Gilbert (1954); W. C. Hayes, *Scepter of Egypt,* 2 vol. (1953–59); P. Fox, *Tutankhamun's Treasure* (1951); B. V. Bothmer, *Egyptian Sculpture of the Late Period* (1960).

Science, Medicine and Technical Processes: O. Neugebauer, *The Exact Sciences in Antiquity* (1952); C. Singer *et al.* (eds.), *History of Technology,* vol. i (1954); R. A. Parker, *Calendars of Ancient Egypt* (1950); H. Grapow *et al., Grundriss der Medizin der alten Aegypter,* 4 vol. (1954–59); B. Ebbell, *The Papyrus Ebers,* Eng. trans. (1937); C. D. Leake, *Old Egyptian Medical Papyri* (1952); A. Lucas, *Ancient Egyptian Materials and Industries,* 3rd ed. (1948); Somers Clarke and R. Engelbach, *Ancient Egyptian Masonry* (1930); S. R. K. Glanville (ed.), *The Legacy of Egypt* (1942); H. Carter and A. C. Mace, *Tomb of Tutankh-amen,* 3 vol. (1923–33).

Literature and Mythology: A. Erman, *Literature of the Ancient Egyptians,* Eng. trans. by A. M. Blackman (1927); J. A. Wilson, "Egyptian Texts," in J. B. Pritchard (ed.), *Ancient Near Eastern Texts Relating to the Old Testament* (1955); G. Lefebvre, *Romans et contes égyptiens de l'époque pharaonique,* Fr. trans. (1949); P. Gilbert, *La Poésie égyptienne* (1949); S. Schott, *Altägyptische Liebeslieder* (1950).

Ancient Religion: A great many books have been written on the subject of ancient Egyptian religion, each reflecting the particular attitude of the writer toward this controversial subject, and embodying widely differing interpretations. A few only are listed here, as a guide to further reading: J. Vandier, *La Religion égyptienne,* 2nd ed. (1949); J. H. Breasted, *Development of Religion and Thought in Ancient Egypt* (1912) and *The Dawn of Conscience* (1933); H. Bonnet, *Reallexikon der ägyptischen Religionsgeschichte* (1952); H. Frankfort, *Ancient Egyptian Religion: an Interpretation* (1948); H. Kees, *Totenglauben und Jenseitsvorstellungen der alten Ägypter* (1926) and *Das Priestertum im ägyptischen Staat* (1958); J. Černy, *Ancient Egyptian Religion* (1952); G. Roeder, *Volksglaube im Pharaonenreich* (1952); S. Sauneron, *Priests of Ancient Egypt* (1960). (M. S. Dr.)

Roman Period: H. Idris Bell, *Egypt From Alexander the Great to the Arab Conquest* (1948); A. C. Johnson, *Roman Egypt,* vol. ii of *Economic Survey of Ancient Rome* (1936); E. G. Turner, "Roman Oxyrhynchus," *Journal of Egyptian Archaeology,* vol. xxxviii (1952); A. C. Johnson and L. C. West, *Byzantine Egypt: Economic Studies* (1949); E. R. Hardy, *Christian Egypt: Church and People* (1952); H. C. Puech *et al., The Jung Codex* (1955). (E. G. T.)

Medieval Period: P. K. Hitti, *History of the Arabs,* 6th ed. (1956), and *History of Syria,* 2nd ed. (1957); J. Tagher, *Coptes et Musulmans* (1952); G. Wiet, *L'Égypte arabe,* and H. Deherain, *L'Égypte turque,* vol. iv and v of *Histoire de la nation égyptienne,* ed. by G. Hanotaux (1937 and 1934); M. Clerget, *Le Caire, étude de géographie urbaine at d'histoire économique,* 2 vol. (1934); A. J. Butler, *The Arab Conquest of Egypt* (1902); Z. M. Hassan, *Les Tulunides* (1933); De L. O'Leary, *A Short History of the Fatimid Khaliphate* (1923); S. Runciman, *History of the Crusades,* vol. ii, iii (1952, 1954); H. L. Gottschalk, *al-Malik al-Kamil von Egypten und seine Zeit* (1958); W. Popper, *Egypt and Syria Under the Circassian Sultans, 1382–1468 A.D.* (1957); E. Combe, *L'Égypte ottomane de la conquête par Selim (1517) à l'arrivé de Bonaparte* (1798), in *Précis de l'histoire d'Égypte,* vol. iii, published by the Institut français d'archéologie orientale du Caire (1933).

Modern Period: M. Rifaat, *The Awakening of Modern Egypt* (1947); F. Charles-Roux, *Bonaparte, gouverneur d'Egypte* (1936), Eng. trans. by E. W. Dickes (1937); E. W. Lane, *An Account of the Manners and Customs of the Modern Egyptians* (1860); Shafiq Ghurbal, *The Beginnings of the Eastern Question and the Rise of Mehemet Ali* (1928); M. Sabri, *L'empire égyptien sous Mohamed-Ali et la question d'orient, 1811–1849* (1930), and *L'époque d'Ismail,* 2 vol. (1932); H. Dodwell, *The Founder of Modern Egypt* (1931); G. Guémard, *Les réformes en Égypte ... (1760–1848)* (1936); G. Douin, *Histoire du regne du khédive Ismail,* 3 vol. (1936–41); A. Sammarco, *Histoire de l'Égypte moderne,* vol. iii: *Le règne du khédive Ismail de 1863 à 1875* (1937) and *Les Règnes d'Abbas, de Sa'id et d'Ismail* in *Précis d'histoire d'Égypte,* vol. iv (1935); D. S. Landes, *Bankers and Pashas* (1958); C. C. Adams, *Islam and Modernism in Egypt* (1933); Annual Reports on the Finances, Administration and Condition of Egypt, British Cmd. papers (1892–1914); *Recueil de tous les documents officiels du gouvernement égyptien* (title changed in 1882 to *Bulletin des lois et décrets*) (1880–); Hayyim Nahoum, *Recueil de firmans imperiaux ottomans adressés aux walis et aux khédives d'Égypte* (1933). (R. L. Hl.)

Lord Lloyd, *Egypt Since Cromer* (1933–34); M. Naguib, *Egypt's Destiny* (1955); Gamal Abd al-Nasser, *Egypt's Liberation: the Philosophy of the Revolution* (1955); Tom Little, *Egypt* (1959); Jean and Simonne Lacouture, *Egypt in Transition* (1958); Anwar al-Sadat, *Revolt on the Nile* (1957); Charles Issawi, *Egypt at Mid-Century* (1954); Taha Husain, *The Future of Culture in Egypt,* Eng. trans. by Sidney Glazer (1954); A. M. Ammar, *A Demographic Survey of an Egyptian Province* (1942); R. Fedden, *The Land of Egypt* (1939); Sayed Marei, *Agrarian Reform in Egypt* (1957); Hamed Ammar, *Growing Up in an Egyptian Village* (1954); H. E. Hurst, *The Nile* (1952); "John Marlowe," *Anglo-Egyptian Relations (1800–1953)* (1954); J. M. Ahmed, *The Intellectual Origins of Egyptian Nationalism* (1960).

Current history and statistics are summarized annually in *Britannica Book of the Year.* (H. S. D.)

EGYPTIAN ARCHITECTURE in this article refers to the architecture that developed in the region around the Nile valley from earliest times until the Greco-Roman period. (For more detailed historical information, *see* EGYPT: *History;* for information on the development of architecture in relation to the other arts of ancient Egypt, *see* EGYPT: *Archaeology.* Later Egyptian architecture is discussed in ISLAMIC ARCHITECTURE and MOSQUE. The ancient architecture of adjacent areas is treated in PRE-HELLENIC ARCHITECTURE.)

The earliest structures built in the Nile valley, many centuries before the beginning of recorded history, were rudimentary dwellings in the form of tent shelters of skin and matting, reed huts, and hutches made of wattle daubed with mud and of wood and matting. When more important buildings such as palaces and

shrines came to be constructed in predynastic times, they were built of very much the same materials. It is doubtful whether brick was ever used for buildings to any extent before the historical period. Borrowed from Mesopotamia, brick architecture was first used in Egypt in funerary monuments of the 1st dynasty (*c.* 3000 B.C.). Stone began to be employed at that time for thresholds and flooring, but stone architecture made its appearance about 300 years later, in the 3rd dynasty, in the step pyramid complex at Saqqara (Sakkara). This vast construction shows that the possibilities of stone as building material were not yet realized. The stone used in it was cut into small blocks which were laid and treated as though they were mud-bricks. Moreover, structural and ornamental elements of the earlier types of reed and wood building were imitated in stone, instead of new designs being created in which the special properties of stone could be utilized to better advantage. The stone dummy doors in the Saqqara complex were copied from wooden originals, the ribbed ceilings imitated the palm logs that had long been in use for roofing, and column forms reproduced the wooden posts of earlier buildings. The 3rd-dynasty small-block masonry was soon superseded. Although the megalithic masonry which developed from it marked a great advance in the structural use of stone as stone, yet it continued throughout the dynastic era to imitate pre-existent architectural forms originally designed for, or executed in, less durable materials. The Egyptian cavetto cornice, the torus molding, and most, if not all, column types reproduce, in stone, primitive forms of reed and wood construction.

In its more striking manifestations Egyptian architecture was an architecture of temples and tombs. The abodes that were erected for the gods and the dead were always more important and monumental, and made of far more durable materials, than the buildings of the transient dwellers upon earth.

TOMBS

The Mastaba.—The earliest type of tomb of real architectural significance was the mastaba, which made its appearance in the 1st dynasty and then became the prevailing mode of burial among the notables of the Old Kingdom. Essentially it consisted of a rectangular superstructure, flat-roofed and pierced by a shaft leading to an underground burial chamber. It was made at first of mud-brick, and subsequently of stone blocks. Set in the east face was a niche or a door-shaped stele, a "false door" through which the deceased might pass to receive the oblations laid upon an adjacent offering table. The place of worship was then outside, in front of the false door, where there was an open courtyard bounded by a little wall, or else, occasionally, a small funerary chapel. Later the place for ritual observances was built as a chamber inside the mastaba, the false door was placed in the west wall of the chamber and additional rooms were added. The inner walls bore inscriptions and paintings and reliefs showing the deceased at his favourite occupations and his workmen at their daily tasks. Walled up in one or more inaccessible cells were statues of the dead owner and, sometimes, of his wife and children. The finest mastabas were built at Giza and Saqqara, among the royal pyramids, for the notables of the time took pride in being buried near their lords.

The Pyramid.—The mastaba was the prototype of the pyramid, the style of tomb adopted by the sovereigns of the Old and Middle Kingdoms. By building a stone mastaba and then superimposing masses of masonry of diminishing size upon its top, Imhotep, King Zoser's architect, created the Step pyramid at Saqqara in the 28th century B.C. The burial chamber was beneath it, deep underground. The Blunted or Rhomboidal pyramid at Dahshur and the Medum pyramid represent further stages in the development toward the true pyramid. The earliest tomb planned from the outset and completed as a true pyramid appears to have been the North Stone pyramid at Dahshur, a vast monument 220 m. square at the base and 104 m. high, doubtfully attributed to Snefru, the founder of the 4th dynasty (*c.* 2700 B.C.). Three kings of that dynasty, Khufu, Khafre and Menkaure, brought pyramid building to its apogee. The famous pyramids of Giza, near modern Cairo, were erected by them. The oldest and largest of them is Khufu's pyramid, originally 230 m. square at the base and 146.59 m. high. The core is formed of huge limestone blocks, once covered by a casing of dressed limestone. The king's burial chamber, entirely built of granite, is inside the pyramid, 42 m. above ground level. Khufu's pyramid was the main feature of a large architectural complex, as were all the standard royal pyramids of the Old and Middle Kingdoms. The pyramid complex comprised, normally, the pyramid itself, which was the tomb proper and stood within a walled enclosure in the desert, a mortuary temple adjoining it, and a causeway connecting the mortuary temple with a pavilion or valley temple situated at the edge of the Nile flood plain. The Middle Kingdom pyramids were smaller and, in general, of much poorer construction than those of former times. The monarchs of the New Kingdom renounced the pyramid tomb and were buried in rock-cut tombs, a style of burial that had already been used by people of lesser rank. (*See* PYRAMID.)

Rock-Cut Tombs.—Private tombs hewn into the face of a desert cliff were constructed as early as the 4th dynasty. They became more common in the late Old Kingdom, when the provincial lords and notables of upper Egypt, forsaking the court, chose to be buried near their own home towns. The finest examples of rock-cut tombs are at Beni Hasan, Bersha and Aswan. Generally the cliff face was cut back to form a small platform or open court in front of the tomb, ending in a porch with columns or pillars carved in the living rock. This was followed by a wide hall, also with columns, at the far end of which a deep recess contained the statue of the deceased and also, often, that of his wife. The sepulchral chamber lay underneath and was reached by a steep shaft. The private tombs of the New Kingdom in the Theban necropolis were also rock-cut tombs. They were entirely hollowed out in the cliff and had an open courtyard in front, roughly quarried in the hillside. Frequently they had, as an additional feature, a small pyramid of unbaked brick which was placed in the exterior, generally above the entrance, more rarely free-standing at the side. The rock-cut tombs of the New Kingdom pharaohs in the Valley of the Tombs of the Kings at western Thebes are very remarkable for the immense amount of mining and stonecutting it took to build them, but are of no great architectural interest. Their general plan was quite simple. It was, broadly speaking, a succession of subterranean corridors leading to a square hall, often pillared, on the floor of which rested a massive granite sarcophagus. Religious and funerary hieroglyphic texts and pictures covered the walls of the tomb from end to end. The mortuary temples of these tombs were built not adjacent to them in the valley but behind the high cliffs that hem it in, facing the Nile.

TEMPLES

Temples for the Dead King.—From the functional point of view temples may be divided into those erected for the worship of the dead king, and those erected for the cult of the gods. To the first type belong the mortuary and valley temples included in every standard pyramid complex of the Old and Middle Kingdoms. The so-called valley temple stood at the lower end of the causeway that led up to the pyramid. It had a columnar hall and storerooms; basins and drainage channels have occasionally been found in the floors. Its exact function is uncertain. It has been suggested that the purification of the dead king's body by ritualistic washing and even the embalming ceremony took place in it. It might have served, moreover, as a landing stage during the inundation. The valley temple of Khafre's pyramid, known as the Granite temple, had a monumental T-shaped hall along the walls of which were ranged 23 seated statues of the king; the floor was of alabaster, and the roof supported by 16 monolithic pillars of red granite. The mortuary temple adjoined the pyramid and had, usually, a central open court surrounded by pillars or columns, a varying number of storerooms, five elongated chambers or shrines supposedly connected in some way with five official names of the monarch, and a chapel containing a false door and an offering table. It was in this chapel that the priests performed the daily funerary rites and presented offerings to the dead king's soul. The 5th dynasty mortuary temples of Sahure's and Neuserre's pyramids at Abusir are noteworthy; they appear to

have been edifices of great magnificence, with fine palm and papyrus columns of granite and walls embellished with excellent reliefs. The only known instance of a mortuary temple not adjacent to the pyramid but enclosing it is the 11th dynasty funerary monument of King Mentuhotep (*c.* 2050 B.C.) at Deir el Bahri, now badly ruined. A ramp led up to the terrace, in the centre of which was the pyramid resting on a podium and surrounded on all sides by a covered hall with 140 polygonal columns. Beyond the hall was an open courtyard flanked by columns, and then a hypostyle hall containing a chapel at the far end. Although no longer buried in pyramids, the New Kingdom sovereigns still had funerary temples built in the vicinity of their rock-cut tombs. Queen Hatshepsut's mortuary temple at Deir el Bahri consists of two successive terraces leading up to a columned court, behind which, in the middle, is the main sanctuary cut back into the cliff. The mortuary temples of the 19th dynasty (13th century B.C.), chiefly represented by those of Seti I, Ramses II, and Ramses III at Western Thebes, were much more like the contemporary standard temples for the gods. Apart from a few special provisions required by the service of the dead, they had the same pattern of pylons, open courts, vestibules, hypostyle halls and shrines.

Temples for the Gods.—The only Old Kingdom temple of this kind which has been preserved is the solar sanctuary built in the 5th dynasty by King Neuserre at Abu Gurab. It consisted originally of a large open forecourt surrounded by storerooms, cult chambers and an altar; in the rear, resting upon a mastabalike podium, an obelisk rose to a height of 56 m. What remains of the temples known to have been erected during the Middle Kingdom is too scanty to merit attention here. The numerous nonfunerary temples of the New Kingdom may be referred to two general basic types. One is the peripteral temple, in the form of a rectangular sanctuary or cella bounded on all sides by a single row of columns or pillars connected by low screen walls; the whole was covered by a flat cornieed roof, and the entire structure was raised upon a plain basement and approached by a flight of steps. The peripteral temple was particularly common during the 18th dynasty (*c.* 1580–1315 B.C.); afterward it appears to have fallen out of favour, only to be revived in Ptolemaic times.

The other basic type of temple which developed in the New Kingdom was intrinsically thus: A pylon or two-towered structure with a gateway in the middle formed the façade. Wooden masts carrying banners were fastened onto the flat fronts of the pylon towers, and obelisks and statues of the royal founder might stand before them. The gateway admitted to an open court with covered colonnades on either side of it. At the far end of the court a ramp led up to a portico or columnar vestibule occupying the whole width of the building. Beyond this lay the hypostyle hall, following which was the "holy of holies," the shrine of the deity to whom the temple was dedicated. Situated on the main axis of the building, the shrine was a relatively small chamber wherein the image of the divinity was kept, usually in a sacred bark. Adjacent to the shrine, which might be flanked by chapels for other deities, were small rooms for ritual practices and for the storage of the paraphernalia needed in the rites. Such were the essential parts and their arrangement in the style of temple which prevailed throughout Egypt from the close of the 18th dynasty to the Greco-Roman period. The pattern is clear and simple in small temples such as those erected in honour of Amen (Amon) and of Khonsu at Karnak by Ramses III in the 12th century B.C. Although obscured and complicated by various innovations, alterations, and additions, the basic pattern still remains fundamentally the same in larger sanctuaries like the Great Temple of Amen at Karnak, the temple of the Theban triad at Luxor, the Ptolemaic temples of Horus at Edfu and of Hathor at Denderah, and even the temple of Seti I at Abydos with its curious L-shaped plan. *See* also TEMPLE ARCHITECTURE: *Egypt*.

DOMESTIC ARCHITECTURE

The prehistoric forms of shelter, the tent, the one-room reed hut, the wattle-and-daub hutch, continued to exist as dwellings of the very poor throughout the dynastic period. More commodious dwellings were developed for the better classes at an early date. Rectangular mud-brick houses of either two rooms or a room and a court were built at the beginning of the Old Kingdom. Other early houses were rectangular shelters open on the front, with the roof supported by two posts driven into the ground; these porticoed houses came to have a second story which was reached by outside stairways. Parallel rectangular chambers under one flat roof with a common court in front were not unusual. At El-Lahun, in the 12th dynasty (*c.* 2000–1785 B.C.), well-to-do officials lived in large one-storyed houses of rather intricate plan. The houses had a central living room with a flat roof supported by four columns, and connected with it a number of smaller private rooms, including a bathroom, for the owner and the male members of his family; the women had a separate suite of rooms or harem. There was also an inner open court with a roofed colonnade on one side which could be used as a reception hall; and then storerooms, and the porter's lodge facing the entrance door. The 18th-dynasty houses of Akhetaton (modern Tel el Amarna) consisted generally of a central, columned living room surrounded by a varying number of chambers, a kitchen, stairs leading up to the flat roof, and a walled court in front of the house. The larger houses had also one or two reception halls opening into the central hall, and stood in sizable plots of ground enclosed by a wall. As far as can be judged from the few New Kingdom royal dwellings that have been found, palaces followed the basic pattern of the better-class mansions, only made more magnificent and bigger by enlarging or reduplicating the essential parts of the house: court, central living room with columns, reception hall, private chambers, women's apartments, and servants' quarters.

See also references under "Egyptian Architecture" in the Index volume.

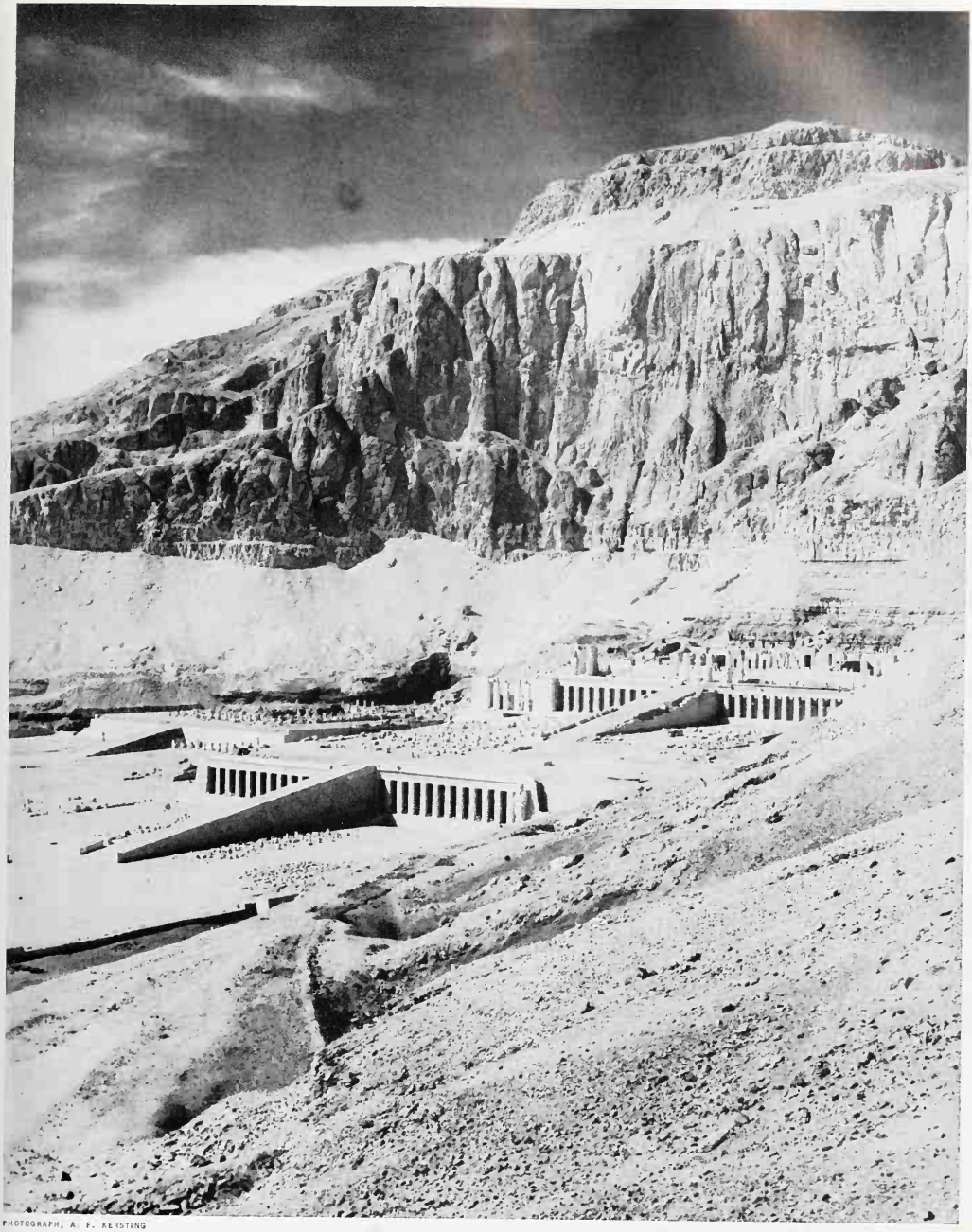
BIBLIOGRAPHY.—E. Baldwin Smith, *Egyptian Architecture as Cultural Expression* (1938); W. M. Flinders Petrie, *Egyptian Architecture* (1938); S. Clarke and R. Engelbach, *Ancient Egyptian Masonry* (1930); J. Vandier, *Manuel d'Archéologie Égyptienne* (1952–55); A. Scharff, "Ägypten," in W. Otto (ed.), *Handbuch der Archäologie*, vol i, 433 ff. (1939). (R. A. Cs.)

EGYPTIAN LANGUAGE was the speech of the Nile valley from the earliest historical times until some time after A.D. 1000. Egyptian forms a branch of the Afro-Asiatic (Hamito-Semitic) language family, the other branches being Semitic, Berber, Cushitic and Chad (which includes Hausa). While the relationship among these is certain, as is demonstrated by a comparison of the personal pronouns and verb affixes, the details are still being worked out.

Only Egyptian and Semitic have records from very ancient times. Even by 3000 B.C., however, these two branches were very distinct. Among the similarities which remained are the possession of two genders (masculine and feminine) and three numbers (singular, dual and plural). The morphologies of both were characterized by consonantal roots which combined with vowel patterns and affixes, something like English *to fall* and *to fell* or *drink*, *drunken* and *drench*. Egyptian had a suffix verb form (the old perfective) which was related to the Akkadian permansive and the West Semitic perfect. It possessed a series of other verb forms, also formed with suffixes, which are peculiar to Egyptian.

History.—There are ample documents for tracing the Egyptian language for well over 4,000 years, so that it has the longest recorded history of any language known. The oldest readable inscription of any length dates from about 2650 B.C.; the latest Coptic compositions are of about the 14th century after Christ. Five major periods are distinguished. (*See* HIEROGLYPHS.)

Old Egyptian.—A number of 1st and 2nd dynasty inscriptions from around 3000 B.C. are known. Most of these are very short, being largely names and titles, and can generally be read. The longer inscriptions of this period are only partially decipherable. By the 3rd dynasty the writing conventions had been established, and the real history of Old Egyptian begins. The main sources for knowledge of the language are the biographic texts (beginning at the end of the 3rd dynasty), the Royal decrees (4th to 6th dynasties) and the Pyramid texts (5th and 6th dynasties). The language of the Pyramid texts, which were magic spells for the well-being of the king's soul after death, shows peculiarities of its

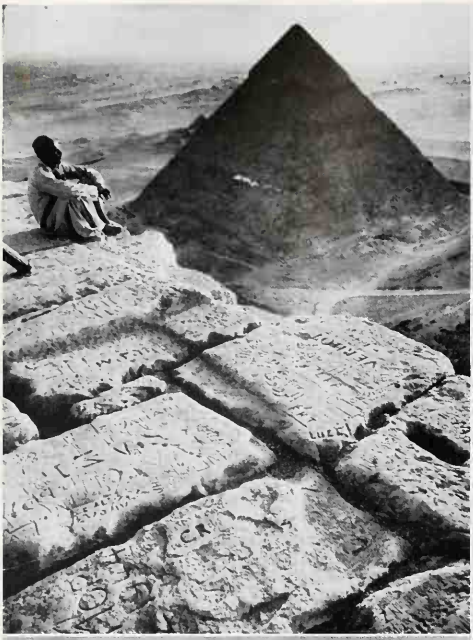


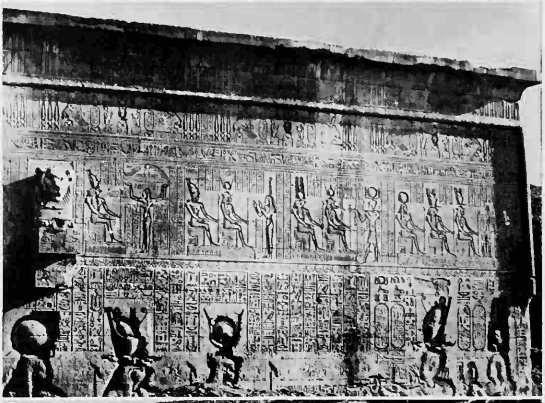
PHOTOGRAPH, A. F. KERSTING

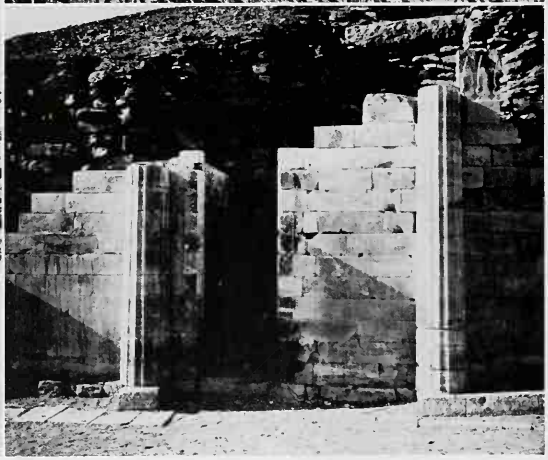
DAYR AL BAHRI

Rock-cut mortuary temple of Queen Hatshepsut, 16th century B.C.

EGYPTIAN ARCHITECTURE







BY COURTESY OF (CENTRE RIGHT) THE TRUSTEES OF THE BRITISH MUSEUM, (BOTTOM RIGHT) CECIL M. FIRTH; PHOTOGRAPHS, (TOP LEFT) PETER SCHMID—PIX FROM PUBLIX, (TOP RIGHT, BOTTOM LEFT) A. F. KERSTING

PYRAMIDS AND RELIGIOUS ARCHITECTURE

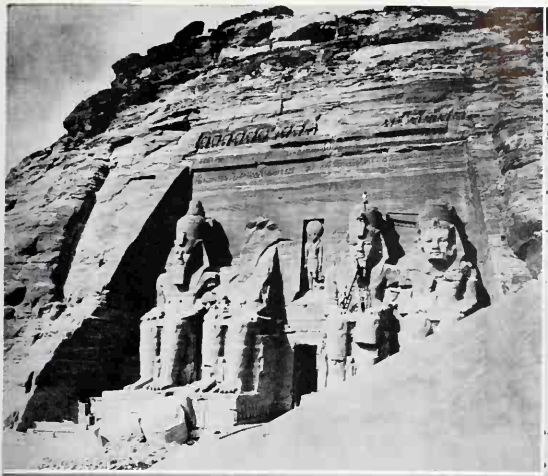
Top left: View of the pyramid of Khafre at Giza from the pyramid of Khufu. 4th dynasty

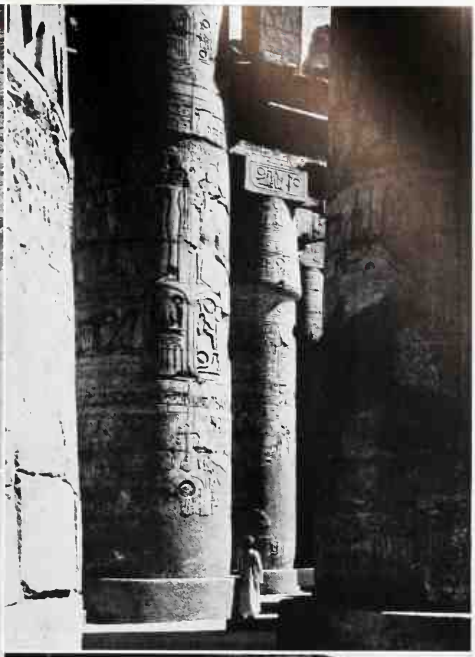
Top right: The Step pyramid at Saqqarah, constructed of stone blocks used as brick. 3rd dynasty

Centre right: South wall of the temple of Hathor at Dandarah. Ptolemaic period

Bottom left: The Great Sphinx at Giza, 189 ft. long. 4th dynasty

Bottom right: Chapel at Saqqarah. 3rd dynasty





BY COURTESY OF (TOP LEFT) THE ORIENTAL INSTITUTE, THE UNIVERSITY OF CHICAGO. (BOTTOM LEFT) PROFESSOR FLINDERS PETRIE. (BOTTOM RIGHT) GADDIS AND SEIF PHOTOGRAPH FROM THE ORIENTAL INSTITUTE. PHOTOGRAPHS (TOP RIGHT) G. E. KIDDER SMITH, (CENTRE LEFT) A. F. KERSTING

THE PYRAMID OF MAYDUM AND TEMPLES OF THE 18TH AND 19TH DYNASTIES

Top left: Expedition members measuring the head of one of the colossi in front of the temple of Ramses II, Abu Simbel. 19th dynasty. The entrance to the temple was cleared after this photograph was taken

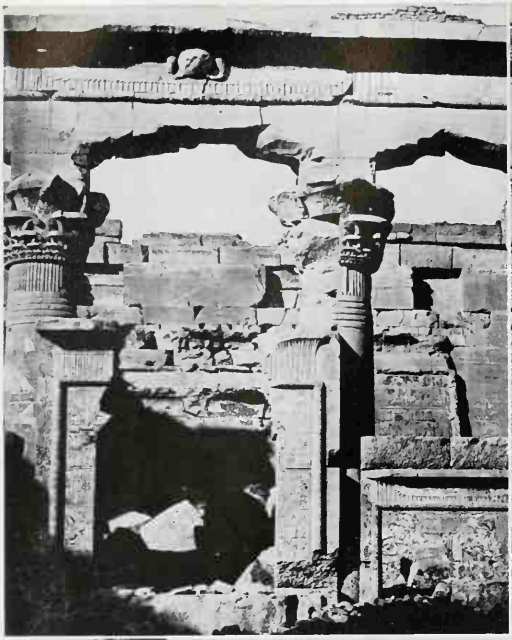
Top right: Carved columns in the hypostyle hall of the Great Temple of Amon at Karnak. 19th dynasty

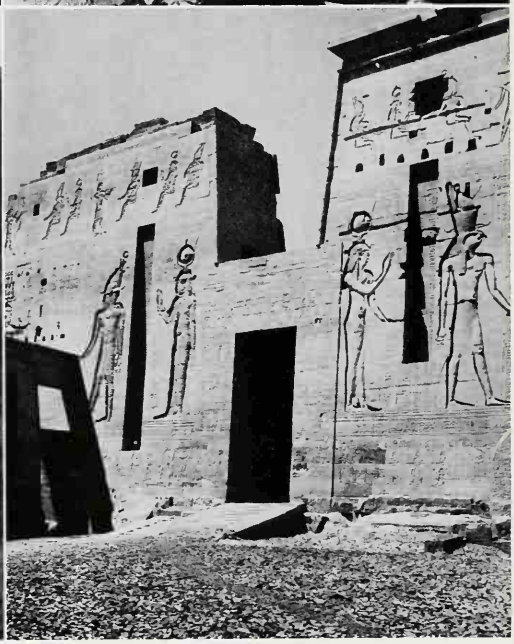
Centre left: Columns of the temple of the Theban triad at Luxor. 18th dynasty

Bottom left: Pyramid of Maydum, a stage in the development of pyramid building between the Step pyramid of Saqqarah and the three at Giza. 3rd dynasty

Bottom right: Red granite pillars in front of the sanctuary in the temple of Amon, Karnak. The inscriptions and reliefs on the pillars are attributed to Thutmose III. 18th dynasty

EGYPTIAN ARCHITECTURE





BY COURTESY OF (TOP LEFT, TOP RIGHT) THE TRUSTEES OF THE BRITISH MUSEUM, (BOTTOM LEFT) THE EGYPT EXPLORATION SOCIETY; PHOTOGRAPH, (BOTTOM RIGHT) A. F. KERSTING

TEMPLES, 18TH DYNASTY TO ROMAN PERIOD

Top left: Temple at Kalabishah, lower Nubia. Capitals, Roman period (1st century A.D.)
Top right: Obelisk of Thutmose III, temple of Amon, Karnak. 18th dynasty
Bottom left: Wall of the temple of Seti I, Abydos. 19th dynasty
Bottom right: Inner pylon of the temple of Isis, Philae (Jazinat Filah), upper Egypt. Ptolemaic period

own, including very archaic features. Some passages are believed to have been composed before the 1st dynasty.

Middle Egyptian.—Middle Egyptian is the classical stage of ancient Egyptian. It developed from Old Egyptian and was based on the spoken language of about 2200 B.C., being used for all purposes from that time until the reign of Ikhnaton (about 1379 B.C.). The continuing development of the language during this period is reflected by the occasional use of forms which were to become standard only in Late Egyptian. The folk tales of the Papyrus Westcar show many such influences. Another example is found in the tomb of the noble Paheri at El Kab, where the workers pictured on the walls speak Late Egyptian, while the rest of the inscriptions are in Middle Egyptian. After it was replaced by Late Egyptian for practical purposes, Middle Egyptian continued to be used for most monumental inscriptions and for some literary compositions into Christian times. The hieroglyphic text of the Rosetta stone (196 B.C.) was an attempt to write the classical language.

Late Egyptian.—The official use of Late Egyptian began as one of the reforms made in the reign of the heretic king Ikhnaton. While Old and Middle Egyptian were essentially the same language, Late Egyptian showed striking differences. Definite and indefinite articles were used (a practice which continued into Coptic), the old verb forms were being replaced, many phonetic changes occurred and numerous new words appeared. The texts are mainly manuscripts, but some inscriptions are also known. Although it was much closer to the spoken language of its day, it never entirely replaced the older language for written use.

Demotic.—This was the ordinary language used for official acts and other documents, including some literary texts, beginning in the Saite period and continuing into Roman times. During the latter part of this period Greek was also used. Contemporary hieroglyphic texts, such as those used in temples and tombs, or for royal decrees, were in a kind of Middle Egyptian, more or less influenced by demotic, which represented the spoken language. In a few inscriptions demotic is written in hieroglyphs. Ordinary demotic documents, including literary works and some religious texts, are in demotic script. The middle section of the Rosetta stone is a good example of a demotic inscription in demotic characters. While it holds a position intermediate between Late Egyptian and Coptic, demotic is in general closer to the former. The script is difficult to read, and much reliance must be placed on comparisons with earlier Egyptian and with Coptic in decipherment. Definite traces of dialect distinctions, which may be related to those of Coptic, are found in demotic texts. (*See* also DEMOTIC LANGUAGE AND WRITING.)

Coptic (From the 3rd Century A.D.).—During the Roman and Byzantine periods Greek was the most common written language in Egypt, being used for governmental purposes as well as for more popular use. As early as the 2nd century A.D. texts were written in Egyptian but in Greek letters, breaking with the old hieroglyphic tradition. These were mostly horoscopes, magic spells and the like. Beginning in the 3rd century a Christian literature appears, written in Greek letters supplemented by seven characters taken from demotic. The language was Egyptian as it was then spoken and is known as Coptic (from the Greek *Aigyptos* "Egypt"). Aside from the difference of linguistic structure between demotic and Coptic, there is a very marked change in vocabulary and general tone. The shift from the one to the other reflects the change from paganism to Christianity, the older priestly words and expressions being largely replaced by ecclesiastical ones borrowed from Greek.

The precise date at which Coptic ceased to be a living language is not known. By the 11th century it was on the decline and gradually gave way to Arabic, which is the present-day speech of Egypt. The traveler P. Vansleb, who visited Egypt in 1672 and 1673, was surprised to find someone who still spoke Coptic. As the historian Maqrizi mentions in the 15th century that Coptic was still spoken by the women and children in upper Egypt, it is generally assumed that it died out between his time and that of Vansleb, that is, during the 16th century. In 1936–37 a few peasants were found near Thebes who could not only read Coptic but could also speak a few phrases. The Bohairic dialect is still the liturgical language of the Coptic Church, but the pronunciation is based on the values of the letters in Modern Greek and is of no value for the history of the language. (*See* also COPTIC LANGUAGE.)

Morphology.—The development of the verbal system from Old Egyptian to Coptic is perhaps the most striking feature of this long linguistic history. Besides the old perfective, Old Egyptian had a series of suffix-conjugations which remained the major verb forms in use through Middle Egyptian. There were also expressions using verbal nouns, such as "he is upon hearing" for "he hears." By Late Egyptian times these verbal phrases had increased both in number and use, and in Coptic they have developed into a complete set of verb forms (present, future, perfect, conjunctive, optative, causative, and others), while only fossil fragments of the older suffix conjugations remain. Curiously, the old perfective, which had been inherited from proto-Afro-Asiatic, survived into Coptic as the qualitative, while suffix conjugations which arose within Egyptian passed away.

Phonology.—Fairly accurate deductions may be made about the phonetic value of the consonants of the older stages of Egyptian, but next to nothing is known about the vowels. Coptic, as the only stage of the language in which the vowels are regularly and completely written, is the main source of information for the reconstruction of the older vocalization. The most important other sources are Egyptian words written in cuneiform documents (particularly the Amarna letters) and in Greek. Since most of the older verb forms are not found in Coptic, the reconstructions are limited in scope.

BIBLIOGRAPHY.—Elmar Edel, *Altaegyptische Grammatik I, Analecta Orientalia*, 34 (1955); Alan Gardiner, *Egyptian Grammar*, 3rd ed. rev. (1957); Adolf Erman, *Neuaegyptische Grammatik*, 2nd ed. rev. (1933); František Lexa, *Grammaire demotique* (1947–51); Walter Till, *Koptische Dialektgrammatik* (1931), *Koptische Grammatik* (1955). (C. T. HE.)

EHIME, Japanese prefecture of northwestern Shikoku, facing the Inland sea (north) and Bungo strait (west). Area 2,182 sq.mi. Pop. (1960) 1,500,687. Its interior is mountainous and most of the population is grouped on shallow coastal plains. Agricultural products include rice, wheat, tea, soybeans and mandarin oranges. Northern Ehime shows rapid industrial advances in cities such as Niihama (heavy chemicals and machinery), Saijo (synthetic fibres), Hashihama (salt) and Imabari (cotton textiles, especially towels). The capital and leading city, Matsuyama, is Shikoku's largest city. Fishing and forestry support Uwajima and other west coast cities. (J. D. EE.)

EHRENBERG, CHRISTIAN GOTTFRIED (1795–1876), German biologist, whose outstanding work on microscopic organisms laid the foundations of systematic bacteriology and protozoology, was born at Delitzsch in Saxony on April 19, 1795. He entered Leipzig university where he read medicine, continuing his studies in the University of Berlin, where he was appointed professor of history of medicine in 1827. However, this appointment was purely nominal, for he devoted most of his time to the study of microorganisms. His most important publication was *Die Infusionsthierchen als Volkommene Organismen* (1838), in which he described—under the name "Infusoria"—all the known bacteria and protozoa, as well as rotifers, diatoms and desmids, all of which were regarded as animals. This monumental work did for the classification of microorganisms what the *Systema Naturae* of Linnaeus did for the higher plants and animals, although Ehrenberg erroneously considered that all the "Infusoria" (Ciliata) were "complete organisms," *i.e.*, multicellular, whose internal structure was similar to that of the higher animals. Nevertheless, the organisms described by him were accurately depicted in the excellent illustrations accompanying this book. Another important work, *Mikrogeologie* (1854), deals with fossil microorganisms.

Ehrenberg's scientific achievements were recognized by his election to many learned societies, including the Berlin Academy of Sciences (1827), the Royal society (1837) and the Institut de France (1860). He died in Berlin on June 27, 1876.

See M. Laue, *Christian Gottfried Ehrenberg* (1895). (C. A. HE.)

EHRENBREITSTEIN: *see* COBLENZ.

EHRENFELS, CHRISTIAN, FREIHERR VON (1859–1932), Austrian philosopher who introduced the expression "Gestalt" into psychology and also made an important contribution to value theory, was born at Rodaun near Vienna on June 20, 1859. At the University of Vienna he came under the influence of Franz Brentano and Alexius Meinong and was licensed as *Privatdozent* in 1888. He was appointed extraordinary professor of philosophy at the German University of Prague in 1896 and was ordinary professor there from 1900 to 1929. He died at Lichtenau in Lower Austria on Sept. 8, 1932.

Ehrenfels' article "Über Gestaltqualitäten," in the *Vierteljahrschrift für wissenschaftliche Philosophie*, xiv (1890), was the starting point of Gestalt psychology. His *System der Werttheorie*, 2 vol. (1897–98), was also a pioneer work, treating the concept of value psychologically, as a function of desire (*see* VALUE, THEORY OF). Ehrenfels also published plays and choral dramas (from 1887 onward); two pamphlets on Richard Wagner (1896 and 1913); *Grundbegriffe der Ethik* (1907); *Sexualethik* (1907) and articles on sexual morality; *Kosmogonie* (1916); *Das Primzahlengesetz* (1922); and *Die Religion der Zukunft* (1929).

EHRLICH, PAUL (1854–1915), German medical research worker, a great pioneer and experimental genius in the application of chemistry to biological and medical science, who was awarded the Nobel prize for medicine in 1908 (shared with Ilya Mechnikov for work on immunity), was born of Jewish parentage at Strehlen (now Strzelin), Silesia, on March 14, 1854. At school, except in mathematics and Latin, his record was undistinguished, and he detested examinations. At the university he followed medicine and was attracted to chemistry, for which he showed great aptitude, being largely self-taught. On qualifying (1878) he became a clinical assistant for some years, set up a small laboratory and quickly aroused attention by his work on the blood cells. Later he transferred to Robert Koch's Institute for Infectious Diseases, and in 1896 was made director of the Institute for Serum Investigation at Steglitz (Berlin), moving in 1898 to the new Institute for Experimental Therapy, Frankfurt am Main. In 1906 he also became director of the adjacent Georg Speyer Haus for Chemotherapy. At first he met considerable opposition, but later received many honours including the foreign membership of the Royal Society of London and the German title of "Excellency." He died at Bad Homburg on Aug. 20, 1915.

Ehrlich's principal researches were successively in histology and blood cytology, immunity and chemotherapy, each occupying a period of 10 to 15 years; in addition, chemical investigations proceeded throughout, and he also made important advances in bacteriology, pharmacology and cancer research. While Ehrlich's prime object was new knowledge, the practical results were epoch-making. The following outlines his chief contributions. For the tinctorial analysis of tissues, Ehrlich began by investigating the biological properties of synthetic dyes, which German science and chemical manufacture produced in profusion. He demonstrated the specific staining characters of various granules in tissue cells and leucocytes and founded modern hematology. He also discovered "vital stains" by which tissues in living animals are coloured without any poisonous side-effects; *e.g.*, nerve fibres by methylene blue. This represented an immense methodological advance. By using dyes susceptible of reversible oxidation and reduction he proved that different tissues of living animals vary in their degree of saturation with oxygen. His demonstration of Koch's bacillus by acid-fast staining provided an indispensable tuberculosis diagnostic method.

The treatment of diphtheria by antitoxin was originally hampered by the uncertain potency of commercial products. Ehrlich developed a method of standardizing such antiserums, which is now in practically universal use. His investigations on the mode of action of bacterial toxins and similar poisons (*e.g.*, snake venoms and cytotoxins), and the neutralizing antibodies developed in the serum of living animals after immunization with antigens, formed a main foundation of immunology. From those researches especially his "side-chain theory" was evolved.

Chemotherapy (*q.v.*), a branch of pharmacology created by Ehrlich, now supports a flourishing chemical industry. The cure of an infected animal by a drug which damages the causal parasites and leads to their sterilization was first announced in 1904. Mice inoculated with *Trypanosoma equinum* were treated successfully by a single subcutaneous injection of the dye trypan red at a time 24 hours before death of untreated controls, when the organisms were abundant in the blood. With his co-workers Ehrlich established the principles of chemotherapeutic biology. The greatest chemotherapeutic achievement was the discovery of the synthetic arsenical compound arsphenamine, once sold commercially as Salvarsan (*q.v.*) or 606. It proved effective in rabbits and apes experimentally infected with the causal organism of syphilis and revolutionized treatment of the human disease, although later superseded. The fact must be emphasized that usually cure by chemotherapeutic drugs results from co-operation with the host's defense mechanisms. Various biological reactions on the part of either host or parasites may account for failures.

According to Ehrlich, his discovery of principles depended on analyzing striking experimental findings. From experiments on the oxygen requirement of the tissues (1885) he postulated that the living protoplasm of cells contains a special central grouping which determines their specific activities; and attached to this are atom-complexes as side chains (receptors). The latter, although subordinate for specific functions, are nevertheless all-important for each cell's collective life. The hypothesis is based on structural chemistry. Ehrlich's researches on immunity and chemotherapy confirmed this view and explained rationally the action of toxins and drugs—these act only on cells whose side chains combine with them. It also accounted for the production of antibodies as free receptors, and allied phenomena. This side-chain theory, redressed, is still dominant.

Ehrlich's collected papers and a complete bibliography, edited by F. Himmelweit, were published in four volumes (1956). The text is in German, English and French.

BIBLIOGRAPHY.—H. Apolant *et al., Paul Ehrlich* (1914); M. Marquardt, *Paul Ehrlich* (1949); C. H. Browning, "Emil Behring and Paul Ehrlich: Their Contributions to Science," *Nature*, vol. 175, pp. 570, 616 (1955). (C. H. Bc.)

EHUD (AOD in the Douai version of the Bible), in the Old Testament, a judge who delivered Israel from the Moabites (Judg. iii, 12–30), by assassinating Eglon, king of Moab, and raising the tribe of Ephraim to seize and hold the fords against the fleeing Moabite garrisons. He is called the son of Gera, a Benjaminite, but since Gera and Ehud are tribal names, it has been supposed that this notice is not original.

EICHELBERGER, ROBERT LAWRENCE (1886–1961), U.S. army officer, commander of the 8th army in World War II, was born March 9, 1886, in Urbana, O., and graduated from the U.S. Military academy at West Point, N.Y., in 1909. From 1918 to 1920 he served in Siberia with the American expeditionary forces, later attended various U.S. army schools, was superintendent of West Point from Oct. 1940 to Jan. 1942, and then led the U.S. 1st corps to Australia. In Jan. 1943 he directed the attack on Buna, New Guinea, one of the first ground victories against the Japanese in World War II.

In Sept. 1944 Eichelberger took command of the newly activated U.S. 8th army in Netherlands New Guinea. He led this force through many severe engagements, including the Philippine Islands campaign, where he made effective use of amphibious tactics. On Aug. 30, 1945, he landed with a small detachment at Atsugi air field to start the occupation of Japan. During the next three years he directed the 8th army military government units throughout Japan. He retired as lieutenant general in Dec. 1948 and was promoted to full general in Nov. 1954. He published his memoirs, *Our Jungle Road to Tokyo*, in 1950.

Eichelberger died in Asheville, N.C., on Sept. 26, 1961.

(R. R. SH.)

EICHENDORFF, JOSEPH, FREIHERR VON (1788–1857), German poet, the one among the followers of the Heidelberg romantics (*see* GERMAN LITERATURE) whose work had the most popular appeal, was born on March 10, 1788, at Schloss Lubowitz, near Ratibor (Silesia). After leaving the Breslau *Gymnasium*, he

studied law at Halle (1805–06) and then at Heidelberg (1807–08), where at that time Joseph Görres was lecturing and Clemens Brentano and Achim von Arnim were bringing out their famous folk-song collection *Des Knaben Wunderhorn.* After leaving Heidelberg, Eichendorff met the leaders of the romantic national movement—Heinrich von Kleist, Gottlieb Fichte and Adam Müller—in Berlin (1809–10). He then went to Vienna to work for the Austrian civil service examinations, but in 1813, when the Prussian War of Liberation broke out, enlisted in the famous *Lützowsche Freikorps.* After the war he held posts in the Prussian civil service in Danzig and Königsberg (1821–31) and after 1831 in Berlin. In 1844 he retired and lived sometimes in Danzig, sometimes in Berlin, Vienna, Dresden and Köthen, and finally in Neisse (Upper Silesia), where he died, Nov. 26, 1857.

The French Revolution was the central event in Eichendorff's life and he returned to it again and again (*Schloss Durande,* 1837; the epic poem *Robert und Guiscard,* 1855; the prose *Der Adel und die Revolution,* 1857). It was the revolution and the Napoleonic Wars which brought about the decline of the Eichendorff family fortunes and the loss of Lubowitz; and it is the sorrow over Lubowitz that gives to Eichendorff's poems their inimitable nostalgia, which echoes so compellingly in Hugo Wolf's setting of "Heimweh" and Schumann's of "Aus der Heimat hinter den Blitzen rot." Eichendorff's poems have indeed caught the imagination of composers from Mendelssohn and Brahms to Hans Pfitzner and Richard Strauss. At the same time they make their full impact only when they are read in connection with the stories in which they are embedded and which Wolf called "landscapes composed in poetry."

The most important of Eichendorff's prose works are the novels *Ahnung und Gegenwart* (1815) and *Dichter und ihre Gesellen* (1834) and the *Novellen Das Marmorbild* (1819), *Die Entführung* (1839), *Die Glücksritter* (1841) and the posthumous *Eine Meerfahrt.* By far the best known, however, both in German and through many translations, is *Aus dem Leben eines Taugenichts* (1826; Eng. trans. *Memoirs of a Good-for-Nothing,* 1955). Of this romantic picaresque novel in miniature Thomas Mann wrote "It is nothing other than dreams, music, delicious ease,—the winding of the posthorn, wanderlust, homesickness, fireworks falling on a darkening park, a foolish bliss—so that the ears echo and the head hums with poetic enchantment and confusion." (*Betrachtungen eines Unpolitischen,* S. Fischer Verlag, Frankfurt am Main, 1918).

As a dramatist, Eichendorff was unsuccessful in his satirical and historical plays but his comedy *Die Freier* (1833), which shows Shakespearian influence, has held its place on the stage. The Christian viewpoint which determined Eichendorff's attitude to poetry is expressed in his various critical works (*Geschichte der poetischen Literatur Deutschlands,* 1857, etc.). He made exemplary translations from the Spanish of Calderón's religious plays (1846–53), Don Juan Manuel's *El Conde Lucanor* and five interludes of Cervantes.

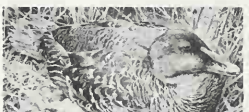
BIBLIOGRAPHY.—H. von Eichendorff, *Eichendorff, sein Leben und seine Schriften* (1864); *Gesammelte Werke,* critical ed. by A. Sauer and W. Kosch, continued by H. Kunisch (1908–); *Werke* (complete, without commentary, ed. by G. Baumann, 4 vol., 1957–58); H. Brandenburg, *J. von Eichendorff* (1922); W. Köhler, *J. von Eichendorff, ein Dichterleben* (1957). (W. E. Ko.)

EICHHORN, JOHANN GOTTFRIED (1752–1827), German biblical scholar and orientalist, one of the first to attempt scientific comparisons between biblical and other Semitic writings, was born at Dörrenzimmern on Oct. 16, 1752. He became professor of oriental languages at Jena in 1775 and of philosophy at Göttingen in 1788. It was Eichhorn's conclusion that "most of the writings of the Hebrews have passed through several hands," and he took for granted that all the so-called supernatural facts relating to the Old and New Testaments were explicable on natural principles. He regarded many books of the Old Testament as spurious, questioned the genuineness of II Peter and Jude, denied the Pauline authorship of Timothy and Titus, and suggested that the canonical gospels were based upon various translations and editions of a primary Aramaic gospel. Eichhorn died at Göttingen on June 25, 1827.

Among his principal works were *Einleitung in das Alte Testament,* three volumes (1780–83); *Einleitung in das Neue Testament* (1804–12); and *Einleitung in die apokryphischen Bücher des Alten Testaments* (1795).

EICHRODT, WALTHER (1890–), German professor who showed the importance to biblical studies of the theology of the Old Testament, was born on Aug. 1, 1890, at Gernsbach. After studying theology at Bethel, Greifswald, Heidelberg and Erlangen he taught at Bethel and Erlangen, then became professor of Old Testament at Basel (1922), where he was later rector (1953–55). His chief work, *Theologie des Alten Testaments* (two volumes, 1933–35; 4th ed. 1957, Eng. trans. 1961), marked the beginning of a new epoch in Old Testament studies. Without reducing the theology of the Old Testament to the history of Israelite religion, Eichrodt made extensive use of the results of literary and comparative analysis to envisage the religion of the Old Testament as a unity of permanent reality throughout the vicissitudes of history. This theology he found, was most aptly characterized by the word "covenant," a term denoting action rather than speculation. The triple aspect of God's covenant, with his people, with the world and with man, formed the plan of his book. By this method he presented the great dogmatic realities in a dialectic appropriate to the Old Testament, preserving both the historical character of the revelation and the unity of the Old and the New Testaments. Eichrodt's other principal works include *Die Quellen der Genesis* (1916), *Die Hoffnung des ewigen Friedens im alten Israel* (1920), *Das Menschenverständnis des Alten Testaments* (1946; Eng. trans. 1951), *Israel in der Weissagung des Alten Testaments* (1951), *Gottes Ruf im Alten Testament* (1951) and *Die Heilige in Israel* (1960). (E. J.)

EIDER, any of several large sea ducks of arctic and subarctic regions, all of which have two to three inches of thick downy plumage. The strikingly handsome males are usually velvety black below and snowy white above, whereas the females are protectively rich-brown and barred. Four to ten greenish to olive-buff eggs are laid in a well-made nest, which the female lines thickly and leaves covered during her absence, with very fine, soft grayish down plucked from her breast. In northern Europe eggs and down, collected at intervals of several days for market, are replaced by the female, who is finally permitted to rear a family.



HUGH M. HALLIDAY FROM NATIONAL AUDUBON SOCIETY

COMMON EIDER (SOMATERIA MOLLISSIMA): FEMALE INCUBATING EGGS

The best-known species is the common eider (*Somateria mollissima*) of northern Europe and Iceland, about 2 ft. long. It is represented in northeastern North America by the northern eider (*S. m. borealis*); a more southern race known as the American eider (*S. m. dresseri*), around Hudson bay and from Labrador to Penobscot bay, Me.; and the Pacific eider (*S. m. v-nigra*), of the coasts of northeastern Siberia and Alaska and the Aleutian Islands. The Pacific eider, 26 in. long, one of the largest of North American ducks, has a striking black V-shaped mark on the white throat. In the same region with the Pacific eider are found the spectacled eider (*Lampronetta fischeri*), with a greenish-yellow crown and large white "spectacles" outlined in black, and Steller's eider (*Polysticta stelleri*), a fast, trim little duck, the smallest of the eiders, about 18 in. long, with black neck and throat and chestnut underparts. The king eider (*S. spectabilis*), a circumpolar species the size of the common eider, is the most strikingly marked of all, with an orange-yellow bill-process or forehead-plate, bordered with black, and a pearl-gray crown. A race of the common eider known as *S. m. faeroeensis* is resident on the Faeroes, north of the British Isles toward Iceland. (G. F. Ss.)

EIDER RIVER, in Schleswig-Holstein, north Germany, rises to the south of Kiel, in Lake Redder, flows first northward, nearly as far as Kiel, then bends westward and flows across the low peninsula in a sluggish, winding course of 117 mi. Tönning stands at the head of its long shallow estuary. It is navigable up to Rendsburg, and is embanked through the marshes across which it runs in its lower course. From the reign of Charlemagne, the Eider (origi-

nally Ägyr Dör, "Neptune's gate") was known as *Romani terminus imperii* and was recognized as the boundary of the empire in 1027 by the emperor Conrad II, the founder of the Salian dynasty. In the controversy arising out of the Schleswig-Holstein question, which culminated in the war of Austria and Prussia against Denmark in 1864, the Eider gave its name to the Eider Danes, the intransigent Danish party which maintained that Schleswig (Sonderjylland, South Jutland) was by nature and historical tradition an integral part of Denmark. The Eider canal, which was constructed between 1777 and 1784, leaves the Eider at the point where the river turns to the west and enters the Bay of Kiel at Holtenau. It was hampered by six sluices but was used annually by about 4,000 vessels, and until its conversion in 1887–95 into the Kaiser Wilhelm (later Nord-Ostsee) canal it afforded the only direct connection between the North sea and the Baltic. *See* KIEL CANAL.

EIDETIC IMAGE is a subjective visual phenomenon that assumes a perceptual character and resembles a negative or positive afterimage in that it is seen in the literal sense of the word. An eidetic person not only can imagine an absent object but really can see it, either with closed eyes or while looking at some surface that serves as a convenient background for his image. A particular object may be seen eidetically either immediately after its disappearance or removal from sight or after a lapse of several minutes, days or even years; there are also spontaneously appearing eidetic images. In some instances, eidetic images and the objects they represent differ in colour, form, apparent size, position in space, richness of details and many other characteristics; in other instances, objects may be eidetically reproduced with almost photographic fidelity. While V. Urbantschitsch's investigations (1907) still suggested a pathological basis for eidetic phenomena, Oswald Kroh's studies (1917) established the fact that eidetic images are frequently found in normal children. E. R. Jaensch advanced the view that most persons pass through an eidetic phase during childhood and, furthermore, that eidetic images, memory images and afterimages are distinct phenomena forming three "levels of memory." Studies in various European countries and the United States brought out marked differences in regard to the number of eidetic persons in different geographic regions. The figures reported for children range from 0 to 100% for different localities. Phenomena corresponding to eidetic images in the visual field are believed to exist in other sense fields as well.

Experimental studies of eidetic phenomena have not been confined to problems of sensory psychology and physiology. They have also been concerned with such problems as the frequency of eidetic imagery at different age levels and in different vocational groups; the relation to intelligence and educational achievement; the relation to personality and constitution, particularly to certain "biotypes"; and the clinical significance of these phenomena, especially in the fields of neurology and psychiatry. Some investigators have assumed an eidetic basis for clairvoyance, second sight and other parapsychological phenomena. *See* also PERCEPTION.

BIBLIOGRAPHY.—E. R. Jaensch, *Eidetic Imagery* (1930); H. Klüver, "Studies on the Eidetic Type and on Eidetic Imagery," *Psychol. Bull.*, vol. 25 (1928), "Eidetic Imagery," in *A Handbook of Child Psychology* (1933), "The Eidetic Type," *Res. Publ. Ass. Nerv. Ment. Dis.* (1933), and "Mechanisms of Hallucinations," in *Studies in Personality* (1942); T. Husén, "Studier rörande de eidetiska fenomenen," *Lunds Univ. Årsskrift, N.F., Avd. 1* (1945, 1952). (H. KL.)

EIFEL, a region of Germany, lies between the Rhine, the Moselle and the Luxembourg frontier. The Eifel is an ancient massif of folded Devonian rocks (slates, sandstones and limestones), with a northeast-southwest structural trend, and with Triassic sandstones, marls and limestones resting unconformably at its margins near Hillesheim, Bitburg and Trier. Permian Rothliegende deposits occur near Wittlich.

The plateau falls into four sections: Hohes Venn, Hohe Eifel, Schneifel (Schnee Eifel) and Voreifel. The Hohes Venn, lying beyond the Belgian frontier, is an extensive plateau of impervious rocks with much bog, gorse, broom and cotton grass. Eastward a trough (about 20 mi. wide) lies between the Hohes Venn and the Hohe Eifel. This is a rolling plateau trending northeast-southwest with Bunter Sandstones and Devonian limestones in shallow synclines; valleys are deeply incised into its rolling surface. In this trough is the Schneifel. Scrub and forest are common, cultivation occurring only on the richer soils. The trough extends from Trier in the south to the Ruhr-Urft valley in the northeast. The Hohe Eifel (1,640–3,280 ft.) is a dissected highland drained to the east by the Ahr river, a vine-growing area. The Voreifel (1,640 ft.) sloping south to the Moselle, has a smooth surface dissected by tributaries of the Moselle. Evidence of volcanic action can be seen in the explosion craters (*Maare*) and small cones. There is further evidence of vulcanism in the eastern Hohe Eifel (Pellenz and Maifeld). Lava and tuffs make up much of this lowland. Basalt from Mayen is used for building and millstones; tuffs and pumice are also quarried in the Pellenz.

Settlement began on the periphery of the Eifel. The plateau from Bitburg to Cologne reveals evidence of settlement from ancient to medieval times; in the middle ages iron, lead and zinc were mined. The Maifeld was an area of early German settlement. Near the Laacher See an abbey was founded in the 11th century. Present settlement dates mainly from the clearing of the forest in the middle ages when monasteries and castles were established. The three-field and common pasture systems remain as evidence of this period. Scrub is still cut, burned and planted with a rotation of rye, potatoes, oats and 10 years' fallow. Holdings are small and land minutely divided. Metalworking industries have disappeared. There has been steady emigration since 1870 to the Ruhr and Aachen. (R. E. DI.)

EIFFEL, ALEXANDRE GUSTAVE (1832–1923), French engineer and bridge builder, whose most notable achievement was the Eiffel tower built for the Paris exhibition of 1889, was born in Dijon on Dec. 15, 1832. He attended the École Centrale des Arts et Manufactures, Paris, from 1852 to 1855. Although he worked on buildings for the Paris expositions of 1867, 1878 and 1889, Eiffel's greatest interest was in bridges. He was the master of airy two-hinged arches, such as that over the Douro river in Portugal (1877; span 530 ft., height 200 ft.), which was constructed from the piles without scaffolding. The Garabit viaduct in southern France (1880–84) is the most daring of his bridges (arch span 550 ft., height 408 ft., total length 1,667 ft.). In 1885 he designed the inner structure for the Statue of Liberty in New York harbour. The Eiffel tower (*see* TOWER and STEEL CONSTRUCTION), embodying Eiffel's lifelong experience of bridge building in Europe, Africa and Indochina, stands as a monument to 19th-century engineering skill.

After having been involved in the unfortunate French Panama canal venture (1893), Eiffel spent his last years studying the effects of air motion on airplanes. He built a wind tunnel (1912) and aerodynamic laboratory and used the tower for experiments. He died in Paris on Dec. 23, 1923.

See Jean Prevost, *Eiffel* (1929). (S. GN.)

EIJKMAN, CHRISTIAAN (1858–1930), co-recipient of the 1929 Nobel prize in medicine and physiology for his discovery of antineuritic vitamins, was born in Nijkerk, Neth., on Aug. 11, 1858. He studied medicine at the University of Amsterdam, from which he graduated in 1883, and then went to Berlin to study bacteriology under Robert Koch. In 1886 he went to Batavia in the Netherlands Indies (now the Republic of Indonesia) and in 1888 became the first director of the research laboratory for pathological anatomy and bacteriology. For the next ten years he carried out his classic investigations on beriberi, a disease long endemic in the Orient, the cause of which was unknown. Eijkman was able for the first time to induce the disease experimentally in fowls by feeding them on polished rice. Together with his collaborator, G. Grijns, he showed that it was possible to cure the disease in birds by adding to their diet the discarded rice polishings or husks. These researches were one of the major steps which led to the discovery of vitamins and to the modern understanding of the nature of nutritional deficiency diseases. In 1898 Eijkman returned to Holland to become professor of hygiene at the University of Utrecht, a position he held until his retirement in 1928. He died in Utrecht on Nov. 5, 1930. (M. C. L.)

EILDON HILLS, three conical volcanic hills in Roxburghshire, Scot., 1 mi. S.S.E. of Melrose. They were once known as Eldune—the Eldunum of Simeon of Durham (d. after 1129)—probably from the Gaelic *aill,* "rock," and *dun,* "hill," or the Cymric *moeldun,* "bald hill." The northern peak, which bears prehistoric remains, is 1,327 ft. high, the central peak 1,385 ft. and the southern peak 1,216 ft. The Roman camp of Trimontium was located nearby. The Eildons have been the subject of much legendary lore. They were said to have been split in three by a demon on the order of Michael Scot. The Eildon tree stone, a large moss-covered boulder lying on the high road as it bends toward the west within a mile of Melrose, marks the spot where the queen of the fairies led Thomas of Erceldoune (Thomas the Rhymer), the earliest Scottish poet of whom any record is held, into her realms in the heart of the hills.

EILEITHYIA, a pre-Hellenic goddess of childbirth. The earliest evidence for her cult is at Amnisos, the port of Knossos in Crete. Her cave there is mentioned in the *Odyssey,* and excavation has shown that she was worshiped there continuously from Neolithic to Roman times. The occurrence of the name of the goddess on Minoan Linear B tablets finally confirms that it is primitive and pre-Greek in derivation. In Homer, Eileithyia appears sometimes in the plural as a personification of the birth pangs and is described as the daughter of Hera, consort of Zeus. The typical legend describes how Hera attempts to prevent the births of her enemies, Heracles, Apollo and Artemis, by hindering Eileithyia's work. Eileithyia's cult is chiefly known from inscriptions, which show that it was plentifully distributed in Crete, but that it was also to be found in the Cyclades islands, north of Crete, at Sparta, at Athens and in Boeotia.

In later times Eileithyia tended to be identified with Hera or Artemis (*qq.v.*), because they were the goddesses concerned with marriage and childbirth.

See M. P. Nilsson, *Geschichte der griechischen Religion,* 2nd ed., vol. i, pp. 312 ff. (1955); H. J. Rose, *A Handbook of Greek Mythology* (1933). (H. W. Pa.)

EILHART VON OBERGE (fl. late 12th century), German poet, a member of a Brunswick family mentioned in records of Henry the Lion, is of interest to students of the early German court epic and the Tristan story. His epic, *Tristrant und Isalde,* a laboured version of an Old French source now lost, dates from the last quarter of the 12th century. Uncertainty about his chronological position in relation to Heinrich von Veldeke (*q.v.*) and the corruptness of the early fragment and later complete but modified versions of his epic make it difficult to assess Eilhart's importance. His epic was popular, for it provided the basis of a 15th-century prose novel, *Tristan und Isalde,* and of a tragedy by Hans Sachs. Its relationship to the greater epic by Gottfried von Strassburg (*q.v.*) is clear but less significant.

The poem was edited by F. Lichtenstein (1877) and K. Wagner (1924). *See* also J. van Dam, *Zur Vorgeschichte des Höfischen Epos* (1923) and G. Cordes, *Zur Sprache Eilharts von Oberg* (1939). (W. W. Cs.)

EINDHOVEN, a town of the southeast Netherlands, province of North Brabant, lies on the Dommel river, 108 km. (67 mi.) S.E. of Rotterdam by road. Pop. (1957 est.) 157,621 (mun.). It was badly damaged in World War II and largely rebuilt. The chief centre is the new 18 September square bordered north by the new railway station, elevated railway yard and post office and south by the main shopping centre. There is a technical university (1957), the van Abbe Museum of Modern Art (1936) and an astronomical observatory founded by A. F. Philips in 1938. Eindhoven is an important railway centre with lines to the north and west and to Belgium and Germany. A helicopter service links it with Brussels and Duisburg, Ger., and it is also served by the Eindhovens Kanaal and Beatrixkanaal. The electrical plant (radio and television sets, electric bulbs, X-ray apparatus, etc.) founded by Philips in 1891 and reconstructed after World War II provides the chief industry. The manufacture of trucks and automobiles is also important, and there are tobacco and textile industries. Eindhoven was given a charter by Henry I, duke of Brabant, in 1232. Its population in 1900 was only 4,730, but in the 20th century it became a prosperous industrial town, the most important south of the big rivers. (F. Gi.)

EINHARD (Eginhard) (*c.* 770–840), Frankish court diplomat and scholar, friend and biographer of Charlemagne and the trusted associate of the emperor Louis I the Pious, was born about 770 and was educated at the abbey of Fulda, where he was a fellow pupil of Rabanus Maurus, afterward archbishop of Mainz. Some years later, as its notary, he was transacting important business for the abbey. Between 792 and 794 he joined Charlemagne's court, where, as one of the younger members (*nutriti*) of the royal household, he was taught by the English scholar Alcuin (*q.v.*). Acquainted with Greek and a sound Latinist, Einhard became a prominent member of the palace academy. He was sent to Rome (806) to obtain Pope Leo III's assent to the documents drawn up at Thionville partitioning the empire between Charlemagne's sons; and as spokesman for the magnates assembled at Aachen (Sept. 813) he formally requested Charlemagne to designate Louis emperor. He was familiar with classical works on architecture, particularly those of Vitruvius: both Charlemagne and Louis I made practical use of his knowledge, and he helped to plan the royal palace and church at Aachen. His most important contribution to court life, however, was his continuation of the role formerly played by Alcuin.

Einhard was not affected by the palace revolution which followed Charlemagne's death (814), and quickly attained the position of trusted counselor to Louis. The new emperor granted him extensive lands (815) and appointed him lay abbot of four great monasteries. He was charged with the education of Louis' eldest son Lothair from about 817 to 822, the years during which he was writing the *Vita Caroli.* He left Aachen about 830, as he did not want to attach himself to Wala or to be involved in the developing intrigues and jealousies of the emperor's sons. He devoted his remaining years to the administration of his abbeys and built one at Mulenheim (afterward Seligenstadt), where he died on March 14, 840.

Einhard's most famous work is the *Vita Caroli,* a source of primary importance for the reign of Charlemagne. A vivid and lively narrative, it is nevertheless marred by the author's deliberate imitation of entire passages from Suetonius. Many other works have been at some time mistakenly attributed to Einhard; but his claim on the attention of historians is sufficiently established by the unrivaled picture of the empire after 815 provided in his voluminous correspondence.

Einhard's letters and his life of Charlemagne are edited by P. Jaffé, *Bibliotheca rerum Germanicarum* vol. iv (1867). There is also an edition of the life of Charlemagne by L. Halphen, with French translation, *Eginhard, vie de Charlemagne* (1923).

See A. Kleinclausz, *Eginhard* (1942). (J. De.)

EINSIEDELN, a town in the Swiss canton of Schwyz, lies on the right bank of the Alpbach at 2,908 ft. above sea level and 41 km. (25.5 mi.) S.E. of Zürich by road. Pop. (1960) 8,792, almost all Catholics and German-speaking. The town is dependent on the Benedictine abbey that rises slightly above it to the east. On its site the hermit St. Meinrad built his cell about 828 bringing with him a small wooden statue of the Virgin. He was murdered by robbers in 861 and after 934, when the abbey was founded by hermits (*einsiedler,* from whom the town derived its name), it was a noted goal of European pilgrims, especially on Sept. 14 during Engelweihe, the chief festival. The statue owes its name, the "Black Madonna," to the discoloration caused by the candles burned in front of it through the centuries. The existing baroque buildings date from the 18th century and the treasury and library contain many precious objects. In 1274 the reigning abbot was made a prince of the Holy Roman empire. Originally under the protection of the counts of Rapperswil and from 1295 under the Laufenburg Habsburgs, after 1386 it belonged to Schwyz. Ulrich Zwingli was the parish priest of Einsiedeln, 1516–18, while near the town Paracelsus (*c.* 1490–1541), the physician and alchemist, was born and there is a monument to him near the monastery. Calderon's *Das Grosse welttheater* ("Great World Drama"), the play about Life and Death adapted from the Spanish, is performed at periodic intervals in an open-air theatre. (R. Fu.)

EINSTEIN, ALBERT (1879–1955), was one of the greatest figures in physics. Apart from his creation of the special and the general theory of relativity, he made great contributions to statistical mechanics and to the quantum theory, especially the quantum theory of radiation.

Einstein was born on March 14, 1879, at Ulm, Ger. The family moved soon, however, to Munich, where his father started a small factory, and the boy began his schooling there. While the school instruction with its pedantic methods bored him, his uncle on his father's side aroused his interest in mathematics, and this had a lasting influence. After the family moved to Italy, Einstein continued his schooling in Aarau, Switz. In 1896 he entered the Swiss Federal Polytechnic school in Zürich to be trained as a teacher in physics and mathematics. He did not find the instruction very inspiring and occupied himself mostly with the reading of the works of Ludwig Boltzmann, James Clerk Maxwell, Hermann Helmholtz, Heinrich Hertz and Gustav Kirchhoff. In 1900 he received his diploma and acquired Swiss citizenship. (He received a Ph.D. degree in 1905 from the University of Zürich.) He searched in vain for an academic position and finally took a post with the patent office in Berne, where his main duty was the preliminary examination of patent applications. His job left him with ample time to contemplate the fundamental problems in physics. The year 1905 saw the appearance of four of Einstein's important papers. Each contained a great discovery in physics: the creation of the special theory of relativity; the establishment of the mass-energy equivalence; the creation of the theory of Brownian motion; the foundation of the photon theory of light. In a symbolic fashion these papers charted the three main directions in physics to which Einstein contributed the most: the theory of relativity, statistical mechanics and the quantum theory of radiation. (For a discussion of these subjects, *see* below.)

In 1909 Einstein became professor at the University of Zürich; in 1910 he joined the German university in Prague (then part of the Austrian-Hungarian empire), returning in 1912 to Zürich as a professor at the Swiss Federal Polytechnic school. Finally, in 1913 he became a professor at the University of Berlin, the director of the Kaiser Wilhelm institute, and a member of the Prussian Academy of Sciences. During this time Einstein came to the conclusion that the correct generalization of the special theory must also furnish a theory of gravitation. His efforts were finally successful, and after several preliminary versions he published in 1916 his great paper on the general theory of relativity. Beside this he contributed to the problems of the theory of radiation, and to problems in statistical mechanics.

As his fame mounted he became, much against his will, an important public figure. He was appointed to the Intellectual Cooperation organization of the League of Nations (1922), and his interest in social problems, his fame and his lucid way of speaking resulted in his appearing many times in public either to draw attention to social problems or to discuss the recent theories of physics. He traveled widely in Europe, the United States and Asia. His visit to the United States in 1921 was for the purpose of supporting the Zionist movement.

In 1922 he received the 1921 Nobel prize in physics "for the photoelectric law and his work in the domain of theoretical physics." The photoelectric law was contained in his 1905 paper on the photon.

Beginning with the 1920s Einstein's approach to the unsolved problems of physics became increasingly divergent from the main developments in physics. In 1926 Erwin Schrödinger, on one hand (wave mechanics), and Werner Heisenberg, Max Born and Pascual Jordan, on the other hand (matrix mechanics), developed the quantum theory of material systems. (*See* QUANTUM MECHANICS.)

Though the theory was eminently successful, the probabilistic interpretation connected with it made Einstein believe that the theory was of provisional nature. This belief was only strengthened by the fact that the methods of the general theory and that of the new quantum theory were quite different. The general theory succeeded in geometrizing the laws of physics in the sense that the laws of physics (at least as far as the laws of gravitation were concerned) were but geometrical propositions concerning the geometry of space-time. Einstein felt that the proper development of physics should lie in the reduction of the other laws of physics as well to geometrical propositions. For this reason he embarked upon the construction of the so-called unified field theories. The name arose because the aim was to demonstrate that both the gravitational field and the electromagnetic field stem from the geometrical properties of space-time, unifying thereby in a sense the fields. For the rest of his life Einstein's main interest in physics was in that direction. However, he also contributed to the problem of the probabilistic interpretation of quantum theory, by drawing attention critically to the difficulties inherent in it.

During the late 1920s the political situation in Germany deteriorated more and more. When Einstein left Berlin in 1932 for a visit to California he was already aware that Adolf Hitler and the Nazi party would soon be in power. Einstein returned to Europe, though not to Germany, resigned from his positions in Berlin and prepared for his emigration to the United States. During the winter of 1933 he joined the Institute for Advanced Study in Princeton, N.J., and settled there for the rest of his life. He became a U.S. citizen in 1940. At the institute Einstein continued his work on general relativity, the unified field theories and the critical discussion of the interpretation of quantum theory. He also co-operated with charitable and social organizations to help the large number of refugees who were arriving in the United States from Nazi Germany.

In 1939 it became known that two German physicists, Otto Hahn (*q.v.*) and Lise Meitner, had discovered the fission of uranium. Enrico Fermi (*q.v.*), an Italian physicist who at that time had arrived in the United States, became aware of the fact that if the fission could be made into a self-perpetuating chain reaction, enormous quantities of energy could be released. Fermi and the Hungarian physicist Leo Szilard, realizing the important military implications, decided to point this out to the U.S. government. Szilard and Eugene Wigner, another Hungarian physicist, asked Einstein, whom they had known in Berlin, to appeal directly to Pres. Franklin D. Roosevelt, pointing out the dangers if Germany succeeded in developing a bomb based on these principles. Einstein's famous letter to President Roosevelt resulted in the Manhattan project and in the development of the atom bomb. In 1945 Einstein retired from his position at the institute but continued to work there until his death in Princeton on April 18, 1955.

Scientific Contributions.—Einstein, like Isaac Newton, at his death left a physics vastly changed by his own contributions. As pointed out previously his main achievements lay in the theory of relativity, statistical mechanics and the photon theory of light. In addition to these discoveries he left to the world of physics a new viewpoint: the aim of the geometrization of physics.

The theory of special relativity arose from the attempt to reconcile the laws of mechanics with the laws of the electromagnetic field. Einstein realized that this can only be accomplished if we analyze carefully the methods observers would use to correlate their time and space measurements while moving with a constant velocity relative to each other (for details, *see* RELATIVITY). In this way he was able to specify general conditions which theories must satisfy to be deemed laws of nature, and he also showed how mechanics should be modified so that this might be accomplished.

From the purely theoretical side the special theory thus furnishes a powerful guiding principle in our search for the laws of nature. Moreover, it shows that our simple concepts of space and time, upon which the laws of Newtonian mechanics were based, require profound modifications. This entailed that the length of a metre rod and the duration of a second on a clock depend on the state of motion of the observer with respect to the instruments. For an observer a moving metre rod will look shorter and the seconds on a moving clock will look longer. For ordinary velocities this change is completely negligible; however, if we approach speeds which are comparable to the speed of light this becomes of great import. The practical significance of these results was immense. The conclusions of this beautiful theory

not only furnished us with a deeper understanding but also supplied us with part of the laws which describe the behaviour of fast elementary particles now so widely used in physical research (*see* ACCELERATORS, PARTICLE). Moreover, it was one of the results of the special theory that mass and energy are equivalent in the sense that if m units of mass could be made to disappear, mc^2 units of energy would be liberated, c being the speed of light. In this way the masses of nuclei furnish the large amounts of energy supplied by nuclear reactions (*see* NUCLEUS: *Nuclear Reactions*). Nature herself makes use of nuclear reactions, and they provide the main energy source in stars.

Already in 1907 Einstein came to the conclusion that the electromagnetic field must be influenced by gravitational fields. During his research he developed his famous principle of equivalence which later became the cornerstone of the general theory of relativity. This principle states that an accelerated observer will see a physical process taking place the same way as if he were not accelerated, but a gravitational field would be present which would produce the same acceleration as that of the observer. In this way the linking of gravitational fields and accelerated observers took place. This principle had great consequences. The geometrical interpretation of the special theory, as proposed originally by Hermann Minkowski, suggested that a possible development of the special theory would lie in the admission of accelerated observers. However, at the outset it was not certain that this could be accomplished. Einstein realized that this would be successful only if the theory also embraced the phenomenon of gravitation. He slowly developed his general theory of relativity. The intellectual beauty of the theory is very great. In it the laws of physics are the laws of geometry in four dimensions, and these laws in turn are determined by the distribution of matter and energy in the universe (*see* RELATIVITY). On the practical side the theory accounted for the anomalous behaviour in the motion of the planet Mercury, and it predicted two new phenomena—the bending of light rays in gravitational fields and the change of frequency of light in gravitational fields. Both effects were later observed. The general theory of relativity gave a new impetus to cosmology, the theory which deals with the properties of the universe in the large. (*See* COSMOLOGY: *Cosmological Models;* COSMOGONY.) Einstein inaugurated this development with his paper in 1917 demonstrating the possibility of a spatially finite though unbounded universe within the general theory. Later many other possibilities were investigated by others. Einstein proposed several generalizations to his general theory. Each attacked from a new angle the possibility of deriving both the gravitational and the electromagnetic fields from the geometrical properties of space-time. None of these attempts met with complete success, though most of them contain mathematical results of great elegance.

We turn now to Einstein's statistical theories and his work on the nature of light. Roughly speaking, there were two phases in his contributions. In the beginning he dealt with the classical problems of statistical mechanics while later he attacked problems in which quantum theory and statistical mechanics were merged. In general Einstein considered the principles of statistical mechanics as having the most secure foundations in a changing physics. This is the reason that Einstein was able to derive several laws of quantum theory (before quantum theory as a theory had emerged) by investigating what these (tentative) laws must be in order to satisfy the condition of thermal equilibrium. For this reason we will discuss here his contributions to quantum theory as well. Einstein's first papers (1902) dealt with the general foundations of statistical mechanics. He immediately applied them to an important problem. It was known for a long time that small particles, bodies visible through a microscope only, undergo a violent agitation if suspended in liquids. This "Brownian motion" Einstein explained as due to the thermal motion of the molecules in the liquid, which bounce against the suspended particle. Einstein showed that with this assumption he could explain quantitatively the motion of the small particles (1905). This theoretical explanation of the Brownian motion was the first visible proof of the molecular constitution of matter.

At the same time he also applied his interest to the investigation of the nature of light (1905) by investigating the thermal properties of light with a low radiation density. He observed that for low intensities the thermal properties of light are of such a nature as if light were composed of independent quanta of energy. This argument suggested to Einstein that there should be experiments where this discrete nature of light would reveal itself directly. These phenomena, the photoelectric effect, photoionization and photoluminescence, were actually known and the experimental results yielded beautifully to this interpretation. At the same time he demonstrated (1907) that if we assume that under certain conditions the energy of a system can take discrete values only, we can satisfactorily explain the specific heat of solid elements and thus remove one of the great stumbling blocks of 19th-century statistical mechanics. Following this, he showed (1917) that the processes of emission and absorption of light must be governed by new laws if light is to have this new discrete property. These new laws were the first known laws of the quantum theory of *matter* derived before the advent of quantum theory. (Some of the quantum laws of *radiation,* of course, had already been discovered by Max Planck [*q.v.*] in 1900.)

Einstein's last great contribution to statistical mechanics was the development of the quantum theory of a monatomic gas, following the idea of Sir Jagadis Chunder Bose, an Indian physicist (1924–25). These two papers stimulated Schrödinger to develop his wave mechanics which marked the coming of age of quantum theory.

Personality.—Einstein's personality was as impressive as his scientific work. He believed in a world of simplicity and harmony which can only be created if people act according to principles founded on experience and consciously clear thinking. His character as revealed through his actions was of noble simplicity. He was full of benevolence, integrity and humour. Except for his humour his traits were impersonal and wholly of an unsentimental nature. Notwithstanding the aloofness inherent in such an attitude, he had a passionate and active interest in social justice and responsibility.

In science he maintained that though the world can be understood in terms of reason the criteria for the acceptance of a theory are, in the last analysis, aesthetical.

He once summed up his general outlook toward the world by saying: *Raffiniert ist der Herrgott, aber boshaft ist er nicht* ("God is subtle but he is not malicious").

Einstein's publications include *The Meaning of Relativity,* Eng. trans. by E. P. Adams and others, 5th rev. ed. (1956); *About Zionism,* Eng. trans. by Leon Simon (1931); *Builders of the Universe* (1932); with Sigmund Freud, *Why War?,* Eng. trans. by Stuart Gilbert (1933); *The World As I See It,* Eng. trans. by Alan Harris (1949); with Leopold Infeld, *The Evolution of Physics* (1938); *Out of My Later Years* (1950).

See ATOMIC ENERGY; BROWNIAN MOVEMENT; GRAVITATION; PHOTOELECTRICITY; RELATIVITY; SPACE-TIME; UNIFIED FIELD THEORY; *see* also references under "Einstein, Albert" in the Index volume.

BIBLIOGRAPHY.—Philipp Frank, *Einstein, His Life and Times,* Eng. trans. by George Rosen (1947); Paul Arthur Schilpp (ed.), *Albert Einstein, Philosopher-Scientist,* "The Library of Living Philosophers" (1949); Carl Seelig (ed.), *Ideas and Opinions,* Eng. trans. by Sonja Bargmann (1954) and *Albert Einstein: a Documentary Biography* (1956). (N. L. B.)

EINSTEINIUM. The synthetic chemical element einsteinium has the chemical symbol Es and the atomic number 99. It occupies a position in the periodic system of the elements as the tenth member of a rare-earth-like transition series, the actinide series, which includes the heaviest known elements in which an inner electronic shell (the 5f shell) is being filled. The seventh transuranium element to be discovered, it was first identified by Albert Ghiorso and his co-workers. For a discussion of its discovery, *see* TRANSURANIUM ELEMENTS.

Since the isotopic composition of any given sample of einsteinium is dependent on its source there is no unambiguous designation of an atomic weight. Where there is a formal requirement for this quantity the use of the mass number of the longest lived iso-

Isotopes of Einsteinium

Isotope*	Half life	Type† and energy of radiation (Mev)
Es^{248}	7.3 min.	EC; α 7.35
Es^{248}	25 min.	EC; α (0.3%) 6.87
Es^{249}	2 hr.	α 6.76
Es^{250}	8 hr.	α EC
Es^{251}	1.5 days	α 6.48 EC>99%
Es^{252}	~140 days	α 6.64
Es^{253}	20 days	α 6.633 (91%) 6.592 (8%) 6.543 (1%) 6.491 (0.3%)
Es^{254m}	36 hr.	β^- 1.1 EC (0.1%)
Es^{254}	480 days	α 6.44
Es^{255}	40 days	β^-
Es^{256}	short (hour or less)	β^-

*The symbol m after the mass number refers to an isomeric form of the isotope.
†EC=electron capture; α=alpha particle; β^-=negative beta particle.

tope might be considered. Because of the existence of the long-lived Es^{254} isomer (480 days half life) weighable amounts of einsteinium are available for detailed physical and chemical studies. The isotope Es^{253} and heavier einsteinium isotopes can be produced by intense neutron irradiation of lower elements, such as plutonium, by a process of successive neutron capture reactions interspersed with beta particle decays until this atomic number and these mass numbers are reached.

By the early 1960s the chemical properties of einsteinium had been studied solely with tracer amounts; it seemed evident that the III oxidation state may be the only one that exists in aqueous solution. Einsteinium, being an actinide element, has chemical properties very similar to those of the other actinide elements in the tripositive state. The major emphasis in einsteinium chemistry up to the early 1960s was on ion-exchange separations of einsteinium from the rare earth and the other actinide elements. Adsorption and elution from columns packed with ion exchange resins are used to separate einsteinium from the other tripositive actinide elements. It is possible to separate einsteinium from the rare earth elements by the use of eluting agents such as concentrated hydrochloric acid, which forms complex ions with the actinide elements to a much greater extent than with the rare earth elements.

See J. J. Katz and G. T. Seaborg, *The Chemistry of the Actinide Elements* (1957); G. T. Seaborg, *The Transuranium Elements* (1958).
(G. T. Sg.)

EINTHOVEN, WILLEM (1860–1927), Dutch physiologist and 1924 Nobel prize winner in medicine and physiology, was born on May 21, 1860, at Samarang, in Java, and was educated at Utrecht State university, Neth., under F. C. Donders (*q.v.*). In 1886 he was made professor of physiology at Leiden university, where he remained until his death on Sept. 29, 1927.

Einthoven's primary interests and greatest contributions were related to electrical phenomena arising in the heart, and in 1903 he devised a type of string galvanometer satisfactory for the registration of very small, rapidly changing voltages. This instrument made it possible, for the first time, to record electrocardiograms simply and accurately, and it was widely used for this purpose for over 20 years. In addition to this great contribution, Einthoven pioneered in development of electrode arrangements (lead systems) for the registration of electrocardiograms in man, and the standard limb leads, almost universally recorded today, were originally described and used by Einthoven. Although his contributions to physiology and medicine were great, he was a very humble and modest man; while traveling in the United States in 1924 he was notified that he had been awarded the Nobel prize in medicine, but he tried to avoid a public announcement of that great honour.
(F. D. J.)

EIRE, a Gaelic name for Ireland, the disyllabic form of the older Gaelic *Ériu* (dative *Érinn*). The Norsemen called the island the land of Éire; *i.e.*, Ireland. The constitution of 1937 abolished the title Irish Free State and restored the name Eire. In 1949 the title was changed again to Republic of Ireland. *See* Ireland, Republic of.

EISENACH, a town of Germany, which after partition of the nation following World War II was included in the *Bezirk* of Erfurt, German Democratic Republic. It is situated on the northwestern slopes of the Thuringian forest at the confluence of the Hörsel and the Nesse, 30 mi. W. of Erfurt city. Pop. (1959 est.) 48,602. Its interesting features include the Romanesque church of St. Nicholas (restored 1887); the Romanesque St. Nicholas gate; the late Gothic St. George's church, several times renovated; the princes' town palace of 1742; the Lutherhaus of the Cotta family where Martin Luther stayed as a schoolboy; the Bach museum in memory of Johann Sebastian Bach who was born at Eisenach; the house where Fritz Reuter lived from 1868 to 1874, now a museum commemorating him and Richard Wagner; the Thuringian museum housed in the palace and in the Dominican church; the Kartaus garden and the municipal park. There are medical and church-music schools. Eisenach is on the railway from Erfurt and Gotha to Frankfurt am Main. It has a thriving tourist industry and the district is rich in natural beauty. On a precipitous hill 600 ft. above the town to the south is the Wartburg (*q.v.*), an ancient castle of the landgraves of Thuringia, famous as the scene of the contest of minnesingers, immortalized in Wagner's *Tannhäuser,* and as the place where Luther, on his return from the diet of Worms in 1521, was kept in hiding and completed his translation of the Bible. Industries include the manufacture of motor vehicles (the Wartburg cars), metal and woodworking, worsted spinning, chemical and electrotechnical works.

Eisenach (Isenache) was founded in the 12th century by the landgraves of Thuringia. The principality fell to the Saxon house of Wettin in 1264. It was a separate Saxon duchy from 1596 to 1741, when it fell to Saxe-Weimar. In 1817 the festival of the Burschenschaft (national political student movement) took place there; in 1859 the Deutscher Nationalverein (German national society) was founded there; and ten years later the Sozialdemokratische Arbeiterpartei (the Social Democratic Workers' party) was established at the congress of Eisenach. The town developed as an industrial centre in the second half of the 19th century and was made the seat of an evangelical protestant bishop in 1921.
(Eb. M.)

EISENHOWER, DWIGHT DAVID (1890–), U.S. army officer, leader of the Allied forces in Europe in World War II and 34th president of the United States, was born on Oct. 14, 1890, in Denison, Tex., the third of seven sons. His forebears had moved west to Kansas as part of a colony of River Brethren, a Mennonite sect. His father was in Texas as a railroad worker in 1890 but the following year returned to Abilene, Kan., to work in a creamery. The family was poor and the young Eisenhowers worked hard. The home saw much Bible-reading, for the parents held firmly to their simple, pietistic faith. A strong, agile, athletic boy, "Ike," as he was early nicknamed, played football and baseball, and maintained an above-average scholastic record. In 1911 he entered the U.S. Military academy at West Point, where he was very popular. From early in life his warm, outgoing, gregarious personality and persistently optimistic nature readily attracted people to him. At graduation, he was academically in the top third of his class.

His life then followed for 25 years the path of a moderately successful but publicly unknown army officer. In 1916 he married Mamie Geneva Doud, daughter of a successful businessman. Two sons were born to them: Doud Dwight in 1917, who died in 1921 of scarlet fever; and John Sheldon Doud, born in 1922.

During World War I, Eisenhower studied at a pioneer tank school and commanded a tank training centre. After the war he was promoted to the permanent rank of major. He had two ambitions: to become a colonel before he retired, and to have the command of troops. These goals were not to be reached until World War II. He served in the Panama Canal Zone from 1922 to 1924, and then achieved a distinction which was to make him a marked man: he graduated first in a class of 275 from the army's command and general staff school in 1926. In 1928 he graduated from the Army War college, then spent a year in France preparing a revision of a guidebook to American battlefields. He was thereafter stationed in Washington, D.C., where he helped to devise

an industrial mobilization plan for use in case of war, played a part in founding the Army Industrial college and served in the office of the chief of staff, Gen. Douglas MacArthur, from 1933 to 1935. In Sept. 1935 he sailed for the Philippine Islands to serve under General MacArthur, military adviser to the commonwealth.

World War II.—When Germany invaded Poland in 1939, and the European phase of World War II began, Eisenhower returned to the United States and rose swiftly to the highest military levels. Brilliant work in the Louisiana war maneuvers in the summer of 1941, where he was chief of staff for the 3rd army, brought him to the notice of Gen. George C. Marshall, U.S. army chief of staff. He was soon promoted to the rank of brigadier general. After the Japanese attack on Pearl Harbor in Dec. 1941, he was summoned to Washington to serve as assistant chief of the war plans division of the army general staff. He helped prepare plans for the invasion of Europe, discussed them with the British in London in May and April 1942, and on June 25, 1942, reappeared in London as commander of the European theatre of operations. In this role his genius for bringing men together, for producing agreement out of conflict, came to full flower. He was soon popular and almost immediately became a world figure. His remarkable combination of qualities—sincerity, genuine friendliness, humanity, a serious dedication to his task, a lack of pageantry, a rugged masculinity tempered by humility—attracted deep devotion, even adulation. As early as the successful invasion of north Africa, which he commanded in late 1942, friends began talking to him of a political career, only to be brushed aside with irritation and mounting puzzlement by Eisenhower.

A summer campaign in 1943 took the island of Sicily and led to the invasion of Italy in September and the signing of an armistice with the defeated Italians. In late Nov. 1943 Eisenhower returned to London to command the enormous force building up to invade France and led it across the channel on June 6, 1944. There followed the hard fighting to hew out a stable bridgehead, then the breakout and the wide-sweeping rush through France. In Dec. 1944 the Germans launched a critically important counterattack through the lightly held front before the Ardennes forest. Fighting down this last-ditch effort to break the western armies, Eisenhower took up the offensive again in mid-Feb. 1945, not ceasing until all Germany was occupied by British and U.S. forces in the west and by Russian forces in the east. He had meanwhile, in Dec. 1944, been promoted to the five-star rank of general of the army.

Postwar Years.—He briefly commanded U.S. occupation forces in Germany and then became chief of staff of the U.S. army. In this post he worked to unify all the armed services. In 1948 he retired from military service and became president of Columbia university. Many persons in both major political parties spoke of Eisenhower as a possible candidate for president, but he stated publicly in Jan. 1948 that he did not wish to be considered as a candidate. In the autumn of 1948 he published his war memoir, *Crusade in Europe,* which was soon translated into many languages. Its sale made him almost overnight a rich man. Reportedly at this time he formed close friendships with a number of men in New York legal and financial circles. Indeed, from the beginning of his postwar public career he found his best friends among men of great wealth. When he became president he took millionaires rather than military men into his cabinet with him.

Eisenhower's career as president of Columbia university was not a great success. Observers remarked that his background was hardly that appropriate to a great academic post. Furthermore, the warmth with which until then many liberals had regarded him began to cool, for in public speeches it became increasingly apparent that, despite his internationalism and his luminous personal qualities, his ideas were firmly conservative on domestic issues. He distrusted intellectuals, looked with alarm upon the growth of a great government, spoke of the dangers of socialism and lauded private initiative.

In Dec. 1950 Pres. Harry S. Truman appointed him commander of the newly established supreme headquarters for the North Atlantic Treaty organization (NATO), with the task of creating an effective military defense against communist aggression in western Europe. He proceeded once again to demonstrate his unrivaled capacity to create a united military organization of representatives drawn from many nations.

Election as President.—Meanwhile, in the United States, a struggle was going on within the Republican party between those of the eastern wing, who were internationalist and moderate in domestic affairs, and those in the middle west, who wished to see the United States reduce its foreign commitments and return as much as possible to the kind of domestic arrangements that had existed before the advent of Pres. Franklin D. Roosevelt. The figure to whom members of the latter group were loyal was Sen. Robert A. Taft of Ohio, and they worked vigorously to secure his nomination for the presidency. To eastern Republicans the nomination of General Eisenhower became progressively more desirable. He, meanwhile, had become convinced that the Republican party should again be given direction of the nation's affairs. Moved by a deep sense of duty, he responded to eager pleas that he run for the presidency by declaring his willingness to accept "a clear-cut call," although he would not campaign for the nomination. After a bitter convention fight he was nominated for president and Richard M. Nixon was nominated for vice-president. In an election campaign keyed to the "crusade" theme he overwhelmingly defeated the Democratic nominee, Gov. Adlai E. Stevenson of Illinois. Great numbers of Democratic and independent voters cast their ballots for him, including enough in the south to swing many states in that region to his support. The Republicans also won control of both houses of congress, though the majorities were narrow.

Conception of the Office.—When he was inaugurated on Jan. 20, 1953, Eisenhower brought to the White House ideas from an earlier era. Educated in the traditional understandings of American government and society, he believed deeply in the separation of powers, and abruptly terminated the practice of executive leadership in the federal government. He felt that his duty as president was to put forth his proposals as logically and persuasively as possible, and leave the disposing to congress. He also applied this concept to partisan politics, insisting that he should be "President of all the people," and "not make of the Presidency an agency to use in partisan elections." He sought to reintroduce into national policy a friendliness toward free enterprise and private development of the nation's resources that had been lacking for many years. He struggled to keep the government's spending within its income, and warned often of the dangers of inflation. Appointing like-minded men to his cabinet, he gave them almost complete freedom to run their departments as they thought proper. He delegated authority to officials subordinate to him, preferring to preside over rather than actively direct the government.

But Eisenhower's deep conviction that the United States must play a strong and constant role in world affairs pushed him steadily in the direction of executive leadership. Maintaining and extending alliances, working with the United Nations, and urging economic and military aid to friendly nations resulted in a constant struggle with congress to secure the necessary funds and a steady use of executive power and prestige in foreign affairs. To get what he wanted, Eisenhower had to learn and use the techniques of party government. Furthermore, his party clamoured for him to join the political campaigning, a plea to which he reluctantly assented. In his second administration the men who came to power with him departed one by one, either by resignation or by death, forcing him to take back the reins of power which he had given to them.

An outstanding aspect of Eisenhower's administration was the fact that he was much more popular than his party, which held control of congress for only two years, 1953–55. Eisenhower won another victory over Stevenson in the presidential election of 1956 but the Democrats kept control of congress. For the first time in 108 years a president had not carried his party along with him to control at least one of the houses. In 1958 the Democrats converted their slim 1956 majorities into huge, almost unwieldy ones. For most of his tenure, therefore, Eisenhower had to work with a congress in the hands of the opposition party. Curiously enough, this was not altogether a handicap.

Presidential Terms.—Eisenhower's great task as president was threefold: he had to quiet a country disturbed by the fears and open hatreds set off by the cold war and the Korean War; keep peace in the world while resisting further communist expansion; and demonstrate to the nation and to the pent-up forces within his party that Republican rule would not mean dismantling an internationalist foreign policy and a system of government devoted to social welfare. To a considerable extent, he was successful. American national life regained a welcome degree of unity and stability. By refusing to feud with political antagonists either in or out of his party, by providing a rallying point to which millions looked, by surrounding the presidency with dignified decorum and by creating an orderly style of governing, he helped establish an atmosphere in which over-heated tempers could subside. He was not, however, wholly successful until the right wing of his party had quieted.

Eisenhower's conception of the presidency greatly limited his capacity to exert control over obstreperous Republicans. In his first year he demonstrated that he had no intention of scuttling the foreign policy his predecessors had built up, and that his frequently asserted concern for the "little people" of America meant continued maintenance and extension of social and economic reform. The result was two years of constant difficulty. The Republican majority in congress, so long out of power that it had practically lost the capacity for anything but obstructive opposition, lashed out in many directions. By the end of summer 1953 he was in despair. He gave serious thought to forming a third party that would better embody the kind of "dynamic conservatism"—a term he could never clearly explain—he believed in. Sen. Joseph R. McCarthy of Wisconsin mounted a steadily growing campaign of attack against him. Eisenhower's reaction was largely limited to public indifference and private fury, for he believed that the good sense of the American public would finally take away McCarthy's support. Critics could find no justification for Eisenhower's persistent refusal to use the great power and prestige of his office to halt this political vendetta that did much harm to the United States both at home and abroad.

Chief Advisers.—Since Eisenhower knew relatively little of domestic affairs, the men he put in charge of the government were all-important. In particular, Secretary of the Treasury George M. Humphrey became the giant of the administration, for it was he, more than anyone else, who established and maintained, through the force of an extraordinarily strong personality, the basic economic policies of the government: tight money controls and credit manipulation to fight inflation. These policies, joined with a fervent dedication to a balanced budget, comprised the economic program to which Eisenhower gave ever stronger support as his presidency proceeded. In good times money and credit were jealously guarded in the hope of keeping dollar values stable. In the two recessions of his administration—in 1954 and 1957–58—the government largely relied upon an easing of credit to bring about recovery. Eisenhower's dedication to a balanced budget was never as passionate as the Republican right wing would have liked, for he refused to sacrifice foreign aid to achieve it, but throughout his administration critics hotly insisted that he was so determined to balance the budget that he slighted essential national needs, such as aid to education, guided missiles and research. In May 1957 Secretary Humphrey resigned and was replaced by Robert B. Anderson, but basic economic policies remained unchanged.

Sherman Adams was almost as important as Humphrey. Head of the staff system that Eisenhower established in the White House, all functions and responsibilities flowed up to him in a fixed order and sequence. Eisenhower would rarely sign any major document pertaining to domestic and political affairs unless Adams had already approved it. Adams brought department heads together, forced them to make decisions or made them himself and rigorously shielded Eisenhower from matters he thought not sufficiently important to take the president from his primary concern; *i.e.,* world affairs. Efficient, hard-driving, rigorously outspoken on ethics in government, he exercised a wide authority. His sudden fall from grace because of accusations that he had used undue influence to help a Boston industrialist came as a shattering blow to Eisenhower, who accepted his resignation in Sept. 1958 with the greatest reluctance.

Within a few months Eisenhower was to lose the third and last of the men who had been the great figures of the early and middle Eisenhower years. John Foster Dulles, secretary of state, had almost no predecessor in the degree to which he was responsible for the conduct of U.S. foreign relations. Eisenhower's views on foreign policy were strong and definite, but so closely did he and Dulles agree and so warm and trusting became their relationship that by 1955 the much more active and verbally skilful Dulles had taken over almost all authority for the formulation and execution of policy. Convinced that communism would ultimately fail, and that therefore no compromises in any quarter should be made, Dulles was unswervingly dedicated to the concept of containment. He spoke cryptically of massive retaliation and liberation of the satellite countries, and moved about swiftly from nation to nation to engage in the personal diplomacy he knew best. Like Humphrey, he was rocklike in his convictions and was widely criticized as too rigid and inflexible. Struck down by cancer early in 1959, Dulles' death led Eisenhower to take over personal direction of foreign policy, just as Adams' fall had required him to do the same in domestic policy.

Eisenhower's almost constant struggle with ill-health was the backdrop to all of this. On Sept. 24, 1955, he was struck down by a heart attack that immobilized him for weeks and led him to warn the American people frankly that he would have to have a "regime of ordered work activity, interspersed with regular amounts of exercise, recreation and rest. . . ." He thenceforth spent much less time at his desk and much more time in extended vacations, golfing, enjoying the bucolic pleasures of his farm at Gettysburg, Pa., and visiting the large estates of his friends. In June 1956 he had major abdominal surgery for ileitis. Mild respiratory ailments troubled him persistently, and in Nov. 1957 he suffered a mild stroke that briefly impaired his speech. In 1958 and 1959, however, he emerged once again into excellent health and his natural vigour began to manifest itself.

Atoms for Peace.—More than anything else, Eisenhower was devoted passionately to the maintenance of peace in the world. He had the extraordinary good fortune of beginning his presidency just before the end of the Stalin regime in the U.S.S.R. and reaped the many benefits that came from the "thaw" in the cold war that set in during the next three years in world affairs. He was able to bring about peace in Korea in July 1953. In Dec. 1953 he proposed in a speech to the United Nations general assembly that the United States and the Soviet Union jointly contribute nuclear fission materials to an international agency to develop the potential of the atom for peaceful use. The Russians were hostile and many American conservatives opposed the idea, but the International Atomic Energy agency was eventually created with more than 80 nations having access to a pool of atomic materials, largely contributed by the United States.

Disarmament.—In July 1953 Eisenhower caused a world-wide sensation by proposing a bold new disarmament plan. This speech eventually led to the Geneva summit conference of July 1955, where the U.S. president met directly with Soviet leaders and proposed that each country allow continuous aerial surveillance of its military activities. This "open-skies" plan caught the imagination of the world, but by the spring of 1956 Nikita Khrushchev had assumed firm control of the Soviet government and the "thaw" was over. Thenceforth, Eisenhower and Dulles had continually to deal with Soviet thrusts outward from the vast Russian perimeter.

Alliances.—The result was the completion of a ring of alliances around the Soviet Union. The situation in Indochina had deteriorated rapidly in 1954, and to prevent further collapse after that country was cut in two, the South East Asia Treaty organization (SEATO) was created in Sept. 1954. Its members were Britain, France, Australia, New Zealand, the Philippines, Pakistan, Thailand and the United States. In Dec. 1954 the United States signed a defensive treaty with the Nationalist Chinese government and pledged itself to defend Formosa and the Pescadores. This was followed by the Formosa resolution, approved by congress in

Jan. 1955, which authorized the president to fight if need be to protect Formosa. But in the fall of 1958, Eisenhower stated that he doubted that the Nationalists would ever return to the mainland.

When the middle east collapsed into war and revolution under the pressure of Soviet activity the Eisenhower Doctrine was promulgated in March 1957. With congressional approval, the president proclaimed that he would use U.S. armed forces against any communist or communist-dominated aggressor in the middle east if the attacked country sought American help. In mid-July 1958 Lebanon appealed for help, and U.S. troops landed to halt further attacks against the Lebanese government. American negotiators helped bring the warring parties together, and U.S. troops were withdrawn in late 1958. In the midst of the crisis, the United States virtually pledged itself to membership in the Baghdad pact, an alliance of Great Britain, Iran, Pakistan, Turkey and Iraq. This alignment was later renamed the Central Treaty organization when Iraq withdrew.

Sputnik.—In Oct. 1957, when the U.S.S.R. launched sputnik into its globe-circling orbit, Eisenhower's popular prestige plummeted. He had been elected primarily on his military record, and now in his supreme area of competence he was accused of failure. Opponents reminded the public that from the beginning there had been much criticism of his military policy. He had early decided that military spending had to be lowered to a level he felt the economy could bear over a long period of years without damage, rather than building to meet major aggression by a forecasted date, as was formerly the policy. Large and small nuclear weapons became the mainstay of U.S. military power; the size of the army was much reduced; and troops were withdrawn from many overseas locations. Emphasis was placed on air power to deliver the atomic punch, on continental air defense, mobile reserves and massive retaliatory power against major attack. Much less money was spent on guided missile research. The Soviet Union's dramatic achievement, therefore, began a year of national crisis in the United States. Khrushchev exultantly mounted a major diplomatic offensive. In response, Eisenhower urged a speed-up in defense spending, especially for missiles and nuclear submarines. He vigorously pushed through a centralization of the defense department. On Jan. 31, 1958, his situation was greatly eased by the U.S. army's success in placing a satellite in orbit. There followed months in which the U.S. slowly narrowed the strategic military gap, culminating in the firing of an air force Atlas intercontinental ballistic missile in Nov. 1958.

Segregation Crisis.—Meanwhile, Eisenhower became directly involved in the dispute over racial segregation in the schools. His most important single appointment had been that of Earl Warren as chief justice of the United States in Sept. 1953. Warren succeeded in bringing unity to a long-divided supreme court. On May 17, 1954, the court handed down its historic decision unanimously ruling that racial segregation in public schools was unconstitutional. In Sept. 1957 Eisenhower took control of the Arkansas national guard away from Gov. Orval Faubus, thus ending its use to prevent the enrollment of Negro students in a white high school in Little Rock, Ark. Furthermore, the president used federal troops to insure compliance with a court order directing integration in Little Rock high school. Deeply convinced, however, that state and local authorities should handle the matter, and devoted to the ideal of gradualism, Eisenhower took no steps beyond insuring compliance with specific court orders and stating publicly in the strongest terms, as he did on Sept. 26, 1958, that racial segregation was contrary to American ideals of equality. Critics insisted that the problem showed most dramatically the weakness of the Eisenhower concept of the presidency and declared that only executive leadership could bring about the concrete changes necessary.

The German Problem.—In Europe, the most persistent problem was the status of Germany. Largely through Dulles' insistence, the German Federal Republic (West Germany) was accorded independence in May 1955, given the right to rearm and brought into NATO, steps which greatly complicated U.S. relations with France. In late Nov. 1958 the Soviet Union suddenly reopened the long quiescent question of West Berlin by calling for the declaration of Berlin as a "free city," the removal from it of western occupation forces and the placing of its contacts with the west under sufferance of the German Democratic Republic (East German) government. Coming at the very time when Dulles fell fatally ill, the Berlin question severely tested Eisenhower's capacity to rise swiftly and decisively to a major crisis. He did so with great success, and by the summer of 1959 was enjoying not only a restoration of his immense popularity with the American people —61% were reportedly in warm support of him—but an unprecedented popularity with the American press.

Evaluation.—Most of the Washington press corps had long found Eisenhower an inept politician, the captive of his advisers and of a coterie of rich men, and indifferent to executive leadership at a time when leadership was needed desperately. He was, they felt, vague in thought, inexpert and without new ideas. Much was made of his reading habits, though those close to him protested that he read general nonfiction literature as well as western stories.

There is little doubt that Eisenhower was in no way an intellectual by nature. Toward religion, for example, he was unreflective, though broadly committed. He brought a strong sense of dedication to his office, introduced a prayer written by himself into his inaugural addresses and began the practice of devoting a minute to silent prayer at the beginning of cabinet meetings. He joined and regularly attended the Presbyterian Church. But the fact that Eisenhower had never been a churchgoing man before he became president laid him open to the charge that he was merely doing what was proper rather than following the urgings of inner thought. Political and economic ideas played an important role in his thinking, for he was deeply committed to such broad concepts as democracy, equality and individual enterprise. But his relationship to these ideas was receptive, not creative. He was not given to contemplation and the working out of new concepts. His talents lay in organization and in active leadership, not in thought.

The "New" Eisenhower.—The "new" Eisenhower who emerged at the beginning of 1959 and delighted the national press was actually an expression of his true basic nature. The one idea he brought to the presidency that did not match his character was that of the relatively passive president. Finally restored to the ruddy vigour of his first two years in the White House, bereft of the great figures who had run his government and convinced at long last by experience that his conception of the presidency had been wrong, he began using the vast powers of his office. He spoke regretfully of the measures he had let die in congress in previous years through a distaste for politics and showed a new zeal for the art of political maneuver that astonished seasoned observers. He held regular press conferences and dinners for reporters, cajoled and threatened congressmen to secure favourable votes, began to speak almost wistfully about the advantages of the British parliamentary system where the head of government has a direct control over the legislative body and reacted angrily at congressional attempts to reduce the powers of the presidency. He succeeded in dominating the heavily Democratic congress, using the veto power to prevent the heavy spending which he repeatedly stated was the greatest danger to the economy. He pushed through a stiff labour relations bill and obtained authority to lift the interest paid upon government bonds so as to ease the government's task of paying its bills.

Out of this background came a bold new approach to foreign relations, one that would have been disapproved of by Secretary Dulles. Premier Khrushchev was invited to the United States for a tour and for face-to-face talks with Eisenhower in Sept. 1959. An unprecedented era of personal presidential diplomacy began. A period of *detente* arrived, and the Eisenhower presidency grew steadily in stature.

In the month of Dec. 1959 President Eisenhower embarked on a much-heralded 11-nation goodwill tour designed to acquaint the world with the American ideal of peace and friendship in freedom. In Italy, Turkey, Pakistan, Afghanistan, India, Iran, Greece, Tunisia, France, Spain and Morocco enormous crowds demonstrated extraordinary enthusiasm and friendship for him. En route home, the president joined in a long-awaited invitation to Premier

Khrushchev to attend a conference in Paris on May 16, 1960. In February and March 1960 Eisenhower undertook a goodwill tour to Brazil, Argentina, Uruguay and Chile, and was again met by tumultuous crowds. He proclaimed a "hemisphere crusade" for economic development and reaffirmed American aims of hemispheric political and economic stability. World-wide hopes were rising that the remarkable personal efforts of the president would be capped by peacemaking accords with the Soviet Union.

Against this background, the shocking events of early May 1960 were a personal tragedy for Dwight Eisenhower. The president whose one dream had been to secure world peace before he left office suffered grievously when the downing of an American U-2 reconnaissance plane in the U.S.S.R. on May 1 became either the cause of, or the pretext for, the breakup of the summit conference in a spectacular tirade by Khrushchev, and the reopening of the cold war. Much of the free world, rightly or wrongly, blamed Eisenhower for the timing of the flight. The proposed Eisenhower trip to the U.S.S.R. was canceled by Khrushchev. Meanwhile, Cuba, Laos and Congo became theatres for cold war hostilities. The president's last goodwill tour, this time to the far east, was blighted when the Japanese government, wracked by tumultuous anti-American riots, asked him to cancel his visit.

The last six months of his term were obscured by the 1960 presidential campaign, by extensive partisan discussion concentrated upon hints of a growing economic recession and by the legacy of frustrated hopes engendered by the collapse of the summit conference. Nevertheless, Eisenhower continued his moderately activist presidential role. In the political campaign he gave his support to Vice-President Nixon. On July 12 he announced a co-operative aid program for Latin America involving expenditure of at least $500,000,000. He provided air transport of troops from Ghana to the Congo when the UN began its tortuous efforts to pacify that country. He visited UN headquarters twice during the stormy September meetings and spoke to the general assembly. He launched a vigorous campaign to halt the flow of gold from the United States and on Jan. 3, 1961, broke diplomatic relations with Cuba. Now the oldest president in the nation's history, 17 days later he applauded warmly as John F. Kennedy, the youngest man to be elected president, delivered his inaugural address. In 1963 he published a volume of reminiscences entitled *Mandate for Change*. *See* also references under "Eisenhower, Dwight David" in the Index.

BIBLIOGRAPHY.—R. J. Donovan, *Eisenhower: the Inside Story* (1956); M. W. Childs, *Eisenhower: Captive Hero* (1958); Kenneth S. Davis, *Soldier of Democracy*, rev. ed. (1952); R. H. Rovere, *The Eisenhower Years: Affairs of State* (1956); Forrest Pogue, *The Supreme Command* (1954). (R. L. K.)

EISENHOWER DOCTRINE, a joint resolution of the United States senate and house of representatives authorizing the president to employ the armed forces of the United States to protect the independence of any nation in the middle east requesting aid against aggression by any nation controlled by international communism; to give military assistance to any nation in the area desiring such aid; and to help any middle eastern nation to maintain its independence by developing its economic strength. The resolution was passed by the U.S. congress on March 7, 1957, in response to an address to congress by Pres. Dwight D. Eisenhower on Jan. 5, 1957.

The doctrine was intended to check the increase of Soviet influence in the middle east as a result of the supply of arms to Egypt by communist countries and of the strong support by such countries to the Arab cause against the Israeli, French and British attack on Egypt in the fall of 1956. Militarily it extended the security frontiers of the United States to the middle east, thus bridging the gap between the North Atlantic Treaty organization and the Southeast Asia Treaty organization. Coming after the Truman doctrine, the Eisenhower doctrine had to be viewed not as an isolated policy but as part of the United States policy of strengthening the noncommunist world against Soviet advances. This policy was exemplified by the landing of marines in Lebanon by the U.S. government in July 1958 upon the request of the Lebanese government.

See J. Campbell, *Defense of the Middle East* (1958). (E. A. SM.)

EISENHÜTTENSTADT, a city of Germany, located in the Frankfurt *Bezirk* (district) of the German Democratic Republic. In 1961 it became capital of a *Kreis* (county) in the Frankfurt an der Oder *Bezirk*. It was formed in 1961 by the merging of the towns of Fürstenberg and Stalinstadt. Fürstenberg, 15 mi. S.S.E. of Frankfurt, lies on the left bank of the Oder with a port (the eastern terminus of the Oder-Spree canal); industrial undertakings including the manufacture of glass, briquettes and coal-tar dyes, and an important frontier trade with Poland. In 1951 the Ost (east) metallurgical complex was founded in a former wooded area about 3 mi. W. of Fürstenberg. Several blast furnaces were erected and by the early 1960s pig-iron production was about 1,250,000 tons annually. The barracks of the metallurgical workers, originally named Stalinstadt, gave place to a modern city with tall buildings, a theatre, hospital and schools. Pop. (1961 est.) 32,970. Eisenhüttenstadt, one of the chief industrial cities of the republic, lies on the railway between Frankfurt and Wilhelm Pieck Stadt (Guben).

EISENSTEIN, SERGEI MIKHAILOVICH (1898–1948), Soviet film director and one of the most influential film technicians, was the son of a prosperous shipbuilder. He was trained as a civil engineer and architect and during the Revolution built trenches for the Bolsheviks, later joining the first workers' theatre, the Proletcult, as art director. Dissatisfied with the theatre's limits after staging Tretiakov's *Gasmasks* in an actual factory setting, he turned to film, beginning with *The Strike*.

At the age of 26 he directed *Potemkin*, a story of mutiny on a battleship in 1905, which had a world-wide influence on makers of realistic films. *October* (also titled *Ten Days That Shook the World*, 1927–28) and *General Line* (*The Old and the New*, 1926–29) increased his reputation. Paramount–Famous–Lasky corporation brought him to Hollywood but the scenarios he wrote for *Sutter's Gold* and *An American Tragedy* were rejected. Sponsored by Upton Sinclair, Eisenstein spent 14 months in Mexico, but he was recalled by his government before the film on which he was working was completed; some of the footage, edited in Hollywood, was released as *Thunder Over Mexico*. Under mounting criticism in Moscow, he had to leave three other sound films unfinished, but *Alexander Nevsky* (1938) won him the order of Lenin. The second part of the projected trilogy *Ivan the Terrible* (1946) was banned. He died in Moscow, Feb. 10, 1948.

Eisenstein's early use of nonprofessionals as actors ("typage") and his crosscutting of images to represent complex ideas (montage) were elaborated theoretically in lectures at the State Institute of Cinematography and in his books *The Film Sense* and *Film Form*.

See Marie Seton, *Sergei M. Eisenstein* (1952). (R. D. MACC.)

EISLEBEN, a town of Germany, which after partition of the nation following World War II was located in Halle *Bezirk* (district) of the German Democratic Republic. It is situated in the southeastern spurs of the Harz mountains 39 km. (24 mi.) W.N.W. of Halle. Pop. (1961 est.) 34,270. The town is divided into old and new towns (Altstadt and Neustadt). Eisleben was the birthplace of Martin Luther. The late Gothic Peter-Paul-Kirche contains the font in which Luther was baptized. The house in which he was born (1483) was burned in 1689 but rebuilt in 1693 as a free school for orphans. The house in which he died (1546) was restored toward the end of the 19th century, and his death chamber is still preserved. Eisleben is connected by railway to Halle, Magdeburg and Erfurt. It is the chief town of the Mansfeld region and the most important centre of copper ore mining in the republic; this industry dates from the 12th century. The town has a mining engineering school and a training institute for mining. Textiles, machinery, furniture, cigars and clothing are also produced and the agriculture of the district is important. The earliest record of the town (Lat. Islebia) is dated 974. In 1045 it belonged to the counts of Mansfeld. In the 18th century Eisleben fell to Saxony but by the congress of Vienna (1815) was assigned to Prussia.

EISNER, KURT (1867–1919), German socialist, writer and statesman, was born in Berlin on May 14, 1867, the son of a Jewish family. From 1886 he studied philosophy and German

literature at Hermann Cohen's Neo-Kantian school at Marburg and in 1892 published *Friedrich Nietzsche und die Apostel der Zukunft*. From 1892 to 1893 Eisner worked on the *Frankfurter Zeitung* and then moved again to Marburg writing for several Berlin journals; he was imprisoned for some months during 1897–98 for a breach of the press laws. From 1898 he was editor of *Vorwärts*, the official Social Democratic paper. Though himself not a revisionist, he resigned in protest when, in 1905, two colleagues were dismissed for their revisionist views. He became cultural editor and theatre critic for the Social Democratic paper in Munich and from 1907 to 1910 edited the one in Nürnberg. Eisner became a Bavarian citizen and from 1910 worked as a free lance writer in Munich. On July 28, 1914, hoping to prevent Germany's coming to the aid of Austria-Hungary against Serbia, he circulated information about the imminent intervention by Russia. In 1917, influenced by pacifist principles, he joined the Independent Social Democratic party (USPD). In Jan. 1918 he was imprisoned in Munich as a strike leader but was later released, after protests by the socialist members of the *Reichstag*. On the night of Nov. 7–8, 1918, he proclaimed the Bavarian republic. He became its first prime minister and minister for foreign affairs and made his views known by contributing to *Die Neue Zeitung*. He was murdered in Munich by the student A. von Arco-Valley on Feb. 21, 1919. A collected edition of Eisner's works was published in two volumes in 1919.

BIBLIOGRAPHY.—F. Fechenbach, *Der Revolutionär Kurt Eisner* (1929); E. Niekisch, *Gewagtes Leben* (1958); W. Herzog, *Menschen, denen ich begegnete* (1959); F. Schade, *Kurt Eisner und die bayerische Sozialdemokratie* (1961). (K. Ko.)

EISTEDDFOD, a Welsh word for "session" or "assembly," was also used, as early as the 15th century, for sessions of bards and minstrels that were in some sense the forerunners of the modern eisteddfod, both local and national, with its great variety of competitions, its concerts and exhibitions.

References in the works of the court poets of the period 1100–1282 suggest that bardic competitions were usual in Wales, and the laws traditionally associated with the name of the 10th century prince Hywel give some information about poetic instruction, even of the payments made to a *pencerdd*, or chief court poet, by his pupils. But the gathering arranged by the Lord Rhys ap Gruffudd in Cardigan castle at Christmas, 1176—often called the first eisteddfod, although it was not so named—was not confined to poetry. The Welsh *Chronicle of the Princes* refers to contests there "between bards and poets, between harpists and crowders and pipers and various classes of music craft." Two chairs were set for the victors and, very like the modern national eisteddfod, "that feast was announced for a year through all Wales, England and Scotland and Ireland and other islands." There is strangely no reference to such meetings during the national revival of the period 1198–1282. A gathering at Carmarthen—1451 and 1453 have been suggested as dates—also provided for vocal and harp contests but is most famous for the new arrangement of the strict metres of Welsh poetry by Dafydd ap Edmwnt, which is still authoritative. Two 16th-century eisteddfodau at Caerwys in Flintshire in 1523 and 1568 were also mainly concerned with the bardic order. At the former, which is associated with the name of Tudur Aled, further reforms were made in metrical rules. At the latter provision was made to give application to Elizabeth I's commission to 20 noblemen of north Wales to redeem the bardic order from the "intollerable multitude of vagrant and idle persons naming theimselfes mynstrelles, rithmers, and barthes," by granting licences only to such as proved themselves worthy. It was for a similar reason that 12 men of north Wales, on May 20, 1594, called for a further eisteddfod, but apparently to no avail. The obvious decline in the merit and standing of poets is still further shown by the derogatory reference of the great Welsh grammarian, Gruffudd Robert, to contemporary eisteddfodau. The decay continued apace throughout the 17th century, and an anonymous poet, about 1650, laments the passing of the old order and bids farewell to the eisteddfod.

Soon after 1700 announcements in annual almanacks show that the name and form of the eisteddfod lingered on in gatherings of rhymesters meeting to compose on impromptu subjects, professing to follow early traditions, and taking the Welsh Grammar (1728) of Sion Rhydderch as their Bible. In 1789 this sterile institution was transformed when Thomas Jones of Corwen, an "excise-man," appealed for and obtained the patronage of a London Welsh society, the Gwyneddigion. The eisteddfod held at Bala in Sept. 1789 was the first arranged by the society. The meetings that followed—at St. Asaph (1790), Llanrwst (1791), Denbigh (1792), etc.—brought the literary life of the country into direct contact with the eisteddfod (a contact that still exists) and served as a convenient nucleus for it. They may be said to have created the eisteddfod *awdl* or ode, which is still for many the main product of the literary competitions, and were clearly a great influence on the modern eisteddfod. Such names as Twm o'r Nant, Dafydd Ddu Eryri and Gwallter Mechain are closely associated with them. A later revival came through the patronage of regional societies in Wales early in the 19th century; each society provided for an annual eisteddfod in its own region, supported in the main by gentry and clergy although providing generously for the bards. Famous among these meetings is the Carmarthen eisteddfod of 1819, arranged by the Cambrian Society of Dyfed, for it was there that the "Gorsedd of Bards," established in 1792 by Iolo Morganwg in opposition to the eisteddfod, first became an integral part of it.

It was at Llangollen in 1858 that the formation of a truly national eisteddfod was discussed and at Denbigh in 1860 that it was first decided that the annual event should be held in north and south Wales alternately. Since 1937 it has been governed by an influential and representative body, the Council of the National Eisteddfod. This largely extended the scope of the eisteddfod—by giving greater prominence to drama, prose and art, for instance—and its influence on Welsh literary and artistic life is obvious. Later developments, based in part on the older body, were *Eisteddfod Genedlaethol Yr Urdd*, a national eisteddfod arranged by *Yr Urdd*, the Welsh youth movement, and, after 1946, the international musical festival at Llangollen.

See T. Parry, *Eisteddfod Y Cymry, The Eisteddfod of Wales* (1943); G. J. Williams, "Eisteddfodau'r Gwyneddigion," in *Y Llenor*, vol. xiv, p. 11; vol. xv, p. 88 (1922–51). (G. H. Hu.)

EJECTMENT, in law, was a form of action to recover possession of real property and to determine its title, superseding the older common law forms of actions for this purpose. Initially, by this method, a lessor could evict a defaulting lessee. By the beginning of the 17th century, an elaborate fiction had developed to make the remedy available for breach of a conveyancing contract. Under this fiction, A made a lease with John Doe, a fictitious person who in theory would seek the possession of the property and bring an action for the purpose. The defendant would deny A's title to the property, and the action would be joined on the issue of title. An essential feature of this action was that the plaintiff must have had the right of possession at the time the suit was brought. This form of action entitled the plaintiff to nominal damages only. Damages, including rents arising from the property while the plaintiff was dispossessed, had to be recovered in another action entitled "trespass for mesne profits" after the successful completion of the ejectment suit.

A related form of action was forcible entry and detainer. This form of action was established by Henry VI and was a remedy for the recovery of possession of land from one who entered forcibly while the plaintiff was in possession. It is to be observed that this action did not determine title to the property but merely the right of possession.

It is doubtful whether the action of ejectment was used in the American colonial courts as it was understood in England. Most of the colonists looked with disfavour on any of the fictions of the common law. As action called ejectment was used in all the states for the simple determination of title to real property held adversely and for the recovery of such property, and in some jurisdictions for damages as well. Ultimately, these fictions were swept away. Today, the action lies for recovery of real property and fixtures and for the determination of the title. To recover, the plaintiff must have good title himself and cannot rely on the insufficiency of the defendant's title. However, prior possession

is sufficient against an intruder with no claim to the title. Although the forms of action have been abolished in name, the essentials of this action have been maintained in modern pleading.

See also LANDLORD AND TENANT; LAWS OF REAL PROPERTY AND CONVEYANCING. (E. C. S.)

EJECTOR, an apparatus for pumping gases or liquids by utilizing the eductive force of a high-velocity jet of steam, air or other fluid. There are also mechanical type ejectors employed for ejecting cartridge cases from guns or for ejecting punched or stamped articles from automatic presses and punching machines. Perhaps the most important application of ejectors is in steam-power plants where they are employed for pumping noncondensable gases from condensers to assure the maintenance of a high vacuum.

The principle of operation of a steam ejector is illustrated in fig. 1. Steam for operating the ejector enters the steam chest under pressure and expands through the steam nozzle to a high velocity. As a result, the pressure at the nozzle outlet is greatly reduced, creating a low-pressure area in the mixing chamber. The gases or vapours to be pumped are drawn into this chamber

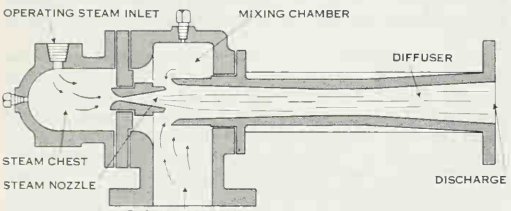


FIG. 1.—SCHEMATIC DRAWING SHOWING THE PRINCIPLE OF OPERATION OF A STEAM EJECTOR

and entrained in the steam jet. In passing through the diffuser the steam jet is gradually reduced in velocity, causing a pressure increase to the value at the diffuser discharge. Ejectors of this type are capable of increasing the pressure at discharge to as high as 10 times that maintained in the suction inlet.

To maintain a vacuum of approximately $\frac{1}{2}$ p.s.i.a. (lb. per square inch absolute) in a steam-power plant condenser, it is necessary to withdraw the noncondensable gases from the condenser and discharge them to the atmosphere at approximately 14.7 p.s.i.a. This represents a compression ratio of about 30 to 1. To accomplish this, two ejectors are employed in series in an apparatus similar to that illustrated in fig. 2. The steam employed in the first and second stages is liquefied in the inter- and after-condensers, allowing the noncondensable gases to escape to the atmosphere.

Steam ejectors are sometimes employed in industrial plants for creating high vacuums and for compressing steam. Another

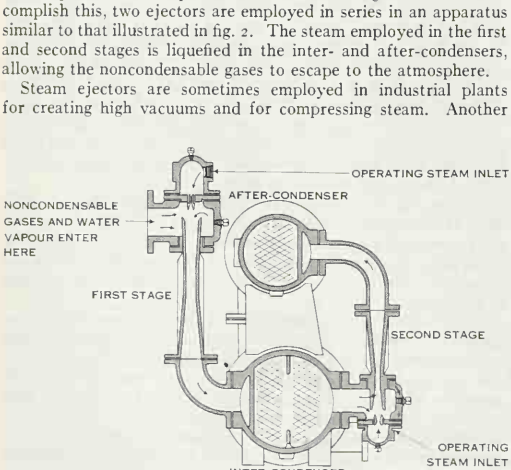


FIG. 2.—SCHEMATIC DRAWING OF A TWO-STAGE EJECTOR FOR REMOVAL OF NONCONDENSABLE GASES FROM A STEAM-POWER PLANT CONDENSER

use is in connection with vacuum refrigeration systems in which water to be chilled is sprayed into a chamber maintained at about 0.25 p.s.i.a. by means of steam operated ejectors. The water is thus vapourized and cooled to approximately 40° F.

See also PUMP. (RD. A. BR.)

EKHOF (ECKHOF), **(HANS) KONRAD (DIETERICH)** (1720–1778), German actor, was born in Hamburg on Aug. 12, 1720. In 1739 he became a member of an acting company managed by Johann Friedrich Schönemann, an association extending over 17 years. In this troupe Ekhof played leading roles in the German translations of the French plays for which the manager was crusading. During the company's stay at Schwerin (1751–56) domestic drama was added to the repertory, giving Ekhof the opportunity to play George Barnwell in George Lillo's *The Merchant of London,* Mellefont in G. E. Lessing's *Miss Sara Sampson* and Beverley in Edward Moore's *The Gamester.* In Schwerin, Ekhof initiated a dramatic academy with fortnightly meetings in which he discussed with his colleagues the problem of the actor's craft and civic responsibility. His prestige lent dignity to the short-lived Hamburg National theatre enterprise. He spent the last three years of his life in Gotha, in charge of the new court theatre, and died there on June 16, 1778. Testimonials as to Ekhof's artistic and personal integrity were written by Lessing, Goethe, Friedrich Ludwig Schröder and August Wilhelm Iffland. (A. M. N.)

EKMAN, VAGN WALFRID (1874–1954), Swedish oceanographer and physicist, best known for his qualitative explanation of ocean-current distribution known as the Ekman spiral, was born in Stockholm on May 3, 1874. After receiving his doctoral degree at Uppsala in 1902, he was assistant at the International Laboratory for Oceanographic Research at Oslo (then Christiania), Nor., 1902–08, where he was associated with the arctic scientist Fridtjof Nansen. During that time Ekman explained the curious hindrance to a ship's progress known as dead water, performed the experiments on which the tables of compressibility of sea water are still based and invented the Ekman current meter, a widely used mechanical deep-sea meter for measuring currents in the ocean. In 1905 he published an oceanographic classic "On the Influence of the Earth's Rotation on Ocean-Currents" (*Arkiv för Matematik, Astronomi och Fysik,* Stockholm), which deals with the currents produced by wind friction on the ocean surface. The theoretical vertical distributions of velocity close below the ocean surface and close above the ocean bottom—the latter applying also to the distribution of wind velocity above the ground—derived in this paper are known in oceanography and meteorology as the Ekman spiral (*q.v.*). Ekman was professor of mechanics and mathematical physics at Lund university, Sweden (1910–39). He died on March 9, 1954. (R. B. My.)

EKMAN SPIRAL, the theoretical vertical distribution of wind or ocean currents, in terms of speed and direction, near the surface under idealized conditions.

Both in the atmosphere and ocean, one can define "layers of frictional influence" in which the force of friction is of the same order of magnitude as the other forces. The depths of such layers vary, depending, for example, on the strength of the surface wind in the case of the water, and on the strength of the upper-air flow in the case of air. Typical thicknesses for layers of frictional influence are 2,000 ft., in the atmosphere.

If friction were produced by molecular viscosity, the frictional layers would only be a few feet thick. The much greater thickness of friction layers in the atmosphere and ocean shows that friction there is produced by eddies in the fluid itself. In the atmosphere, such eddies bring the slow air close to the ground in contact with the more rapidly moving air higher up, thus slowing it down. This slowing down, accompanied by a turning due to the earth's rotation, can be accounted for by a force, which in analogy to molecular friction is called the force of eddy friction.

In the ocean the uppermost layers are dragged along with the surface wind. Due to the vertical exchange of momentum, brought about by the eddies, the water in the lower layers is dragged along as well. Again, a turning of the current with increasing depth

accompanies the changed velocity distribution.

Qualitatively, the wind distribution in the friction layer of the air starts from a relatively slow wind near the ground, blowing toward low pressure; with increasing height, the wind speeds up and blows more and more parallel to the lines of constant pressure (isobars). The end points of the wind vectors (drawn from a common origin) form a spiral, as indicated in the figure. In the ocean the surface current makes an angle of the order of 45° with the surface wind (in the absence of deep-water current), and the speed of the current drops gradually with depth, turning clockwise (counterclockwise in the southern hemisphere) at the same time. If there are deep-water currents, the frictional currents just described are added vectorially to the deep-water currents.

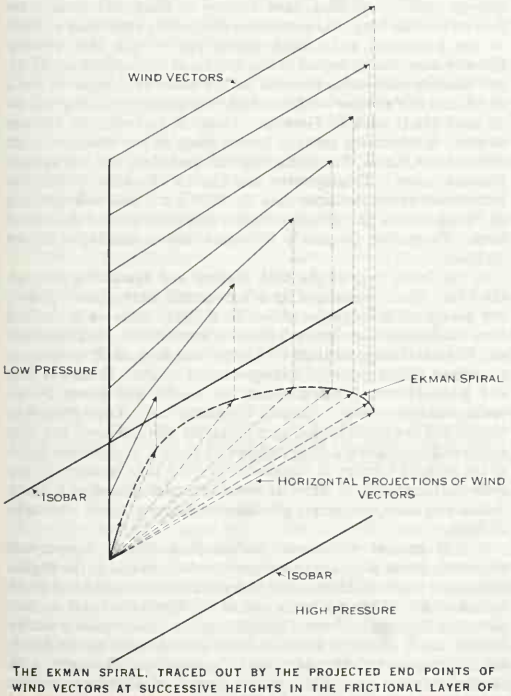
The current distribution in the oceanic friction layer was first explained qualitatively by V. W. Ekman. He integrated the equations of motion with respect to height under the following assumptions: (1) no acceleration; (2) constant eddy viscosity; (3) the slope of isobaric surfaces invariant with height; and (4) infinitely deep water.

Ekman's theory agrees qualitatively with observations. The theory was modified later by C. Rossby and R. Montgomery to allow for finite ocean depth and variable eddy viscosity.

F. A. Åkerblom carried out the analogous integration for the atmosphere, subject to similar assumptions as those made by Ekman. As a result, the vector difference between the wind at any level and the free air current decreases exponentially with height, and turns clockwise (in the northern hemisphere) at a constant rate. The wind becomes parallel to the isobars at a height H, commonly called the height of the "gradient wind level." This height in Åkerblom's theory is proportional to the square root of the eddy viscosity.

In general, the theories by Ekman and Åkerblom, and their modifications, account qualitatively for many properties of the current distributions in the friction layers of atmosphere and oceans; quantitatively, however, all the theories are inexact, and, in particular, do not account for the variation of the wind distribution from day to night. This variation was explained qualitatively in 1841 by J. P. Espy and in 1888 by Wladimir Köppen, and quantitatively by A. Blackadar and K. Buajitti in 1957.

(H. A. A. P.)



THE EKMAN SPIRAL, TRACED OUT BY THE PROJECTED END POINTS OF WIND VECTORS AT SUCCESSIVE HEIGHTS IN THE FRICTIONAL LAYER OF THE ATMOSPHERE IN THE NORTHERN HEMISPHERE

EKRON, an ancient Palestinian city often identified with 'Agir, about 25 mi. S.E. of Tel Aviv–Jaffa in modern Israel, though other sites are possible. Although included by the Israelites in the territory of Judah and mentioned in Josh. xix, 43 as a city of Dan, it is specifically mentioned as one of the cities not captured by Joshua (*cf.* Josh. xv, 45 and xiii, 3). It was occupied by the Philistines in the days of Samuel. The sanctuary of Beelzebul, "Lord of the Abode," was at Ekron, and the restoration of the Ark of the Covenant to Israel was by way of Ekron to Beth-Shemesh. According to Assyrian annals Padi, king of Ekron, was imprisoned by his subjects and handed over to Hezekiah, but regained his throne after Sennacherib had besieged Jerusalem (701 B.C.) and released his loyal vassal. In 147 B.C. the town was ceded to Jonathan Maccabaeus by Alexander Balas. (J. S. I.)

EKWALL, (BROR OSCAR) EILERT (1877–1964), Swedish linguist, the leading authority on the place-names of England, was born on Jan. 8, 1877, at Sandsjoe, Swed. He was professor of English in the University of Lund (1907–42). Ekwall died on Nov. 30, 1964. His earlier work was in English linguistic history, but after 1917 he devoted himself principally to the study of English toponymy. His chief works in this field are: *Scandinavians and Celts in the North-West of England* (1918), *English River-Names* (1928) and *The Concise Oxford Dictionary of English Place-Names*, 3rd ed. (1947). His work exhibits thorough command of the sources and sound philological judgment.

(M. S. Br.)

ELAGABALUS (corrupted to HELIOGABALUS by the etymological fancies of some Greek writers), Roman emperor A.D. 218–222, was originally called Varius Avitus, and was the son of Varius Marcellus and of Julia Soaemias, daughter of Julia Maesa, the sister-in-law of the emperor Septimius Severus. He was born in 204 at Emesa (Homs) in Syria, where his mother's family were hereditary high priests of the local Baal, Elah-Gabal. In 217 his cousin Caracalla, Severus' son and successor, was murdered and replaced by the praetorian prefect Macrinus. Maesa and Soaemias won the support of the neighbouring legion for the boy, whom they passed off as an illegitimate son of Caracalla, and the remainder of the eastern armies soon deserted Macrinus (218). The senate then acknowledged Elagabalus, who assumed the honourable name of Marcus Aurelius Antoninus. He imposed the worship of his Baal upon the Roman world, executed a number of resentful generals, and pushed into high places a number of favourites distinguished by their personal beauty and their humble and alien origin. The boy's effeminate sexual behaviour, practised openly in the palace where the grave philosopher Marcus Aurelius had ruled less than 40 years before, outraged Roman opinion. He was persuaded to adopt his docile cousin Alexander as his son and heir (221); when he changed his mind and sought to depose Alexander, the imperial guards mutinied, killed him and made Alexander emperor.

EL ALAMEIN, BATTLE OF, one of the decisive engagements of World War II. It opened on the night of Oct. 23–24, 1942, with a large-scale assault westward from the line of El Alamein (al 'Alamayn) by Gen. B. L. Montgomery's British 8th army on Field-Marshal Erwin Rommel's German-Italian force entrenched behind antitank minefields about 60 mi. W. of Alexandria in Egypt. It ended on Nov. 4 with the Axis army's precipitate retreat.

Montgomery had assumed command of the 8th army on Aug. 13, 1942. He found it holding a north-and-south line of 35 mi. from the sea near Tall al 'Isa to Qarat al Humaymat; it was short of effective tanks and of antitank guns and had lost confidence in its higher command. The Axis army, though superior in armour, was lacking in reliable infantry and in transport and also suffered from supply difficulties caused by heavy shipping losses on the long line of communications from Naples to Tripoli and Benghazi. Having strengthened his position on the key 'Alam al Halfa' ridge

to protect his weakly-held southern flank, Montgomery set about planning an offensive operation while the process of re-equipping and retraining his army continued. Rommel attacked on the night of Aug. 30–31, penetrated XIII corps's front in the southern sector, but failed to capture the 'Alam al Halfa' feature, and after seven days' fighting the Axis forces were compelled to withdraw with considerable loss in men and vehicles.

Montgomery's original intention was to attack in the northern sector with four infantry divisions of XXX corps and, having secured a passage through the enemy's antitank minefields, to pass the two armoured divisions of X corps through the gap so as to position themselves astride the Axis supply line. In view, however, of the British force's state of training the plan was modified, and it was decided that X corps, advancing behind the infantry of XXX corps, should take up defensive positions to hold off the enemy armour while the British infantry set about destroying the enemy infantry. In the south, XIII corps was to make a subsidiary attack so as to pin down enemy armour and prevent its intervening in the decisive northern sector. Elaborate measures were taken to achieve tactical surprise by persuading the enemy that the main thrust would be delivered in the south. The Royal Air Force undertook a series of day and night bombing raids against Axis airfields which virtually drove enemy aircraft out of the sky. By Oct. 23 the R.A.F. had unchallenged command of the air.

At 9:40 P.M. on Oct. 23 more than 1,000 British guns opened a bombardment of the enemy artillery positions. At 10 P.M. they switched to a moving barrage on the Axis forward defenses. In bright moonlight four divisions of XXX corps advanced behind the barrage, and by dawn they had gained most of their objectives. The two armoured divisions of X corps passed through the two corridors cleared of mines by the Royal Engineers and took up positions on the Mutayriah ridge. In the south, XIII corps had only limited success and was ordered not to press the attack after Oct. 25.

There followed days of confused and hard fighting, but German tank counterattacks were repulsed. On the extreme right, the 9th Australian division attacked northward toward the coast with considerable success in the nights of Oct. 25–26, 28–29 and 30–31. These local surprise thrusts attracted the German reserves to the northern sector near the coast, so that the Italian infantry were left unsupported by German armour.

Meanwhile Montgomery had been regrouping by withdrawing troops into reserve, and on the night of Nov. 1–2 he launched the 2nd New Zealand division westward against Italian positions in "Operation Supercharge." The attack was successful and enemy tank counterattacks were beaten off. Next day there were signs of enemy withdrawal, and on the night of Nov. 3–4 a hastily mounted attack by the 51st Highland division penetrated to the Tall al 'Aqaqir track against very slight opposition. This stroke virtually isolated the German forces in the north from the Italian divisions in the centre and south. Next morning X corps was despatched in pursuit, the Germans under air attack withdrawing along the coast road and leaving the Italians, without transport, to be captured at leisure by XIII corps.

The German-Italian army lost about 10,000 killed and wounded with 30,000 prisoners, 1,000 guns and more than 500 tanks. Four German and eight Italian divisions were virtually destroyed. The 8th army had about 13,000 casualties and lost about 450 tanks, but many of these latter could be repaired. The battle was the starting point of the 8th army's spectacular advance toward Tunis.

(M. E. S. L.)

ELAM, the biblical name given to a country in southwestern Persia approximately equivalent to the modern Iranian province of Khuzistan (*q.v.*). The name is derived from the native (Elamite) name *Ha(l)tamti,* Assyro-Babylonian *Elamtu.* Mentioned in ancient sources are four prominent geographical names within Elam: Awan, Anshan (*qq.v.*); Simash and Susa (*q.v.;* Shushan, Susiane in Greek; modern Shush), but only the last has been identified as a city and partially excavated.

Geographically and culturally, part of Elam belonged to the southern lowlands of the Tigris-Euphrates river systems; at various times its peoples fought the inhabitants of Sumer and Akkad. Periodically, Elam also included river and mountainous areas to the northeast and southeast of the Mesopotamian lowlands; its exact limits, however, cannot be defined.

Throughout the late prehistoric and predynastic eras, Elam was closely tied culturally to Mesopotamia. Historical and economic documents of Sumer refer to the area at about 2450 B.C., by which time the Elamites had attempted to write their own language in pictographic symbols. Perhaps because of domination by the Mesopotamian dynasty of Akkad (*c.* 2350–2230), Elamites adopted the Sumero-Akkadian cuneiform script, with which they wrote Elamite and Semitic (Akkadian) inscriptions on clay tablets and stone steles. Simultaneously with this cultural advance, an independent dynasty of Elamite kings (most notably, Puzur-Inshushinak) briefly ruled an area including Susa and Awan. The dynasty ceased, perhaps because Elam was invaded by northern neighbours from the Zagros mountains, the Guti.

Following the Guti, the 3rd dynasty of Ur controlled much of Elam, though with some difficulty, during the 21st century B.C. Sovereigns of Ur maintained governors in many Elamite centres, warred against Elamite recalcitrants, sent political and trading missions into the area and married off royal daughters to Elamite officials. As the power of Ur in turn declined and independent rival kingdoms emerged in Mesopotamia, Elamites reasserted their independence, sacked and destroyed the capital city and carried off to Elam one of the last kings of Ur (*c.* 2030 B.C.).

In this turbulent period Elamite rulers (*e.g.,* Idaddu) were known as "governors" (*i.e.,* kings) of Elam and Simash and became embroiled in Mesopotamian affairs. Subsequently, a cultural feature emerged which remained to the very end of Elamite history and was unique in the area. This feature involved an aspect of matrilinear succession: sovereignty was hereditary through women, in that a new ruler was always "son of a sister" of some member of an older sovereign's family. Owing probably to the regulations of this succession, each sovereign had first been viceroy (*sukkal*) of Susa, then viceroy of Elam and Simash before he became king. As cuneiform documents excavated at Mari on the Euphrates make clear, during part of the 18th century Elamite mercenaries served in the armies of Mesopotamian kings, and Elamite emissaries traveled as far afield as Aleppo in Syria and Hazor in Palestine—a fact which may help to elucidate ch. xiv of the biblical book of Genesis. There is, however, an absence of exact synchronism between known kings of the Mesopotamian cities of Isin, Larsa, Eshnunna, Babylon and Mari, and the various Elamite rulers. This dynasty and the 1st dynasty of Babylon terminated about the same date (*c.* 1600 B.C.), and new invaders of Mesopotamia, the Kassites (*q.v.*), may have caused the fall of both. Thereafter, for nearly 300 years, almost nothing is known of Elam.

In the latter part of the 13th century and continuing through the 12th, Elam re-emerged as a substantial international power, but nearly all information about its military prowess is derived from inadequately understood Elamite inscriptions. Hubanumena and Untash-Huban, as kings of Anshan and Susa, built temples to a score of deities, some of Mesopotamian origin. Untash-Huban and Kidin-Hutran began a search for wealth and power by invading Mesopotamia. Shutruk-Nahhunte and Kutir-Nahhunte almost, but not quite, turned such invasions into conquest, but only succeeded in securing a large number of ancient monuments (such as the Stele of Victory of Naram-Sin and the stele bearing the law code of Hammurabi). Most of these were rededicated to Elamite deities and set up in temple precincts of Elamite cities, especially at Susa.

A still greater conqueror, Shilhak-Inshushinak, campaigned vigorously in the Babylonian lowlands, in Assyria and in the Zagros mountains north of Elam. For at least a short period his domain included most of Mesopotamia east of the Tigris river and reached almost to Persepolis; from it flowed into the Susa capital a sizable income, much of which seems to have been devoted to the building of temples and temple towers (ziggurats). However, with dramatic abruptness, this greatest period of Elamite conquest ended when Nebuchadrezzar I of Babylon (*c.* 1130 B.C.) captured

and despoiled Susa. For almost 300 years thereafter, Elamite history is unknown.

The years from 750 to 650 witnessed a new and partially independent Elam, but Assyria, the dominant power, showed skill in playing one claimant to the Elamite throne off against another. With the Babylonian rebel Merodach-baladan, the Elamites once interfered in Babylonian affairs against Sennacherib and even sacked the city of Sippar, but this episode had no sequel except that Elam became a refuge for recalcitrant Babylonian subjects of Assyria. Assyrian patience was exhausted during the latter days of the reign of Ashurbanipal, who in 640 invaded Elam, sacked Susa and nearby cities and deported some of their leading citizens to far-off Samaria in Palestine. Subsequently, at least a part of Elam was controlled by the neo-Babylonian kings (including Nebuchadrezzar II), while another part appears to have been subject to the Persians.

From the rise of the Persian empire (550), Elam was merely a satrapy, paying tribute to Persian kings. Susa, however, was one of the three most important cities of the realm; there the kings resided for about three months each year and built massive palaces and audience halls, since destroyed, which undoubtedly resembled those at Persepolis. Susa, at one end of the Great Royal road leading from Sardis in Anatolia, a distance of less than 90 days by royal post, was well known to Greek authors such as Herodotus, Ctesias and Arrian. The episodes recounted in the biblical book of Esther reputedly occurred at Susa in the days of Ahasuerus (presumably a Xerxes).

Elam's cultural accomplishments do not appear to have been extensive. During the days of empire the only materials written by known Elamite scribes comprise documents composed at royal command to recount military achievements or temple installations; such business or economic affairs as received written form were inscribed in the borrowed Akkadian. Still less is known of Elamite religious beliefs, and no epic or religious materials in the language have been discovered. Prominent among native Elamite deities were Inshushinak ("the Lord of Susa"), Huban, Kiririsha, Nahhunte and Shimut, but little is known of the particular powers or features with which each was associated.

The Elamite language is not clearly understood. It was of a type which has often been called agglutinative, but had no known ancient relatives and assuredly no modern descendants. In the Persian period all royal texts (*e.g.*, that of Bisitun) were inscribed in three languages, of which Elamite was one; several thousand documents, excavated at Persepolis, were written at a time when Elamites served as Persian court accountants. These were helpful in deciphering the Elamite of the Persian period, but the structure of the language had undergone considerable change from that current during the classical period of Elamite history (beginning *c.* 1250), and the documents of this earlier period are translatable only with difficulty and the help of conjecture.

Elam's art and architecture, though known only from inadequate excavations, strongly resembled those of Babylonia, and its temples and temple towers, in ground plan and decoration, could equally have been erected in Babylon. Minor variations existed (especially in the designs on cylinder seals), but Mesopotamia clearly provided much of the inspiration.

BIBLIOGRAPHY.—L. Vanden Berghe, *Archéologie de l'Iran ancien* (1959); R. Ghirshman, *Iran* (1954); A. T. Olmstead, *History of the Persian Empire* (1948); E. Herzfeld, *Iran in the Ancient East* (1941); George G. Cameron, *History of Early Iran* (1936), *Persepolis Treasury Tablets* (1948). (G. G. CA.)

ELAND, the largest of the African antelopes, *Taurotragus* species. They are somewhat oxlike in appearance and have a distinctive dewlap (a loose flap of skin drooping from the neck); old bulls have a thick forelock. The bull of the common eland (*T. oryx*), a species akin to the kudu, but with horns present in both sexes, and their spirals much closer, may reach a height of 6 ft. at the shoulders and a weight of up to 1,200 lb. This species, in herds of 50 or more animals, grazes on the grassy plains from south-central to southern Africa. In the southern form the body is wholly pale fawn, old bulls being blue-gray in colour, but north of the Orange river (Republic of South Africa) the body is often marked by narrow, vertical white stripes.

ARTHUR W. AMELER FROM NATIONAL AUDUBON SOCIETY

ELAND (TAUROTRAGUS ORYX)

North of the west African rain forest, from Senegal to the Bahr el Ghazal district (Sudan), a larger species, the giant or Derby eland (*T. derbianus*), occurs, characterized by a reddish coat, a dark neck and longer, more massive horns. *See* also ANTELOPE.

ELASMOBRANCH, a term, after Elasmobranchii, given to the modern sharks, skates and rays (all of which are also sometimes called selachians), for which several classification schemes have been proposed. *See* CHONDRICHTHYES; RAY; SHARK.

ELASTICITY. The word *elastic*, derived originally from the Greek verb *elaunein*, to drive, has acquired its present meaning largely from the writings of Robert Boyle. In 1660, in a description of his experiments relating to the spring of the air he says, "There is a spring or elastical power in the air in which we live." This spring he attributed to an effort made by each particle of air to drive its neighbours away. For the sake of illustration, a particle of air was at that time pictured with little springs attached to it, these springs being compressed by those of neighbouring particles.

The chief characteristic of a spring is its power of recovering a former durable shape and size when the load causing a temporary deformation is removed. The spring possesses this power, whether it is stretched or compressed. A similar power of recovery, when deforming forces are removed, is possessed to some extent by any solid body. This power is described in a general way by saying that the body is elastic. If recovery is prevented by some other body or bodies, the effort which the body makes to recover its original durable form may be represented by an elastic force or system of elastic forces acting on the obstruction. This application of the word elastic to the forces exerted by deformed solid bodies was made by Boyle. In an essay on effects of motion, written in 1685, he says, "By the elastical force of the bent bow the string is brought into a violent state of tension."

LAWS OF ELASTICITY

Boyle's Law.—The first law of elasticity was published in 1662 by Boyle, and in 1676 by Edmé Mariotte. It is now called the law of the isothermal expansion or contraction of a "perfect" gas, and is generally written in the form

$$pv = f(\theta)$$

where v is the volume of unit mass of the gas and p the pressure intensity. The quantity on the right-hand side depends only on the temperature θ, which was kept constant in Boyle's experiments. When θ is the *absolute* temperature, as measured by an air-thermometer, $f(\theta)$ becomes simply $R\theta$ where R is a constant depending on the nature of the gas. The quantity v is the reciprocal of the density (ρ), and is frequently called the *specific volume*, while p is called simply the *pressure*. The law indicates that, under isothermal conditions, the pressure is proportional to the density. For a real gas this law has only a limited validity; it fails at high pressures and low temperatures, *i.e.*, under conditions close to those at which liquefaction occurs. If dp and dv denote small changes in p and v at constant temperature, Boyle's law (*q.v.*) gives the equation

$$p = -v\frac{dp}{dv} \ (=k)$$

The quantity on the right is called the *volume-elasticity;* under isothermal conditions it is equal to the pressure, but under adiabatic conditions, *i.e.*, when there is neither gain nor loss of heat, it is equal to γp where γ, the ratio of the specific heats at constant pressure and constant volume, is always greater than 1 and is about 1.4 for air. In Boyle's law the pressure is supposed to be constant throughout the unit mass of gas, but, since the choice of

this unit is arbitrary, the law may be applied to a small portion of the gas, whose size may be diminished indefinitely. The limiting value of the ratio dv/v is called the *dilatation* (Δ), and $-\Delta$ is called the *compression*. The volume-elasticity k is thus the ratio of the increase of pressure to the resulting compression. With a suitable definition of pressure this law is also applicable to solids, and to prevent misunderstanding the pressure so defined is often called *hydrostatic pressure*.

Hooke's Law.—The first law of elasticity for solid bodies was discovered by Robert Hooke in 1660, and published in 1676 in the form of an anagram meant to represent the words *Ut tensio sic vis*. *Tensio* is understood to mean what is now called *extension*. This, in the case of a wire stretched to a new length L, is the ratio of the gain in length, $L-l$, to the original length l. Interpreting *vis* as the tension in the wire, the law states that the tension is proportional to the extension. If the tension is regarded as a force, the factor of proportionality is found not to be a physical constant for the substance of which the wire is made but to depend upon the area of the cross section of the wire as well as upon the properties of the substance. For this reason it is more convenient to regard tension as a *stress*, *i.e.*, the ratio of a force to the area across which it is transmitted. A more general and precise definition of stress, proposed by Augustin Louis Cauchy, will be given later, but the present one will suffice to indicate the physical dimensions of a stress and the most convenient units of stress. Extension will be regarded as a particular type of *strain*, a physical quantity which is always the ratio of two lengths and therefore a mere number, so no unit of strain is needed.

Hooke's law may be expressed in the form of an equation

$$S = Es$$

where S is the stress, s the strain and E a modulus of elasticity, now called *Young's modulus*, in honour of Thomas Young, who gave it a physical meaning. This law does not hold for very great values of S, and even with small values of S it does not give a complete description of the elastic behaviour of the substance, as there is a lateral contraction when a rod or wire is stretched. For a long time, however, problems in the theory of elasticity were treated with the aid of this law alone, the first use of the law being made by Mariotte, who discovered it independently. He pointed out that, when a beam is bent, some of the longitudinal fibres are stretched and others contracted.

Bending of Rods.—This idea was developed by Jacques (Jakob, James) Bernoulli and Leonhard Euler into a theory of thin elastic rods, which involved the idea of a *couple* whose moment resists the *bending moment* and is proportional at each place to the curvature of the rod when bent. If the deflection y at a distance x from one end of the rod is small, the equation for the bending moment M is approximately

$$M = EI\frac{d^2y}{dx^2}$$

where EI is a constant called the *flexural rigidity* of the beam, E is Young's modulus for the material and I is a constant depending on the shape of the cross section of the rod. In the case of a rod, of length l, free to turn about its ends and bent by forces acting along the line joining these ends, the bending moment is $-yP$, and this is balanced by the resisting moment if

$$EI\frac{d^2y}{dx^2} + Py = 0$$

Since y is zero when $x = 0$ and when $x = 1$, an appropriate solution is $y = A \sin \alpha x$, where $EI\alpha^2 = P$ and $A \sin \alpha l = 0$.

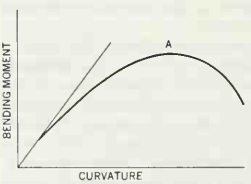
The last equation gives either $A = 0$ or $\alpha l = n\pi$, where n is an integer. A more complete theory, in which an exact expression for the curvature is used, indicates that the straight form, given by $A = 0$, is stable if $Pl^2 < \pi^2 EI$ and unstable if $Pl^2 \geqq \pi^2 EI$. In the latter case a bent form is stable. This result can be extended, with suitable modifications, to columns and frameworks supported and loaded in various ways. The fact that the criterion for the failure of a rod, through instability of the straight form, depends upon the modulus E adds to the importance of this modulus as a physical quantity. The theory has been confirmed by the very careful experiments of Andrew Robertson. In practice a rod generally buckles under a load less than Euler's critical load, because the load is badly centred or because the rod has a slight initial curvature. The applicability of Euler's theory was discussed by Theodore von Karman in 1906.

In the theory of the flexure of a thick beam, given by C. A. Coulomb, a line through the centroid of the cross section, called the *neutral line*, is found to possess the property that the extension of a longitudinal fibre through a point Q of the section is represented by Cy, where y is the distance of Q from the neutral line and C is the curvature of the line of centroids at the position of this section. This relation is derived on the assumption that the cross sections remain plane after bending.

The couple required to maintain flexure is then parallel to the plane of bending if $\iint xydxdy = 0$, where x is measured parallel to the neutral line. The couple is then given by the formula

$$M = \iint (ECy)ydxdy = ECI$$

where I is the moment of inertia of the cross section of the beam about the neutral axis. The stress in the fibre under consideration is a tension or compression, according as y is positive or negative; it is given by the quantity ECy which may be expressed in the more convenient form My/I. There may be an additional longitudinal stress in the fibre due to longitudinal end loads. If the total stress in a fibre is greater than the material can sustain there will be rupture. When the problem of flexure is treated by more exact methods, it is found that the bending moment is not always proportional to the curvature of the line of centroids. This was first noted by Karl Pearson, for a special distribution of lateral loads. In the case treated by Barré de Saint Venant, in which the proportionality was confirmed, the bending was supposed to be produced by forces applied to the terminal sections. The proportionality fails even in this case if the beam is a very thin tube and is consequently capable of large displacements. A theory developed for these conditions by L. G. Brazier gives a relation between bending moment and curvature of the type shown in the accompanying figure. If the bending moment is increased above the maximum value, corresponding to the point A, the beam must collapse. This is a new form of instability and is an indication of the new results which may be expected from the general theory of elastic stability which has been developed by R. V. Southwell and applied to rods, tubes and plates.



TYPICAL RELATIONSHIP BETWEEN BENDING MOMENT AND CURVATURE OF A BEAM

Torsion and Shearing Stress.—The advantage of using the idea of stress instead of that of force is seen again from a consideration of the torsion of a shaft of length l. If θ is the angular displacement of one end relative to the other, and T the total twisting couple transmitted along the shaft, there is, for small values of θ, a relation of the type

$$lT = K\theta$$

where K is a constant which may be called the *torsional rigidity* of the shaft. This constant K is not, however, a physical constant for the material; it depends in a rather complicated way on the shape and size of the cross section. Cauchy's theory showed that, in the case of a cylindrical shaft of uniform cross section, K could be written in the form μJ, where J is the moment of inertia of the cross-sectional area about the axis of the cylinder and μ is a physical constant called the *rigidity* of the material. Saint Venant later gave a general theory of torsion, in which K has the same form as before except that J is now a geometrical constant, the determination of which for a given section depends on the solution of a problem of potential theory. The stress which, in this theory, is supposed to be transmitted across any element of area of a cross section, is a *shearing stress*, *i.e.*, a stress derived from

a force tangential to the area. This stress is expressed in terms of geometrical quantities by means of an extension of Hooke's law appropriate for shearing stresses. This and other extensions of Hooke's law will be given later when the ideas of stress and strain are made more definite. A mechanical method of determining K by means of a soap film stretched across a hole whose shape is the same as that of the cross section of the shaft has been developed by A. A. Griffith and Geoffrey I. Taylor from a suggestion made by Ludwig Prandtl.

The Spiral Spring.—In 1848 James Thomson pointed out that the action of an ordinary spiral spring of circular cross section depends mainly on torsion. The spring consists of a thin wire coiled into a helix of small pitch and radius c. When one end is fixed at a point O and the other attached to a weight W, it may be assumed that the axis of the coil is vertical and along the line OW The twisting moment is then Wc, and the twist for a length l of coil is given approximately by the equation

$$lWc = K\theta$$

The vertical displacement due to a twist $\theta\delta s/l$ in an element of length δs is $c\theta\delta s/l$. The total vertical displacement is therefore

$$c\theta = lWc^2/K$$

There is no appreciable horizontal displacement of the weight, because the horizontal components due to the different elements will on the whole neutralize one another. Unless the pitch is small it is necessary in a more exact theory to take into consideration the bending moment and the flexural rigidity. The spring then tends to uncoil when stretched.

Units of Stress.—The centimetre-gram-second (c.g.s.) unit of one dyne per square centimetre is rather small, and so a unit of one kilogram per square centimetre is used by many scientists. This unit will be denoted here by the symbol κ. Other convenient symbols which will also be used are τ for a stress of one ton per square inch, π for a stress of one pound per square inch and α for a unit pressure of one atmosphere. The transformation from one unit to another may be made with the aid of the relations

$$1000\alpha = 6.5611\tau = 14700\pi = 1033\kappa = 1.013 \times 10^9 \text{ c.g.s. units}$$

The unit used in the International Critical Tables is one kilomegabyre, a megabyre being equal to 14.5π. Young's modulus for steel is a little greater than $13{,}200\tau$, consequently a tensional stress of 13.2τ will produce an extension of only about 0.001.

LIMITS OF ELASTICITY

Behaviour of a Substance Under Great Stress.—When a rod or wire has been stretched by a load, it recovers its original length and elasticity after the removal of the load, provided the stress produced by the load does not exceed a certain threshold value. This value is called the *elastic limit;* it generally gives very closely the limit of the range for which the strain is proportional to the stress, but may be slightly less than the proportional limit. With other kinds of stress the reverse may be the case. In Albert A. Michelson's experiments with twisted prisms the elastic strain produced by a stress p was found to be given by an expression of type Cpe^{hp}, where C and h are constants, while in P. W. Bridgman's experiments with substances subjected to hydrostatic pressures up to $12{,}000\kappa$ no elastic limit was detected, though there was always a limit of proportionality. As the pressure increases the compressibility decreases, the decrease for solids being generally less than for liquids.

With stresses beyond the elastic limit the behaviour is different for different substances, but generally the material becomes plastic when the load is applied, and the magnitude of the extension corresponding to a given load is indefinite and varies with the time. With a sufficiently large load the extension may increase with time in such a manner as to lead to rupture; with a smaller load the rate of increase, after quickly reaching a maximum value, may gradually decrease to zero, just as if the motion were resisted by a kind of viscous drag. In this case a state of plastic equilibrium is approached. On the removal of the load the rod may not return to its original length, but may assume a slightly greater length. Sometimes the old or a new length is approached gradually, as if a viscous drag were again operative. This gradual creeping to a different length is called *elastic afterworking;* it was discovered by Wilhelm Weber in 1835.

If, after receiving a permanent set or deformation, a rod is reloaded, the strain may not be exactly proportional to the stress, but in some cases it is, and the proportional limit may be either greater or less than before. Sometimes the proportional limit is surprisingly low, but if the rod be allowed to rest for some days after the overstrain it recovers the property of linear elasticity. The stress at which the strain begins to increase very rapidly as the load is increased is called the *yield point.* For mild steel it may be about 18τ. The cross section of the bar now diminishes in area, and just before rupture a narrow neck forms. The extension may then be about 30.6% and the reduction in area 67.4%.

In order to avoid calculation of the stress at each stage, engineers generally use a load-extension diagram in preference to a stress strain diagram, and the stress at rupture, or *ultimate strength,* is a nominal stress calculated on the basis of the original cross section. The actual stress is much greater than this; and it increases right up to the moment of rupture, while the load may actually decrease after the yield point has been passed. The ultimate strength of ordinary mild steel is about 30τ, that of piano wire about 120τ, while the ultimate strength of wood is generally less than 8τ.

The Principle of Least Work.—The stresses in the members of an over-rigid, that is a redundant, framework of elastic rods may generally be obtained by means of the principle of least work introduced by L. F. Menabrea and C. A. Castigliano. The method is to represent the stresses in a certain number of rods by algebraic symbols, the number being just sufficient to determine uniquely the stresses in the remaining rods when the framework is acted upon by external forces which, if compounded together, would form a system in equilibrium. The total strain energy of the framework is then calculated by summation, using the theorem that the strain energy of a rod of length l and uniform cross-sectional area A is $lF^2/2EA$, where F is the force acting along the rod and E is Young's modulus. This is in fact just l times the area under the curve in the force-extension diagram, and is the volume of the rod multiplied by the area under the curve in the stress-strain diagram. The values to be given to the algebraic symbols are now those which will make the total strain energy a minimum. The conditions for a minimum take the form of linear algebraic equations with a single set of solutions.

The stresses thus found do not necessarily form a physically possible system, for it may happen that the calculated stress in one or more of the rods exceeds the proportional limit. One method of treating the problem in such a case is to make the simplifying assumption that, when the stress in a rod reaches a certain yield point, it cannot exceed this value. Assuming that certain rods behave as plastic and the rest as elastic members, the stresses in the plastic rods are given by the known yield points, while the stresses in the elastic rods may be determined by the methods of statics or by an application of the principle of least work.

The stresses determined by the ordinary methods of statics or by the principle of least work may fail to represent a physical possible system for another reason. If the stress in a member is compressive and is greater than a certain critical value, the member will buckle. This critical value depends on the modulus E, and may be less than the yield point of the material. For some purposes it is useful to compare an elastic solid with a structure of elastic rods, and one of the important problems of the theory of elasticity is the determination of the critical loads for struts, structures and elastic plates of various shapes. These loads may be expected in general to depend not only upon E but also upon the other elastic constants just as in the case of a structure they depend upon the geometrical form. In some cases the problem can be treated by the elementary methods, in which the displacements of particles are treated as small, but for a full elucidation of some points a more exact theory is needed.

The Growth of the Structure Theory.—In the 17th and 18th centuries some problems relating to the bending and vibra

tion of beams and columns were partially solved, although there was then no general theory of elasticity. The treatment of other problems was, however, erroneous and the need of a comprehensive theory became apparent. Curiously enough the general equations of the mathematical theory of elasticity originated not from the more or less empirical attempts to solve special problems but from a combination of definite mathematical hypotheses with certain speculations regarding the structure of matter.

Following up Boyle's idea of particles of air endowed with the properties of a spring, Sir Isaac Newton endeavoured to explain the simple proportionality between air pressure and density by introducing the idea of a repulsive force between two molecules depending only on their mutual distance. Newton actually used an inverse distance law, but was careful to state that the force must be regarded as operative only when the distance does not exceed a certain length. An atom or molecule was thus regarded as having a more or less definite sphere of action, and attractive forces of various intensities were regarded as the cause of chemical action. The idea of molecular attractive forces with a limited range of action was used also by P. S. Laplace in his theory of surface tension.

The use in physics of forces depending only on the distance and acting only over a definite range had thus already received some sanction when C. L. M. H. Navier used the idea in 1821 to obtain a set of general equations for the equilibrium and vibration of an elastic solid. Navier's molecules, like those of R. G. Boscovich, were practically centres of force, and his equations for an isotropic substance involved only a single elastic constant analogous to Young's modulus. The theory was carried a step further by Cauchy, who imagined the molecules to form a homogeneous assemblage or lattice. He thus obtained a generalization of Hooke's law involving 15 elastic constants, and regarded it as applicable to any crystalline or aeolotropic substance.

The need for a greater number of elastic constants for both crystalline and isotropic bodies was realized by S. D. Poisson who, prior to Navier's work, had used the Newtonian ideas in some work on the equilibrium of an elastic surface. Poisson proposed to go back to the Newtonian conception of a molecule with a definite shape and size, and to consider the effect of changes in the orientation as well as the position of the molecules of an elastic body. The calculations were completed by Woldemar Voigt in 1887. Shortly afterwards Lord Kelvin (William Thomson), in an attempt to raise the number of elastic constants for a crystalline substance to 21, represented a molecule by a pair of centres of force. This work has been extended by Max Born and his co-workers. By using a number of centres of force as the structural element, it is possible to calculate the elastic constants with fair accuracy and to estimate also the thermal characteristics of a crystalline substance. The hypotheses upon which the structure theory is based are only rough approximations to the truth, and so the general equations of elasticity have also been derived from general mechanical and thermodynamical principles, a solid body being treated as continuous.

THEORY OF STRAIN

In the general theory of strain, let P, Q be two neighbouring points of an elastic body, and P', Q', the new positions of these points when the body is deformed. Using a system of rectangular axes to specify the positions of these points, and denoting their co-ordinates by (x, y, z), $(x+dx, y+dy, z+dz)$, (x', y', z'), $(x'+dx', y'+dy', z'+dz')$ respectively, the differences

$$u=x'-x, \quad v=y'-y, \quad w=z'-z,$$
$$u+du=x'+dx'-x-dx,$$
$$v+dv=y'+dy'-y-dy,$$
$$w+dw=z'+dz'-z-dz$$

are regarded as the *component displacements* of the points respectively.

If now ds represents the small distance PQ, and ds' the small distance $P'Q'$, the ratio of $ds'-ds$ to ds is defined to be the *strain* of the element ds. In the mathematical theory, a limiting process is supposed to be carried out in which the length PQ is made indefinitely small, and the strain is defined in very much the same way as a differential coefficient. The strain is then associated with the point P, but depends upon the direction of the element PQ. Indeed, if the space occupied by the body is assumed to be Euclidean before and after deformation, the relations

$$ds'^2=dx'^2+dy'^2+dz'^2, \quad ds^2=dx^2+dy^2+dz^2$$

give rise to a relation of type

$$ds'^2=ds^2+2(adx^2+bdy^2+cdz^2+2fdydz+2gdzdx+2hdxdy)$$

in which the coefficients a, b, c, $2f$, $2g$, $2h$ depend on the partial derivatives of the displacements u, v, w. These coefficients are called the six components of strain, and may be rewritten as s_1, s_2, s_3, s_4, s_5, s_6, or as e_{xx}, e_{yy}, e_{zz}, e_{yz}, e_{zx}, e_{xy}, according to circumstances, one notation being sometimes more convenient than another.

The six components of strain are connected by certain relations which imply that the expression for ds'^2 is suitable for the representation of the square of the element of length in Euclidean space (x, y, z, are regarded now as generalized co-ordinates). When the squares and products of the partial derivatives are so small in comparison with the derivatives themselves that they may be neglected, the expressions for the components of strain may be written in the form:—

$$a=\frac{\partial u}{\partial x} \qquad 2f=\frac{\partial w}{\partial y}+\frac{\partial v}{\partial z}$$
$$b=\frac{\partial v}{\partial y} \qquad 2g=\frac{\partial u}{\partial z}+\frac{\partial w}{\partial x}$$
$$c=\frac{\partial w}{\partial z} \qquad 2h=\frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}$$

and the relations are easily found. These are called the *conditions of compatibility.*

The strain represented by a single component of type a, the other components being zero, is a *simple extension* parallel to the axis of x. The strain represented by the single component $2f$, the other components being zero, is called a *shearing strain.* When f is constant this shearing strain may be regarded as made up of two simple shears for which the corresponding displacements are, respectively,

$$u=0 \qquad v=0 \qquad w=fy$$
$$u=0 \qquad v=fz \qquad w=0$$

The displacement in a simple shear may be regarded as composed of two parts; viz., a rotational displacement,

$$u=0, \qquad v=-kz, \qquad w=ky, \qquad (2k=f)$$

for which there is no corresponding strain, and the displacement

$$u=0, \qquad v=kz, \qquad w=ky$$

for which the three components of rotation defined by the equations

$$2\xi=\frac{\partial w}{\partial y}-\frac{\partial v}{\partial z} \qquad 2\eta=\frac{\partial u}{\partial z}-\frac{\partial w}{\partial x} \qquad 2\zeta=\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}$$

are all zero. This resolution of a displacement into a rotation and an irrotational part has an analogue in the case when the strain is not constant, for we may write

$$du=(adx+hdy+gdz)+(\eta dz-\zeta dy)$$
$$dv=(hdx+bdy+fdz)+(\xi dx-\xi dy)$$
$$dw=(gdx+fdy+cdz)+(\xi dy-\eta dx)$$

Invariants.—When the axes of co-ordinates are changed, the six components of strain are changed also, but certain combinations of them remain unaltered in form. These invariants may be obtained by remarking that the ratio $ds':ds$ for any direction through P is inversely proportional to the central radius vector, in that direction, of the quadric surface whose equation in rectangular co-ordinates is

$$x^2+y^2+z^2+2(ax^2+by^2+cz^2+2fyz+2gzx+2hxy)=1$$

the lengths of the principal axes of this quadric are consequently invariants. The principal axes of the quadric are called the principal axes of strain for the point P, and the extensions of a linear

element for these directions are called the principal extensions $\epsilon_1, \epsilon_2, \epsilon_3$. The values of $1+\epsilon_1$, $1+\epsilon_2$, $1+\epsilon_3$ are the positive square roots of the three values of k for which

$$\begin{vmatrix} 1+2a-k & 2h & 2g \\ 2h & 1+2b-k & 2f \\ 2g & 2f & 1+2c-k \end{vmatrix} = 0$$

The coefficients of the different powers of $\frac{1-k}{2}$ in this equation are naturally invariants. These invariants are respectively

$$I_1 = a+b+c$$
$$I_2 = bc-f^2+ca-g^2+ab-h^2$$
$$I_3 = abc+2fgh-af^2-bg^2-ch^2$$

An invariant expressible in terms of these is obtained by putting $k=0$ in the determinant. This invariant may be written, in an alternative form, as the square of a Jacobian

$$J^2 = \left(\frac{\partial(x',y',z')}{\partial(x,y,z)}\right)^2 = (1+\Delta)^2$$

where Δ is a quantity called the *cubical dilatation*. When squares and products of the derivatives of u, v and w are neglected, the formula for the dilatation becomes simply

$$\Delta = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}$$

The quantity J represents the ratio of the volume of a small portion of the body after and before deformation. When the approximate expressions for the strains are adopted, it is clear that there is no change of volume in the shearing strain for which f is the only component different from zero.

The Strain Energy Function.—In 1837 George Green obtained a generalization of Hooke's law, suitable for the most general type of crystalline substance, by starting from the assumption that when the components of strain in a deformed body are defined by considering displacements from an initial state in which the body is free from strain, and consequently also free from stress, and when the components are also small, there is a strain energy function W which can be supposed to represent the density of strain energy at each point (x, y, z) of the body. In 1855 Lord Kelvin gave a thermodynamical argument which made Green's hypothesis seem plausible in two cases, first when the deformation takes place at a constant temperature and secondly when it takes place without loss of heat. Green assumed further that there is in general a relation of type

$$2W = \sum_{m=1}^{6}\sum_{n=1}^{6} c_{mn}\, s_m\, s_n$$

where the coefficients c_{mn} are elastic constants characteristic of the material. For convenience it will be assumed that $c_{nm}=c_{mn}$; this may be done without loss of generality because the sum of these two quantities occurs as the coefficient of $s_m\, s_n$ in the complete expression for W.

Lord Kelvin's thermodynamical argument indicates that it is necessary to distinguish between two sets of elastic constants, the *isothermal* and the *adiabatic;* it also shows that the quantities

$$S_n = \frac{\partial W}{\partial s} = \sum_{m=1}^{6} c_{mn}\, s_m$$

specify the state of stress in the body. The relation

$$2W = s_1S_1+s_2S_2+s_3S_3+s_4S_4+s_5S_5+s_6S_6$$

is, in fact, a natural extension of the formula $2W=sS$ already obtained for the case in which there is only one component of stress which is constant throughout the body and of the nature of a simple tension.

The quantities $S_1, S_2, S_3, S_4, S_5, S_6$ may be regarded provisionally as the *six components of stress*. They are usually denoted by another set of symbols such as X_x, Y_y, Z_z, Y_z, Z_x, X_y, and this notation is completed by introducing quantities Z_y, X_z, Y_x defined by the relations $Z_y=Y_z$, $X_z=Z_x$, $Y_x=X_y$. The general expression for W is suitable for a crystalline body or a substance like wood, whose physical properties are related to certain definite directions in the material. For an isotropic substance, *i.e.*, a substance in which all directions are alike as far as structure is concerned, the expression for W should have the same form in all systems of rectangular co-ordinates, and so should depend only on the invariants I_1, I_2, I_3. The natural assumption is thus

$$2W = (\lambda+2\mu)I_1^2 - 4\mu I_2$$

where λ and μ are elastic constants. The corresponding relations between stress and strain are now of type

$$X_x = \lambda\Delta + 2\mu a, \qquad Y_z = 2\mu f$$

where now $\Delta = a+b+c$.

The relations between stress and strain may be solved for the components of strain giving, in particular,

$$Ea = X_x - \sigma(Y_y+Z_z)$$

where $$E = \frac{\mu(3\lambda+2\mu)}{\lambda+\mu}, \qquad \sigma = \frac{\lambda}{2(\lambda+\mu)}$$

In these equations E is *Young's modulus*, σ is *Poisson's ratio*, μ is the *modulus of rigidity*, and the quantity $k=\lambda+\frac{2}{3}\mu$ is the *modulus of compression*. Under ordinary conditions the adiabatic constants for a solid substance are practically the same as the corresponding isothermal constants.

It must be remembered that considerable differences are found in the elastic constants of different samples of nominally the same substance. The assumption that a metal is homogeneous and isotropic is not very accurate, for modern metallurgical studies have shown that metals which have been subjected to certain treatments may be composed of long narrow crystals which impart to a specimen a kind of fibrous structure. Much work has indeed been done recently on the strength and elastic properties of single crystals, and these are found to differ considerably from those of a solid composed of an aggregate of small crystals and an amorphous substance which binds them together. The fibrous structure of wood is well known and several elastic constants are really needed to describe its properties, and these depend largely upon the amount of moisture in the wood. The value of E generally used corresponds to tension along the grain. *See* also CRYSTALLOGRAPHY; CRYSTALS, DISLOCATION OF.

Elastic Constants
(E is expressed in units of $10^6\pi$)

Substance	E	σ
Wrought iron	27–29	.28
Mild steel	29–31	
Cast iron	14–23	.28
Low carbon steel	28	
High carbon steel	27	
Aluminum (cast)	12.5	.34
Aluminum (wire)	18	
Duralumin	9.7–10.7	
Brass	12	.33
Copper	15	.35
Glass	8–10.6	.23
Rubber	250–350π	.5
Spruce (15% moisture)	1.3	
Ash	1.46	
Oak	1.4	
Walnut	1.3–1.5	
Birch	1.8–1.95	
Mahogany	1.3–1.4	
Hickory	1.9	
Celluloid (nitrocellulose)	0.355	.41

Effect of Changes of Temperature and Pressure on the Elastic Constants.—The effect of changes in temperature on Young's modulus of metals was investigated by H. L. Dodge (*Phys. Rev.*, 1913–18). It was found that E generally decreases as the temperature T increases and at an increasing rate. Widder (*Phys., Z.*, 1931) proposed the formula

$$\log\,(E/A) = B(T_0-T)$$

where T_0 is the temperature of the melting point and A and B are constants. Theoretical studies of the dependence have been

made by L. Brillouin (*Indian Acad. Sci.*, 1931–38).

The effect of pressure on the elastic properties of rocks and minerals has been studied by L. H. Adams and E. D. Williamson (*J. Franklin Inst.*, 1923).

EQUATIONS OF EQUILIBRIUM

From the Principle of Least Work.—The conditions for the equilibrium of an elastic solid deformed by given forces may be obtained from an earlier form of the principle of least work. Assuming, for simplicity, that the body is deformed only by surface tractions and that the displacements of points on the surface are prescribed, the internal displacements are supposed to be of such a nature that the total strain energy of the body is a minimum when the energy associated with the actual deformation is compared with the energy of a fictitious deformation consistent with the same surface displacements.

Giving arbitrary small increments to the component displacements u,v,w of a point (x,y,z), and assuming that the strains are so small that their expressions (1) may be used in the expression for the total energy, the Eulerian equations of the Calculus of Variations give in this case

$$L \equiv \frac{\partial X_x}{\partial x} + \frac{\partial X_y}{\partial y} + \frac{\partial X_z}{\partial z} = 0$$
$$M \equiv \frac{\partial Y_x}{\partial x} + \frac{\partial Y_y}{\partial y} + \frac{\partial Y_z}{\partial z} = 0$$
$$N \equiv \frac{\partial Z_x}{\partial x} + \frac{\partial Z_y}{\partial y} + \frac{\partial Z_z}{\partial z} = 0$$

These will be regarded as the differential equations for the equilibrium of an element of the body which occupies the position (x,y,z).

The first equation gives

$$\int\int\int L dx dy dz = 0$$

where the integration extends over an arbitrary portion of the body. Transforming this volume integral into a surface integral by the usual method, the resulting equation may be written in the form

$$\int\int (lX_x + mX_y + nX_z) dS = 0$$

where (l,m,n) are the direction cosines of the normal to the surface-element dS. The quantity under the integral sign is now interpreted to be the x-component of the surface traction acting across the surface element dS, while the quantity within brackets is regarded as the x-component of the stress across this element. Denoting this quantity by X, the preceding equation may be written in the form

$$\int\int X_\nu dS = 0$$

and there are similar equations involving the y and z-components $Y_\nu Z_\nu$. These may be regarded as three of the conditions of equilibrium of the surface tractions on the elements of the closed surface S. The other conditions for the equilibrium of these surface tractions are obtained by taking moments about the axes of co-ordinates; they are

$$\int\int (yZ_\nu - zY_\nu) dS = 0 \qquad \int\int (zX_\nu - xZ_\nu) dS = 0$$
$$\int\int (xY_\nu - yX_\nu) dS = 0$$

Transforming the surface integrals into volume integrals, it is found that the three equations are satisfied in virtue of the differential equations of equilibrium and the three equations

$$Y_z = Z_y \qquad Z_x = X_z \qquad X_y = Y_x$$

which were introduced to complete the notation.

With the present definition of stress, the stress components across an area perpendicular to the axis of x are X_x, Y_x, Z_x, while those across an area perpendicular to the axis of y are X_y, Y_y, Z_y. The equation $X_y = Y_x$ is interpreted as a relation between complementary shearing stresses. The usual convention in interpreting X_ν is that it is a stress exerted by the part of the body towards which the normal is drawn. It is clear from the definition that the component stress in the direction (l', m', n') across a plane at right angles to (l, m, n) is

$$l'(lX_x + mX_y + nX_z) + m'(lY_x + mY_y + nY_z) + n'(lZ_x + mZ_y + nZ_z)$$

The state of stress at a point (x,y,z) is thus completely specified by the six components of stress.

Body Forces.—When an elastic solid is acted upon by a body force like gravity, which acts right through the body, the equations of equilibrium take a more general form, and are of type

$$L + \rho X = 0$$

where X is the x-component of the body force per unit mass. When the body is in motion the force of inertia may be treated as an additional body force, and the equations of motion are of the type

$$L + \rho X = \rho \frac{\partial^2 u}{\partial t^2}$$

These equations are particularly useful for a study of the vibrations of solid bodies and the propagation of waves through them.

It was shown by Poisson that there are two distinct velocities with which waves can be propagated without change of type through a homogeneous isotropic elastic solid. The so-called irrotational waves (or waves of compression) travel with velocity $\sqrt{(\lambda + 2\mu)}/\sqrt{\rho}$, while the equivoluminal waves (or waves of shear) travel with velocity $\sqrt{\mu}/\sqrt{\rho}$. Lord Rayleigh (John William Strutt, 3rd baron) discovered that a type of surface wave, compounded in a rather complex way from waves of the two types, can travel with a velocity slightly less than that of the equivoluminal waves. The vibrations of elastic solids are of great interest in the theory of sound and in the design of engineering structures (*see* SOUND: *Sources of Sound*).

Thermoelasticity.—A theory of thermal stresses was given by J. C. M. Duhamel about 1838, terms proportional to the change of temperature being introduced into the expressions for the strains in terms of the stresses. The theory was applied first to cylinders and spheres strained radially but has been extended to some two-dimensional and three-dimensional problems. The stresses in the earth produced by sinusoidal fluctuations of temperature have been discussed by some writers (for example H. Jeffreys, *R. Soc. London*, 1921–22; H. Arakawa, *Geophys. Mag.*, Tokyo, 1932). The corresponding stresses in walls and slabs have been discussed by N. S. Boulton (*Phil. Mag.*, 1933). The buckling of a thin circular plate by heat has been considered by F. E. Relton (*Phil. Mag.*, 1928). The transient thermal stresses in concrete cylinders have been discussed by G. Pickett (*Am. Soc. Testing Materials*, 1939). The thermal stresses in the piston heads of internal-combustion engines are discussed in books on the subject (*e.g.*, L. O. Lichty, 1939). The thermal stresses in the barrel and fins were considered by A. M. Binnie (*Phil. Mag.*, 1926). The thermal stresses in cylindrical tubes and pipes have been discussed by many writers. Reference may be made to the work of C. H. Lees (*R. Soc. London*, 1922), C. H. Kent (*Trans. Am. Soc. Mech. Eng.*, 1931) and H. Poritsky (*Phil. Mag.*, 1937).

PROBLEMS IN THEORY

Some of the classical problems relating to the deformation of beams and columns lead to ordinary differential equations and can be regarded as one-dimensional problems. As architecture, shipbuilding and bridge construction progressed, the flexure of continuous beams with many points of support needed to be studied and an equation of three moments was developed for the treatment of such problems. In modern times this equation has been extended to the case in which the beam is under the influence of end thrust or tension. Problems relating to the vibration of beams and columns have become prominent and the classical problem of the stability of a column has led to many analogous investigations.

The problems leading to a partial differential equation in two variables form another important class in which much progress has been made. In the problem of the deflection of a loaded plate Poisson's equation is of the form $B\nabla w = P$, where B is the biharmonic operator often written as ∇^4 and P is generally a function of the two variables giving the distribution of load. The solution of this equation can be built up by integration from the solution of $B\nabla w = 0$ that corresponds to the case of a single concentrated load and the specified boundary conditions. This same partial

differential equation $Bw=0$ occurs in the problems of plane stress which are particularly interesting because very often the correctness of a solution can be checked by photoelastic methods. The problem of a tension member with an elliptic hole or inclusion has been much discussed, estimates being obtained for the concentration of stress. The problem of a notched plate under tension is of similar interest. Among the other problems of plane elastic theory those relating to dams, loaded plane surfaces, rectangular members in shear and annular members have been much studied.

Saint Venant's theory of the torsion of a prism likewise leads to a partial differential equation in two independent variables and solution of the problem has been found for prisms of various cross section. The case of a shaft with a keyway is particularly interesting. The problem of a shaft of varying section also leads to a partial differential equation in two variables when the section is circular and solutions have been found for some laws of variation of section.

Saint Venant's theory of the bending of a beam of constant section and the problem of a beam rotating about a longitudinal axis furnish additional cases in which the treatment depends on partial differential equations in two variables and much progress can be made. When torsion and flexure are combined the theory is more difficult. The related problem of the vibrations of a rotating shaft is of much importance, and is generally solved by numerical methods in which changes in section are lumped and the vibration problem is solved by means of difference equations. The word "whirling" is used to refer to a fundamental type of flexural vibration of the rotating shaft, while the word "whip" is applied to any one of the higher flexural modes of vibration.

Among the three-dimensional problems which have been much discussed are the statics and vibrations of plates and shells, pipe bends and helical springs. The stresses in hooks, links, rings and spoked wheels have been much studied and innumerable problems have arisen in connection with the design of airplanes. The introduction of light metals, plywood and plastics has brought many new problems. The rolling of metals has been discussed mathematically.

The heat treatment of metals also introduces new problems such as those of initial strain. There is the important problem of the shrink fit and it has been found that initial stress must be considered in problems of elastic instability and in the propagation of elastic waves. Elastic instability and wave propagation are of much importance in geophysics. Some information regarding the internal structure of the earth has been derived from measurements of the travel times of earthquake waves. Further information has been derived from measurement of the elastic properties of the substances forming the earth's crust and from pendulums used in various ways and locations. The results of gravimetric and seismologic observations agree in fixing the average value of the rigidity of the earth at a value close to that of steel. This estimate has been confirmed by the experiments of Albert A. Michelson (*Astrophys. J.*, 1914) on earth tides by the interference method. The aim was then to find the changes in direction of the gravitational vertical as modified by the distortion of the earth and the plasticity of its material. Preliminary values found were: rigidity $=8.6\times10^{11}$ c.g.s. units; viscosity $=10.9\times10^{16}$ c.g.s. units.

Theories of Elastic Afterworking, Creep and Lag.—In phenomena of hysteresis the behaviour of the substance depends upon the stress at times previous to the one under consideration. In 1826 Simeon Denis Poisson made use, in the theory of magnetism, of a type of integral depending on the past history of the substance; and in 1874 Ludwig Boltzmann adopted a somewhat similar idea to account for the phenomena of elastic afterworking. Like Poisson he used the principle of the superposition of effects, the validity of which has been questioned when efforts are made to account for the effects of cyclic variations of load. Most theories of elastic afterworking can be reduced to Boltzmann's by a suitable choice of the "afterworking function" that appears in the integral.

In the work of James Clerk Maxwell (1868) use was made of a time of relaxation τ and the residual stress S in a specimen at time t was given by an equation of type $S=S_0\,e^{-t/\tau}$. In the theories of Sir Joseph Thomson (1888) and E. Wiechert (1889) several states of variation with different times of relaxation τ were supposed to occur simultaneously. The deformation was then represented by a sum of terms of the exponential type, the exponents being linear functions of the time.

Attempts have been made by H. V. Wartenberg (*Deutsch Phys. Ges.*, 1918), G. Masing (*ibid.* 1919), and others to base a theory of elastic afterworking and plastic deformation on the crystalline structure of metals. In the theory of the first two of these writers the angle of twist ϕ is supposed to vary with the time according to a law of type

$$\tau.d\phi/dt=A-\phi$$

where A is a constant residual angle of twist which is gradually approached. The wire is supposed to be composed of a large number of crystals some of which are plastically deformed even when the stress is below the elastic limit for the whole piece. In support of this view Wartenberg cited his experiments with single-crystal zinc wires in which there was apparently no trace of elastic afterworking.

L. Prandtl (*Z. Angew. Math. Mech.*, 1928) starts with similar ideas and uses a working model consisting of two lattices with parallel plane faces and particles in between which are more strongly bound to one lattice than the other and have various positions of stable and unstable equilibrium. His analysis of a simple case leads eventually to a hyperbolic sine law for the creep or velocity of flow v

$$v=v_0 sinh(S/S_0)$$

where the factor v_0 is different for different "hardness" of the model. This hyperbolic sine law was proposed also by A. Nadai and seems to occur also in the chemical reaction rate theory of plastic flow proposed by H. Eyring (*J. Chem. Phys.*, 1936). In the theory of L. H. Donnell (*Trans. Am. Soc. Mech. Eng.*, 1942) plastic flow is regarded as an unstable process. The mode of growth of the plastic region is discussed. In the theory of internal friction in solids developed by C. Zener (*Phys. Rev.*, 1937) inhomogeneities in stress are supposed to give rise to fluctuations in temperature and to local currents of heat which increase the entropy of the substance and are a cause of internal friction. This is considered with good reason to be the chief cause of internal friction. In the work of B. P. Haigh (*Trans. Faraday Soc.*, 1924) the inclusion of foreign materials such as slag is considered as a possible cause of fatigue, and N. O. Myklestad (*Trans. Am. Soc. Mech. Eng.*, 1942) has calculated the thermal stresses for needle-shaped and disk-shaped inclusions, finding that for a given temperature variation these forms give rise to higher temperature stresses than a spherical inclusion.

The rolling of metals is a much discussed subject in which plasticity plays an important part and the effects of temperature need consideration. Discussions have been given by E. Orowan (*Inst. Mech. Eng.*, London, 1943) and B. F. Decker (*Am. Soc. Testing Materials*, 1943).

A mathematical treatment of some problems has been based on a condition of plasticity derived from the theory that any definite volume of the material has only a limited capacity for absorbing and storing energy in the plastic form (*see* MATERIALS, STRENGTH OF).

Mention may be made of the work of W. Prager (1932–33) on the spreading of the plastic region and the theory of Lüders' lines which is sketched in the discussion of Donnell's paper (*Am. Soc. Mech. Eng.*, 1943).

Many writers are content with empirical expressions for the rate of creep. Rather complicated expressions were given by A. A. Michelson (*Nat. Acad. Sci.*, 1917) for the elastic viscous displacement, the viscous displacement that varies with time and the lost motion. L. H. Adams and E. D. Williamson (*Optical Soc. America*, 1920) gave a law

$$1/S-1/S_0=At$$

for the relation of the stress S at time t to the initial stress S. An

empirical expression for deformation involving simple powers of the time and force was proposed by P. G. Nutting in 1921 and in the plotting of the results of creep tests it is customary to use logarithmic scales so that the points lie nearly on portions of straight lines. For the relation to the hyperbolic sine law reference may be made to the paper of A. Nadai and P. G. McVetty (*Am. Soc. Testing Materials*, 1943).

Much work has been done recently on elastic theory for large rotations, displacements or strains. Reference may be made to the work of F. D. Murnaghan (*Bologna Math. Congress*, 1928; *Nat. Acad. Sci.*, 1944); M. A. Biot and K. Maguerre (*Intern. Congr. Appl. Mech.*, 1938); G. E. Hay (*Am. Math. Soc. Trans.*, 1942); R. A. Fairthorne (*Phil. Mag.*, 1938); and B. R. Seth (*Indian Acad. Sci.*, 1941). A nonlinear theory of elasticity has been developed by M. A. Biot (*Phil. Mag.*, 1939). *See* also MECHANICS; METALLURGY; SOLID STATE PHYSICS.

BIBLIOGRAPHY.—E. G. Coker and L. N. G. Filon, *A Treatise on Photoelasticity* (1931); M. M. Frocht, *Photoelasticity* (1941); B. Gutenberg, *Internal Constitution of the Earth*, National Research Council (1939); R. Houwink, *Elasticity, Plasticity and Structure of Matter* (1937); A. L. Kimball, *Vibration in Engineering* (1932); A. E. H. Love, *The Mathematical Theory of Elasticity* (1927); A. Nadai, *Plasticity* (1931); J. F. Nye, *Physical Properties of Crystals: Their Representation by Tensors and Matrices* (1957); J. Prescott, *Applied Elasticity* (1924); R. V. Southwell, *Introduction to Theory of Elasticity for Engineers and Physicists* (1941), *Relaxation Methods in Engineering Science* (1940); S. Timoshenko, *Theory of Elasticity* (1934), *Theory of Elastic Stability* (1936), *Theory of Plates and Shells* (1940); *Vibration Problems in Engineering*, 3rd ed. (1955); W. Ker Wilson, *Practical Solution of Torsional Vibration Problems* (1940). (H. BN.)

ELATERITE, also termed elastic bitumen and mineral caoutchouc, an asphaltic pyrobitumen originally discovered in the Odin lead mines at Castleton, Derbyshire. It occurs in a few other localities, but always in small quantities, and is of merely scientific interest. A natural hydrocarbon, it varies somewhat in consistency but is typically soft and elastic, often closely resembling India rubber, and is a derivative of petroleum. A substance of similar physical character is found in the Coorong district of South Australia and is known as coorongite, but some doubt exists as to whether this is a derivative of petroleum or a vegetable product. *See* also BITUMEN.

ELAZIG, the chief town of the *il* (province) of Elazig in the southern part of eastern Anatolia, Turk., is situated at the foot of a plateau overlooking a fertile plain, 55 mi. E.N.E. of Malatya. Pop. (1960) 60,438. It is one of the newer urban centres of Turkey and was founded in the mid-19th century as a military garrison and administrative centre. The town developed rapidly as a result of its favourable position and later became the capital of the old province of Harput. Once famous as both a cultural and an administrative centre, the town of Harput lost its importance and its population to become a commune of Elazig. The town of Elazig is on the railway between Adana and Lake Van, and is an important road junction.

ELAZIG IL lies between Diyarbakir *il* to the south and Tunceli *il* to the north. Area 3,574 sq.mi. Pop. (1960) 278,106. It is bounded on the north by the Peri river and on the west and southwest by the Euphrates (Firat); the Murat crosses from the east to join the Peri. Lake Hazar lies about 13 mi. S.E. of Elazig town. The climate is continental with severe winters. The chief economic activities of the area are agriculture and stock raising. Chromium is mined at Guleman and copper at Maden.
(N. TU.; S. ER.; E. TU.)

ELBA, an island off the west coast of Italy, is part of Leghorn (Livorno) province in the region of Tuscany and lies 7 mi. S.W. of Piombino on the mainland. Pop. (1951) 27,877. Area 86 sq.mi. The eastern part consists largely of mica schists, gneiss and limestone, with important deposits of the iron ores magnetite and limonite; the centre is of diabase Eocenic serpentine, porphyry and granite; the west is made up of the massif of Monte Capanne, the highest point (3,343 ft.), with scattered finds of tourmaline, beryl and garnet.

The mild climate supports a varied vegetation of Mediterranean type with rich olive groves and vineyards. There are important fisheries of anchovies, sardines and tunny. The iron ore of Elba was smelted by the Etruscans at Populonia, their nearest mainland settlement. The Greeks called the island Aethalia, "the smoky place," probably because of the smelting furnaces. The Romans quarried granite and mined iron ore on Elba, which they knew as Ilva. In modern times the ore is mined in four places on the east coast, some being smelted on the island and some shipped to Porto Baratti on the mainland. By the early 1960s the reserves of accessible ore were nearing exhaustion and it seemed unlikely that the deeper deposits would repay the increased costs of mining.

During the middle ages Elba was ruled by Pisa. In 1399 it passed to the dukes of Piombino who in 1548 ceded it to Cosimo I of Florence. From 1596 until 1709 in Spanish hands, Elba was next ruled by Naples. In 1802 it was given to France, and when in 1814 Napoleon I abdicated he was exiled to Elba. The island was recognized as an independent principality with Napoleon as ruler until Feb. 26, 1815, when he returned to France for the Hundred Days. After his fall Elba was restored to Tuscany, with which it passed to Italy in 1860.

The chief town is Portoferraio (pop. [1957 est.] 10,230) situated on a bay on the north coast, with a harbour built and fortified by Cosimo. It has blast furnaces and coke ovens. Napoleon's chief residence, the Palazzina dei Mulini, overlooks the sea; his summer resort, Villa San Martino, lies 6 km. (4 mi.) S.W. and contains a museum and collection of paintings. With the decline of the iron industry, tourism has assumed increasing importance. Popular summer resorts are Biódola, Procchio, Marciana Marina and Poggio di Marciana in the north, Marina di Campo on the south coast and Porto Azzurro with its great Spanish fort (1602) facing the mainland. Elba has bus services and is connected by motor-vessel services with Portovecchio di Piombino and with Leghorn. (G. KH.)

ELBASAN, a town in central Albania and capital of Elbasan *rreth* (district), lies on the northern bank of the Shkumbi river, at the eastern end of a rich, well-watered plain. Pop. (1955) 23,616. Until World War II about 85% were Muslim and the rest Albanian Orthodox, with a few Uniate. The town is on the old Roman Egnatian road, now the main Durres-Korce motor road, and it is also the terminus of the railway from Durres. The streets are largely narrow and cobbled, with picturesque bazaars; little remains of the old walls except for one well-preserved archway and a tower. Olives, maize and tobacco are the chief local products, and there are olive oil, soap and timber factories. A few miles west is the new petroleum refinery of Cerrik, opened in 1957. Elbasan was one of the principal centres of Albanian nationalism under the Ottoman empire, and it is noted for the purity of its language, which links the northern Geg (Gheg) and the southern Tosk dialects. (D. R. O.-H.)

ELBE, a major European river, is the longest and most important navigable waterway in eastern Germany. Its upper drainage basin is in the Bohemian massif of Czechoslovakia; from there it flows into eastern Germany, and between Wittenberge and Lauenburg the boundary between the Federal Republic of Germany and the German Democratic Republic lies in midstream.

The Elbe rises on the southern side of the Giant mountains near the German-Czechoslovak frontier. It flows south and west forming a wide arc for about 225 mi. in Czechoslovakia (where it is known as the Labe) to its confluence with the Vltava (Moldau) and is joined 18 mi. downstream by the Ohre (Eger). It then cuts northwest through the Ore mountains, in a gorge 4 mi. long, into Germany. Between Dresden and Magdeburg it receives many long tributaries, of which all except the Schwarze Elster are left-bank streams. These are the Mulde, the Saale and its tributaries the Weisse Elster and the Pleisse, the Unstrut and the Ilm. These left-bank tributaries rise in the Ore mountains or the Thüringer Wald and form the drainage basin of the middle Elbe with its geographic foci in Halle and Leipzig. Halle is on the Saale, just below the confluence of the Pleisse; Leipzig lies along the Pleisse and the Elster which flow close and parallel to each other on the west side of the city. Below Magdeburg, the Elbe receives most of its water from its right bank. Its tributaries, including the Havel, rise in the uplands of Mecklenburg.

The river enters the north German plain at Riesa 25 mi. below

Dresden; below Riesa it meanders in a wide flood plain and has some abandoned loops. Dikes begin there and continue as far as the confluence of the Mulde. Between Wittenberg and Dessau the east-west valley floor narrows to 5 mi. in width and hilly land rises to the north and south. From Dessau to Magdeburg the flood plain widens and dikes have been constructed continuously down to the sea. In its course below Magdeburg, between the Lüneburg and the Fläming heaths, the flood plain is 2 mi. wide down to the confluence with the Havel. The river keeps to the left of its flood plain and sometimes cuts into the low hills on its banks. Below the confluence with the Havel the river flows southeast-northwest; the flood plain widens and has distributaries and backwaters often flanked by low sandy hills (geest). Reclaimed salt marshes begin at Lauenburg. Above Hamburg, which the Elbe transverses in two arms, the Norder Elbe and the Süder Elbe, the flood plain is 8 mi. wide but narrows to 4 mi. between the sandy geest of Schleswig-Holstein and the Lüneburg heath. From Hamburg to Cuxhaven (56 mi.) is the Unterelbe or estuary proper. It varies in width from 1 to 2 mi. but much of it is occupied by mud flats and sandbanks. The main channel is buoyed and dredged. Vessels drawing 36 ft. can normally reach Hamburg on the flood tide. The south or left bank is low and marshy and the river has sandbanks; the right bank is steep below Hamburg but farther downstream there are marshes, diked and drained and intensively cultivated. Hamburg grew up on the Alster river on low sandy hills above the marshes. The modern port has spread to the low-lying south bank of the Elbe. The tidal limit of the Elbe is at Lauenburg; the flow at low water is 5,300 cu.ft. per second, mean level 24,700 cu.ft. and the highest level 135,600 cu.ft. The tidal range at Cuxhaven averages 9.1 ft. and at Hamburg 7.5 ft.; the highest recorded range is 28.4 ft. at Cuxhaven and 27 ft. at Hamburg.

The Elbe has a total length of 723 mi. of which 347 mi. are in the German Democratic Republic. Its total drainage area is 55,620 sq.mi. The lengths and drainage areas of its main tributaries are: Schwarze Elster, length 117 mi., drainage basin 2,154 sq.mi.; Mulde and Zwickauer-Mulde, 156 mi. and 2,780 sq.mi.; Saale 265 mi. and 9,073 sq.mi.; Unstrut 117 mi. and 2,452 sq.mi.; Weisse Elster 153 mi. and 1,969 sq.mi.

There are 15 dams on the German Elbe and its tributaries. The largest are on the Saale at Bleiloch and at Hohenwarthe.

The regime of the Elbe has a strong seasonal variation. Its upper drainage basin in Bohemia receives precipitation in winter, but relatively dry and hot conditions prevail in summer, with occasional heavy, concentrated falls of rain and a high evaporation rate. On leaving the Bohemian massif the river contains 28% of the total precipitation of its catchment area. High level occurs in February, March and April (with the melting of the snows); minimum level is in midsummer (August–September) with occasional sudden rises after summer storms. This regime remains characteristic of the river throughout its course, although the summer minimum is not quite so low. Additional supplies are brought from the left-bank tributaries from the well-watered Ore mountains in late spring. The river loses much water, however, in its wide flood plain through evaporation and infiltration. The stretch in Saxony, above the confluence of the Mulde and Saale, has the least depth of water and can present difficulties to navigation in late summer.

The mouth of the Elbe, a little below Cuxhaven, has a width of about 9.3 mi. The seaward approaches to the estuary contain extensive sandbanks and shoals, some of which form sandy islands. Channels are marked by buoys and light vessels. Above Cuxhaven, the estuary narrows in a funnel shape and at the entrance to the Kiel canal at Brunsbüttel it is 1¼ mi. wide.

Commerce.—The Elbe is navigable to Usti nad Labem (Aussig) in Bohemia for barges with a capacity up to 700 tons, and for smaller barges to Prague. The river is connected with the Weser and the Rhine via the Mittelland canal and its continuation, the Havel, below Magdeburg, each taking 1,000-ton barges. The Saale is navigable upstream to Naumburg for 200-ton barges, and for 1,000-ton barges as far as Trotha (below Halle). Ice hinders traffic in winter; 28 days are averaged for fast ice at Dresden and 15 at Hamburg; the extreme limits for fast ice are Nov. 10–March 18, and Nov. 17–March 9 respectively. The estuary is open through the winter though sometimes choked by ice. Incidence of fog is high (80 days October–February), as is that of seafloods in the Deutsche Bucht, at the estuary entrance.

The Elbe carries a great volume and variety of traffic, for most of which Hamburg in the Federal Republic of Germany is the natural terminus. Commerce, however, was disrupted after the partition of Germany following World War II when east-west traffic was reduced. In 1938 Hamburg handled by river about 10,000,000 tons of coal, coke, petroleum, wheat, timber, iron ore, fertilizers and foodstuffs. In that year the section from Hamburg to Lauenburg (31 mi.) handled 8,800,000 tons, the middle stream to Dessau (192 mi.) 6,000,000 tons, and the upper course to the Czech frontier 1,000,000 tons.

Hamburg is also a port of transshipment from ocean-going vessels to river barges. By the mid-20th century its distribution of trade with the hinterland had changed considerably. In 1914 about 80% of its inland trade was by water and 20% by rail; in 1938 water transport accounted for only 40%. Moreover, in 1938 38% of this inland trade was with the territories now in the German Democratic Republic and 20% with eastern Europe. Thus, more than 50% of its trade was with territories now in the Soviet bloc. Ten years after World War II less than 20% of inland trade was by water, the bulk being transported by rail and road. Hamburg's trade with eastern Europe was only about 20% of its pre-World War II level.

In the late 1950s more than 14,000,000 gross tons were carried annually on inland waterways in East Germany according to official data. The principal cargoes included coal and coke, metals, fertilizers, building materials, grains and seeds. There was also a great diversity of small consignments. This large and diverse traffic was concentrated at Berlin (Osthafen), Magdeburg, Halle and Dresden and revolves around the production of the middle Elbe basin, south of Magdeburg, an important centre of industry and agriculture. Heavy industries are situated around Halle, Leipzig and in lower Lusatia (lignite mining, chemical and thermal electrical production). Saxony, on the slopes of the Ore mountains, is highly industrialized, with textile and machine industries based on electricity derived from the lignite field to the north.

The chief cities actually on the Elbe are Dresden, Magdeburg and Hamburg. Major cities of the middle Elbe are Leipzig, Halle and Dessau; and in Saxony, Karl-Marx-Stadt (Chemnitz) and Zwickau.

Historical Importance.—The Elbe-Saale line has played a significant part in the history of the north European lowland. It marked in the 8th century the limit of Charlemagne's conquests, and later, especially under the Saxon dynasty, became the line of departure for the German conquest of Slavonic lands. After World War I the Elbe was internationalized by the treaty of Versailles but this was subsequently repudiated by Nazi Germany. In World War II a point on the Elbe, near Torgau, was the meeting place of the U.S. and Soviet armies. Later the river formed part of the demarcation line between divided Germany. (R. E. Dr.)

ELBERFELD, a former city in the Prussian Rhine province (now in North-Rhine Westphalia *Land* of the Federal Republic of Germany), straddling the Wupper valley, noted especially for its silk and red-dyeing industries. In 1929 it was joined with Barmen and neighbouring towns to form one city, known since 1930 as Wuppertal (*q.v.*).

ELBEUF, a town of northwestern France, Seine-Maritime *département,* lies on a loop of the Seine, 19 km. (12 mi.) S.S.W. of Rouen by road. Pop. (1954) 16,900. Elbeuf is a mixture of old streets, modern quarters and factories. Part of the town had to be rebuilt after World War II. Behind Elbeuf rise wooded hills which stretch downstream through the cliffs of Orival with their grottoes and panoramic views. The principal churches are those of St. Étienne (15th–16th centuries) and St. Jean (17th century). Both contain fine stained-glass windows; one in St. Étienne (1514) shows contemporary Elbeuf drapers at work. The town hall has a natural history and archaeological museum. The educational establishments include a *lycée* and a textiles school.

Elbeuf is on the railway from Rouen to Le Mans and Tours and has a river port handling coal and other products. It is the centre of a group of industrial towns, Caudebec-lès-Elbeuf (pop. 9,336), St. Pierre-lès-Elbeuf (3,230) and Orival on the left bank of the Seine and St. Aubin-les-Elbeuf (5,289) on the right. Since the 15th century these towns have been an important centre of the woolen industry and of the manufacture of cloth and textiles whose reputation is world-wide. At Elbeuf Jean Baptiste Colbert (*q.v.*) set up a royal cloth factory in 1667 and a royal tapestry factory in 1685. Other industries include the making of automobiles.

The history of the area goes back to the Roman occupation when Caudebec-lès-Elbeuf is mentioned as Uggade. Elbeuf itself is first recorded as Wellebou in 950. Later it was known as Huellebuef. After the English occupation (1418–49) Elbeuf became part of the De Lorraine family possessions and was successively a barony, marquisate and duchy. The last duke of Elbeuf emigrated at the Revolution and died in Vienna in 1825. Elbeuf was in German hands during the Franco-German War of 1870–71 and in World War II from June 1940 to Aug. 1944. (Ch. Br.)

ELBLAG (Ger. Elbing), chief town of the Elblag district in Gdansk (Danzig) *wojewodztwo* (province), Pol., situated 56 km. (35 mi.) E.S.E. of Danzig. Pop. (1960 census) 77,000. Area 31 sq.km. (80 sq.mi.). It lies between two types of landscape, the Elblag elevation, which is part of the Masurian lakelands, and the Zulawy, an alluvial region of the Vistula delta. The oldest part of the town stands on the small Elblag river, which flows into the Vistula bay, while the more modern part lies on the slopes of the elevation. The town is a centre of the metallurgical industry. The proximity of the fertile Zulawy district has also meant development of trade in agricultural products.

Elblag arose on the site of the prosperous Prussian market settlement of Truso, which was in existence till the end of the 10th century. The Teutonic Knights built a castle there in 1237, and the settlement which developed round it obtained municipal rights in 1246 and was early admitted to the Hanseatic league. In 1454 Elblag and other Prussian towns rose against the Teutonic Knights and submitted to the king of Poland. In the 15th and 16th centuries, when the Vistula sandy spit (*mierzeja*) was still a number of separate islands, Elblag was an important commercial port on the Baltic. From 1560 to 1626 it was the seat of the English Eastland company. However, rapid silting up between the separate islands of the spit and the obstruction of Vistula bay reduced Elblag to an inland port. After the first partition of Poland (1772) Elblag was handed over to Prussia, with unfavourable effects on its economic development. During World War II about 60% of the town was destroyed. It was returned to Poland in 1945.

(K. M. Wi.)

ELBOW: *see* Joints and Ligaments.

ELBURZ MOUNTAINS (Persian Reshteh-ye Kuhha-ye Alborz), a chain of mountains in northern Iran which separates the Caspian depression from the Iranian plateau and extends for 650 mi., encircling the southwestern and southern shores of the Caspian sea, to Khurasan, where it passes into the southern of the two principal ranges of that province. The chain's width varies from less than 15 mi. at the Soviet border near Astara to 75 mi. east of Teheran. The Caspian slope is well defined, forming an imposing feature by its steepness and height; the interior edge of the chain is less clearly defined. West of Kazvin and east of Semnan its limits are undetermined. The geological structure also merges into the interior highlands.

Most conspicuous are limestones of the Upper Jurassic and Cretaceous, and Lower Jurassic shales with coal seams. Dark Permocarbonic limestones, Red Devonian sandstones, together with schists and granites also occur. Igneous rocks of different periods are common. Eocene lava flows are most extensive in the west while contemporary tufaceous greenbeds abound in the central parts. The Miocene is represented by clays, marls and conglomerates. Elevations are considerable, over 12,000 ft. in many ranges, and are distinctly alpine in character; the lower ranges generally display gentler forms such as rounded ridges, plateaus and open valleys. Most valleys are deep and narrow, with steep sides, and frequently take the form of gorges.

The Elburz may be subdivided into three sections, separated by the gorge of the Safid Rud and by the Gaduk pass (7,540 ft.) near Firuzkuh. (1) On the west, the Talish mountains (Kuhha-ye Tavalesh) run, curving slightly, in a north-northwest–south-southeast direction for 120 mi. From 15 to 25 mi. wide, they consist of one asymmetric ridge, the long slope extending to the Caspian. Altitude rarely exceeds 10,000 ft. but decreases to about 5,000 ft. east of Ardabil. The Safid Rud gorge is the only one entirely crossing the chain; it carries the combined waters of the Qizil Uzun (Qezel Owzan) and Shahrud rivers and offers, combined with the low pass of Kazvin, the best though not the easiest passage from Gilan to the interior. (2) The central section, 240 mi. long, strikes roughly west-east. With the exception of the extinct volcano Mt. Demavend (*q.v.*; 18,580 ft.), this includes the highest ranges of the whole chain on both sides of the Kendavan (Kanduan) pass (10,000 ft.) and the Chalus road. The only peaks to bear glaciers apart from Mt. Demavend are the Takht-e Soleyman group culminating in Kuh-e Alamdar (or Alam, 15,817 ft.). This section is split into several parallel ridges of varying importance by longitudinal valleys or depressions which run along both sides of the main ridge. These depressions are: on the north side, Deylaman, Eshkevar, Kelardasht, Kojur and Nur; on the south side, Shahrud, Talaqan (with the side branch of Alamut), Lar and Firuzkuh valleys, and the depression followed by the Mazanderan road east of Teheran. Important valleys run straight down from the main ridge: Karaj, Jajrud and Hableh Rud on the south; Sezar, Chalus and Horaz on the north side. (3) The eastern section extends approximately southwest-northeast for 290 mi., its width decreasing to 30 mi. as the longitudinal valleys disappear. The height falls to about 8,000 ft., except for the Shahkuh range between Gorgan and Shahrud, which rises to 12,000 ft.

The Caspian and interior slopes differ markedly in climatic and other aspects. While the interior shares the arid character of the highland, the Caspian side is exceptionally humid and the constant moisture from the sea has given it a cover of luxuriant deciduous forest. Snow is plentiful on both sides, blocking traffic until April and May, but also providing an abundant water supply. The Elburz mountains support many flourishing villages, and cultivation is extensive, wherever relief permits. Terracing is used on steep slopes, or for rice growing in the valley bottoms of the Caspian side. The villages at higher altitudes are largely abandoned in winter for the lowlands. On high levels lush summer grazing attracts both the villagers and the nomads of the plains with their cattle, sheep and goats. Some upland regions have been reserved for specific tribes.

The mineral wealth of the Elburz is small. Bituminous coal and iron ore are of some importance; two modern coal mines are worked at Shamshak and Zirab. There has always been exchange of goods between the Caspian and the interior provinces. Traffic was formerly by mule tracks. By the 1960s there were several motorable roads and the Trans-Iranian railway crossing the range. Dam projects are important. Two were constructed in the 1960s: the Safid Rud dam (Manjil) for power and irrigation in the Gilan area, and the Karaj dam for power and water for Teheran.

The population is of old Persian stock with considerable additions of Kurdish, Turkish and Arab tribes, settled or nomadic. The Elburz mountains and the Caspian lowlands have played a role apart from the interior highlands in the history of the region. They were a refuge of the Zoroastrians, the seat of the Dailamite kingdom, and a stronghold of the Ismailian assassins. (H. Bo.)

ELCANO, JUAN SEBASTIAN DE: *see* Cano, Juan Sebastián del.

EL CENTRO, a city of southeastern California, U.S., 120 mi. E. of San Diego and 58 mi. W. of Yuma, Ariz., the seat of Imperial county, is a major commercial and transportation centre for the productive irrigated desert area known as the Imperial valley (*q.v.*). The Mexican border and the Mexican city of Mexicali lie about 10 mi. to the south. Since its beginnings about 1906, El Centro has grown steadily with the development of the Imperial valley and has become a modern city of air-conditioned homes, motels, offices, stores and public buildings set in the year-round green of the irrigated desert.

Like the surrounding area, El Centro is below sea level, its altitude being −52 ft. For comparative population figures *see* table in CALIFORNIA: *Population*. (R. M. GL.)

EL CERRITO, a city of Contra Costa county, Calif., U.S., on the eastern shore of San Francisco bay, north of Oakland, and 14 mi. N.E. of San Francisco via the San Francisco-Oakland bay bridge. The early village, named Rust after a German blacksmith and storeowner who lived there, did not grow until the establishment of a railroad terminal and harbour at neighbouring Richmond (*q.v.*) in the early 1900s. The San Francisco earthquake and fire of 1906 resulted in a spurt of settlement, and thereafter a slow growth took place, culminating in the incorporation of the city of El Cerrito ("the little hill") in 1917. Subsequent growth accompanied the economic development of the San Francisco bay area but was particularly affected by that of Richmond. (For comparative population figures *see* table in CALIFORNIA: *Population*.) El Cerrito is almost exclusively residential, with the greater part of the working population employed in Richmond, Oakland and San Francisco.

The city has had a council-manager form of government from 1948. (M. H. Mo.)

ELCHE, a town of southeastern Spain, in the province of Alicante, lies on the Vinalopó river, 21 km. (13 mi.) S.W. of Alicante, to which it is linked by road and rail. Pop. (1950) 56,341. Its narrow streets, flat-roofed, whitewashed houses and, above all, the innumerable date palms give the city, standing on a low hill in the midst of a sandy plain, a strikingly oriental aspect. The palms are of Carthaginian origin and this is their most northerly cultivation in Europe. Both dates and palm leaves are exported. Other products are pomegranates, figs, olives, almonds and cereals, and Elche manufactures oil, soap, esparto fabrics and leather and, for export, rope-soled shoes. A famous example of Iberian art, the polychrome stone statue known as the "Dama de Elche," was found on an archaeological site. A local custom is the annual presentation, on Aug. 14 and 15 in the church of Santa Maria, of the medieval drama *Misterio de Elche*. (M. B. F.)

ELDAD (BEN MAHLI) **HA-DANI,** 9th-century Jewish traveler, author of a fanciful geographical narrative which exerted an enduring influence throughout the middle ages and, most probably, gave rise to the legend of Prester John. Eldad visited Babylonia, Qayrawan and Spain, and caused a great stir by his colourful accounts of the lost ten tribes, he himself claiming to be a descendant of the Danites (hence ha-Dani) who—together with Naphtali, Asher and Gad—were to have established a Jewish kingdom "beyond the rivers of Abyssinia" (Kush). His veracity was frequently challenged, not least because of the halakhoth, or ritual prescriptions, which he alleged to have brought from his native country and which diverged sharply from those of the Talmud.

His narrative (*Sefer Eldad*) has appeared in several languages and in widely deviating versions (1st ed., Mantua, 1480). It established his reputation as a philologist, while his ritual idiosyncrasies have caused modern critics to consider him either as a Karaite missionary or an Ethiopian Jew (Falasha). But his language reveals no traces of Ethiopic, nor does his narrative betray a first-hand knowledge of Abyssinia. He shows, however, more than a casual acquaintance with Arabia, and his Hebrew offers much evidence of an Arabic substratum. It is therefore likely that he was a Jew from south Arabia where Jewish colonies had existed from early times.

See Abraham Epstein, *Eldad ha-Dani* (1891); S. Krauss, "New Light on Geographical Information of Eldad Hadani," *Tarbiz* (1936–37). (E. U.)

ELDER, a ruler or official in certain political or ecclesiastical systems of government. The use of the title, in origin political, has been retained from a period when age was considered an indispensable condition of leadership in the community.

The term elder in the early Christian Church, though possibly influenced by the use of the title for secular magistrates in Asia Minor, was derived from the Israelites, who shared it with other Semitic peoples. Moses appointed 70 elders as intermediaries between himself and the people (Num. xi, 16). In the New Testament elders (Gr. *presbyteroi*) are mentioned together with bishops (Gr. *episcopoi*) as leaders of local churches; in some passages the two terms seem interchangeable. Later the word *presbyteros* came to mean "presbyter" (*i.e.*, priest). It is thus difficult to decide on its exact significance in the early church (*see* MINISTRY, CHRISTIAN). After the threefold ministry of bishop, priest and deacon became fully accepted in the 2nd and following centuries, the office of elder lapsed; it was revived at the Reformation in the 16th century by certain churches, notably the Presbyterians. According to Presbyterian theory of church government, there are two classes of elders: the teaching elders, called ministers, specially set apart to the pastoral office, and the ruling elders, who are laymen chosen generally by the congregation to assist the minister in the oversight and government of the church. Among other churches which have the office of elder are the Disciples of Christ and the Moravians. (J. G. DA.)

ELDER, the popular designation of the deciduous shrubs and trees constituting the genus *Sambucus* of the family Caprifoliaceae (*q.v.*). For the box elder *see* MAPLE.

The common elder, *S. nigra,* is found in Europe, north Africa, western Asia, the Caucasus and southern Siberia; in sheltered spots it attains a height of over 20 ft. The bark is smooth; the shoots are stout and angular, and the leaves glabrous, pinnate, with oval or elliptical leaflets. The flowers, which form dense flat-topped clusters (corymbose cymes), with five main branches, have a cream-coloured, five-lobed corolla, five stamens and three sessile stigmas; the berries are purplish-black, globular and three- or four-seeded, and ripen about September.

The elder thrives best in moist, well-drained situations, but can be grown in a great diversity of soils. It is found useful as an ornamental for making screen fences in bleak, exposed situations, and also as a shelter for other shrubs in the outskirts of plantations. By clipping two or three times a year, it may be made close and compact in growth.

ROCHE

AMERICAN OR SWEET ELDER (SAMBUCUS CANADENSIS)

The young trees furnish a brittle wood, containing much pith; the wood of old trees is white, hard and close-grained, polishes well, and was employed for shoemakers' pegs, combs, skewers, mathematical instruments and turned articles. Young elder twigs deprived of pith have from early times been made into whistles, popguns and other toys.

Several varieties of *S. nigra* are known in cultivation: *aurea,* golden elder, has golden-yellow leaves; *laciniata,* parsley-leaved elder, has the leaflets cut into fine segments; *rotundifolia* has rounded leaflets; forms also occur with variegated white and yellow leaves, and *virescens* is a variety having white bark and green-coloured berries.

The elder was known to the ancients for its medicinal properties, and in Europe the inner bark was formerly administered as a cathartic. The flowers contain a volatile oil, and serve for the distillation of elder-flower water, used in confectionery, perfumes and lotions. The leaves are employed to impart a green colour to fat and oil, and the berries for making wine, a common adulterant of port. The leaves and bark emit a sickly odour, believed to be repugnant to insects.

According to German folklore, the hat must be doffed in the presence of the elder tree; and in certain of the English midland counties a belief was once prevalent that the cross of Christ was made from its wood. It was also a common medieval tradition that the elder was the tree on which Judas hanged himself. In Denmark the tree was supposed by the superstitious to be under the protection of the "Elder-mother": its flowers may not be gathered without her leave; its wood must not be employed for any household furniture; and a child sleeping in an elderwood cradle would certainly be strangled by the Elder-mother.

The scarlet-berried elder, *S. racemosa*, is the handsomest species of the genus. It is a native of various parts of Europe, Asia and North America, growing in Great Britain to a height of over 15 ft., but often producing no fruit. The dwarf elder or danewort (supposed to have been introduced into Great Britain by the Danes), *S. ebulus*, a common European species, reaches a height of about six feet. Its cyme is hairy, has three principal branches, and is smaller than that of *S. nigra;* the flowers are white tipped with pink.

Besides the red-berried elder, found in rocky soil from Newfoundland to Alaska and southward to Georgia and California, several other species are native to North America. The best known of these is the American or sweet elder (*S. canadensis*), common in moist soil from Nova Scotia to Manitoba and south to Florida, Texas and the West Indies, with deep-purple or black edible fruit. The blue elder (*S. caerulea*), found from British Columbia to Montana and south to lower California and New Mexico, usually a thicket bush 6 ft. to 10 ft. high but sometimes a tree 25 ft. high, bears blue fruit covered with a white bloom.

ELDON, JOHN SCOTT, 1ST EARL OF (1751–1838), as lord chancellor of England during most of the period from 1804 to 1827 was memorable for the length of his service, the soundness and slowness of his judgments and his opposition to reform. Born at Newcastle upon Tyne on June 4, 1751, he was educated at Newcastle grammar school and University college, Oxford. In 1772 he made a runaway marriage with Elizabeth Surtees, the daughter of a wealthy Newcastle banker. The following year he was admitted to the Middle Temple, and was called to the bar in 1776, practising in London and on the northern circuit. He became friendly with Lord Thurlow, who in 1783 procured his return to parliament for Lord Weymouth's nomination borough of Weobley, Herefordshire, for which he sat until 1796. He was appointed solicitor general and knighted in June 1788.

In 1793 he became attorney general and in that capacity made himself hated by the reformers, being largely responsible for Pitt's "gagging acts" which aimed at suppressing political meetings and the publication of seditious literature, and for conducting the trials for high treason of the group of reformers, including Thomas Hardy and Horne Tooke, who had expressed sympathy with the ideas of the French revolutionaries.

Worn out with these labours, he asked for and was appointed to the office of chief justice of the common pleas, and was given a peerage (1799), with the title of Baron Eldon. He became lord chancellor in 1801 in the Addington ministry, and was known during his long service in that position for his repressive policy toward internal dissent and his uncompromising opposition to Catholic emancipation. He subsequently served in the Pitt ministry of 1804–06. Eldon was again lord chancellor in the Portland ministry of 1807, retaining the office, under two more prime ministers (Perceval and Liverpool) for 20 more years. George III then told him that he wished and hoped he might keep the great seal until he died.

His position, and that of his colleagues, was threatened when George III was judged permanently insane in 1810 and the prince of Wales became prince regent (1811). They had been the friends and advisers of the prince's worst enemy, the princess of Wales, and, too, the Regency act embodied their view that his power should be restricted for one year. In the end, however, the prince retained his father's ministers. Eldon had already abandoned the princess, and now devoted himself to the service of his new master as he had done to that of his old. George IV confessed that at first he had hated his lord chancellor, but finally became as much attached to him as George III had been. Describing his resignation in April 1827 (he refused to serve under Canning, the advocate of Catholic emancipation), Eldon said that the king was so much affected that he threw his arms around his neck and shed tears.

Eldon had held the great seal for a longer period than any of his predecessors. In 1821 the king promoted him to an earldom. Although he had more than once, before 1827, talked of retirement because of advancing years, Eldon was seriously offended with Wellington in Jan. 1828, not merely because he was not offered a place in the new ministry (he would have liked to be president of the council), but because he was not even consulted about arrangements. "They might have given me an opportunity of offering my services to the country," he said reproachfully. He protested that he was "no enemy to real reform," but his definition of real reform was notoriously a narrow one: he opposed Catholic emancipation, the repeal of the Test and Corporation acts, the commutation of the tithe, measures of law reform, the abolition of imprisonment as a punishment for debtors and the abolition of the slave trade, as well as the reform of parliament itself. He died in London on Jan. 13, 1838.

His two sons predeceased him, and his grandson John (1805–54) succeeded to the earldom.

As a judge, Eldon's greatness is indisputable, though his judgments, as voluminous as they are profound, are notably clumsy in form. It was under his influence that the injunction became so important a weapon in the chancellor's armoury, and the rules for its use became settled. He had one fatal fault, his slowness, which outweighed all his virtues. The notorious slowness of the chancery reached its climax under him, and it is said that he would keep papers by him for years, though he rarely altered his first estimate of a case.

BIBLIOGRAPHY.—Horace Twiss, *Life of Lord Chancellor Eldon* (1844); W. E. Surtees, *A Sketch of the Lives of Lords Stowell and Eldon* (1846); Lord Campbell, *Lives of the Lord Chancellors*, vol. 7 (1847); W. C. Townsend, *Lives of Twelve Eminent Judges* (1846); Sir William S. Holdsworth, *Some Makers of English Law* (1938).

(A. AL.; X.)

ELDORADO (Span. "the gilded one"), a name first applied to a chief of an Indian town near Bogotá, Colombia, who according to legend plastered his naked body with gold dust during annual festivals, then plunged into Lake Guatavita to wash off the dust. At the conclusion of the ceremony his subjects cast jewels and objects of gold into the lake. Spanish conquerors heard of this before 1530, and one of them told of being entertained by Eldorado himself at a city called Omagua. In 1538 Spaniards from Peru and by way of the Magdalena valley and the Germans overland from Venezuela converged on the Bogotá highlands in search of the "gilded man." No trace of him was found, but the Spaniards remained as conquerors of the area.

As the search continued into the Orinoco and Amazon country, "Eldorado" came to mean a whole fabulous country of gold, with legendary cities named Manoa and Omagua. In this quest, Gonzalo Pizarro crossed the Andes from Quito (1539), Francisco Orellana sailed down the Napo and the Amazon (1541–42) and Gonzalo Jiménez de Quesada explored eastward from Bogotá (1569–72). Sir Walter Raleigh tried to find Manoa in the Orinoco lowlands (1595), while Spaniards sought Omagua nearby. In 1603 the Portuguese Pero Coelho de Sousa explored northward from Pernambuco, and the golden city of Eldorado was shown on maps of Brazil and the Guianas for years thereafter.

Eldorado was only one of many mythical regions of great wealth —Cibolá, Quivira, The City of the Caesars, Otro México being among the more famous of the others—the search for which contributed so significantly to the rapid overrunning of a major portion of the hemisphere by the Spaniards and other Europeans.

Since then Eldorado has come to be used for any place where riches could be had easily and quickly. The name was given to a California county and to towns in Latin America and in the United States. The story is often mentioned in literature, as in Milton's *Paradise Lost* (xi, 411) and Voltaire's *Candide* (ch. 18, 19).

(J. J. J.)

EL DORADO, a city of southern Arkansas, U.S., about 100 mi. S. of Little Rock, the seat of Union county, is an oil, lumber and cotton centre. De Soto explored the area in 1541. Founded in 1843 by Robert Black, John Hampton and Green Newton, commissioners who were instructed "to centrally locate the seat of Union county," and incorporated July 15, 1845, El Dorado grew slowly until the railroads came in 1889. Lumber and cotton were major products before oil was discovered Jan. 10, 1921, in the Busey well. Conservation methods pioneered in the nearby Schuler field are followed by the national oil industry. More than 30 plants, including refineries and a bromine plant tapping oil-field brine, provide employment. The state's oil and gas commission is

located there. Railroads, airlines, buses and nearby Ouachita river afford transportation. A boys' club, stadium, teen-club library, auditorium and armory, built 1947–59, form El Dorado's civic centre. For comparative population figures *see* table in ARKANSAS: *Population.* (A. L. SP.)

ELEA: *see* VELIA.

ELEANOR OF AQUITAINE (*c.* 1122–1204), queen consort of Henry II of England, was the daughter and heir of William X, duke of Aquitaine, by Aenor of Châtellérault. She succeeded to the duchy in April 1137, and at Bordeaux in the following July married Louis le Jeune, afterward Louis VII of France, who succeeded his father Louis VI less than a month later. Eleanor bore Louis two daughters but no sons. She accompanied him on the second crusade and while they were in the east (1147–49) the estrangement between them began, which, despite papal efforts at reconciliation, led to the annulment of their marriage on grounds of consanguinity (March 1152). Louis, though devoted to his wife, was undoubtedly troubled by their kinship, while Eleanor resented her husband's being "more monk than king." In May 1152 she married Henry, duke of Normandy and count of Anjou, who two years later vindicated his claim to the English throne. Thus by far the greater part of France was brought under the rule of the king of England; but the loss of Aquitaine, then regretted, was really a blessing in disguise for the French monarchy, not yet capable of exploiting possession of the southwest.

There were three daughters of Eleanor's second marriage, and five sons, of whom only two, Richard I "Coeur de Lion" and John, survived their father. Eleanor was 11 years older than Henry and had too much spirit to tolerate his persistent infidelity. She supported her sons in their rebellion (1173), in alliance with her former husband; in consequence, she spent most of the years 1174–83 in confinement, but after 1185 she again shared in the government of Aquitaine. On Henry's death she reasserted her hereditary position in Poitou and Gascony, where her role from 1189 to 1204 was of crucial importance. She supported King Richard (1189–99) against her youngest son John and the king of France, while in 1199 her opposition to her grandson Arthur of Brittany was decisive in securing Anjou and Aquitaine for John. In 1202, aged about 80, she held Mirabeau near Poitiers against Arthur until John, coming promptly to her relief, was able to take him prisoner. After the loss of Normandy (1204) it was essentially her ancestral lands and not the old Norman territories which remained loyal to the English crown and formed the *casus belli* of the Hundred Years' War. She died on April 1, 1204, and was buried at Fontevrault. Eleanor was renowned in the Frankish and Mediterranean world as an able and forceful queen and a beautiful and passionate woman, the patron of troubadours and the inspiration of many love poems and of courtly literature generally.

BIBLIOGRAPHY.—K. Norgate, *England Under the Angevin Kings* (1887); A. Kelly, *Eleanor of Aquitaine and the Four Kings* (1952); H. G. Richardson, "The Letters and Charters of Eleanor of Aquitaine," *English Historical Review,* vol. lxxiv (1959). (G. W. S. B.)

ELEANOR OF CASTILE (1246–1290), queen consort of Edward I of England, was the daughter of Ferdinand III of Castile and his second wife Joan, daughter of Simon, count of Ponthieu. She married the Lord Edward at Burgos (Oct. 1254), her half brother Alfonso X of Castile renouncing in Edward's favour his claims to Gascony. Eleanor came to England, residing at Windsor castle, where her first child Eleanor was born (June 1264), soon after the barons had defeated the royalist party at the battle of Lewes. She was sent for safety to France and returned (Oct. 1265) after Edward's victory at the battle of Evesham. She accompanied Edward on his crusade (1270–73) and was reputed to have saved his life when he was wounded at Acre (1272). She visited Alfonso X in Castile on her way home, and was crowned with Edward on Aug. 19, 1274. She inherited the county of Ponthieu from her mother (1279). Eleanor incurred some criticism by using Jewish moneylenders to obtain lands, and by the harsh conditions imposed on some of her tenants, but was personally pious and gentle. Edward never fully recovered from her early death at Harby, Nottinghamshire, on Nov. 28, 1290, for they were completely devoted to each other: he erected the famous "Eleanor" crosses at each place where her body rested on its journey to London. (R. F. T.)

ELEANOR OF PROVENCE (1223–1291), queen consort of Henry III of England, was the second daughter of Raymond Berengar IV, count of Provence, and Beatrice, daughter of Thomas I, count of Savoy. Designed to further Henry's continental diplomatic ambitions, his marriage with her (Jan. 14, 1236) ultimately produced friendship with Louis IX. It brought to England many Provençals and Savoyards seeking official posts, benefices, lands, pensions and marriages in Henry's gift; the most notable were Eleanor's uncles Boniface and Peter of Savoy. Energetic, decisive, but not politically ambitious, she was greatly loved by Henry and her children and influenced Henry powerfully. She accompanied Henry to Gascony in 1242, was co-regent while he was in Gascony in 1253–54 and brought her son Edward (afterward Edward I) to Gascony in 1254 to marry Eleanor of Castile. During the Barons' War the Londoners attacked her barge (1263), forcing her to flee to St. Paul's, an insult which Edward never forgave. Withdrawing to France (Oct. 1263), she became the leader of the royalist exiles after the baronial victory at Lewes (May 1264). She implored King Louis's help and pledged Henry's jewels and his rights in Limoges, Cahors and Périgord under the treaty of Paris (1259) for loans to raise an invasion force, but her fleet was wrecked at Sluys. After the defeat of the baronial party at the battle of Evesham she returned to England (Oct. 1265), and when Henry died she retired to Amesbury convent, taking the veil (1286) and dying there on June 25, 1291. (R. F. T.)

ELEATIC SCHOOL, a Greek school of philosophy, which existed during the 5th century B.C. It was named after Elea, a Greek colony in southern Italy and the home of Parmenides. As a result of Aristotle's systematizing, Xenophanes of Colophon has often been counted as the school's founder; but its chief tenets undoubtedly appear first in Parmenides (*q.v.*). Zeno of Elea was the pupil of Parmenides; and Melissus was the third and last leader of the school.

While the chief tenets of the Eleatics may be studied in the three articles PARMENIDES, ZENO and MELISSUS, a summary of certain features in Parmenides' doctrine will explain its outstanding importance to philosophy. It argues that reality is one undifferentiated whole; the "things" and their changes which we observe through our senses are mere appearances or names. In opposition to the Ionian natural philosophers who accepted the reality of change but looked for material stuff underlying it, Parmenides' position was reached by an abstract, logical or generalized analysis of what we mean by "to be." He was therefore correctly regarded as the founder of metaphysics. His use of the law of contradiction meant the introduction of logic as the prime method of philosophy. Melissus defended him along the same lines, while Zeno's work forms a complement by defending him negatively or indirectly, through a *reductio ad absurdum* of his opponents' position. With Zeno, however, the analysis focuses less on "being" than on the co-extensive term "unity." He may have intended to deal with the pluralism of Pythagoreans, whose materialization of numbers made them specially aware that plurality implied units. Here was the weak point of Eleaticism, and it was noticed by Gorgias and by Plato: the absurdities demonstrated in the alleged units of the pluralists had only to be turned against the unitary reality of the Eleatics.

By Plato's time the heirs of Zeno were so-called Megarian thinkers, who concentrated on logical puzzles involved in the notions of predication, contradiction, possibility and motion; we can see how they compelled Plato to attend to purely logical matters in the *Euthydemus, Parmenides* and *Sophist.* The influence of the Eleatic school was due, in fact, to the very generality of its analysis. Melissus' remark, that if there were a "many" these would have to be like the Eleatic "one" and not as they appear, was acceptable to Democritus and to Plato alike. The properties of the atom were almost certainly suggested by those of "the one." The Eleatic or monistic strand in Plato is conspicuous in the Neoplatonists and, through their medium, was to be consciously woven by Hegel into his objective idealism.

See general works listed under PARMENIDES.

For discussions of modern specialists' interpretations, *see* P. Albertelli, *Gli Eleati* (1939). (A. C. Ld.)

ELECAMPANE (*Inula helenium*), a perennial plant, family Compositae (*q.v.*), common throughout central and southern Europe, many parts of Britain, and in Asia as far eastward as the Himalayas. It is also widely naturalized in North America, occurring along roadsides and in fields. Elecampane is a coarse, stout and rigid herb, the stem of which attains a height of three to six feet; the leaves are large and toothed, the lower ones stalked, the rest stalkless; the flowers are yellow, two inches broad, and have many rays. The root is thick, branching and mucilaginous, and has a warm, bitter taste and a camphoraceous odour. Besides the compound inulin (*q.v.*), isomeric with starch, the root contains helenin, a drug formerly used in medicine. Inulin is used in modern medicine in tests to determine the efficiency of kidney excretion. (N. Tr.; X.)

ELECTION: *see* Electoral Systems; Representation.

ELECTORAL COLLEGE, U.S. The U.S. constitution provides that the president and vice-president shall be elected by electors chosen within each state. Considered as a group, these electors are known as the electoral college. The number of electors from each state is equal to the total of the state's representation in congress (senators and representatives combined). Electors were originally chosen by state legislatures but this practice was soon abandoned in favour of popular election. In providing for the electoral college the framers of the constitution expected that each elector would exercise his own judgment in casting his vote but, with rare exceptions, this expectation was not realized. When presidential and vice-presidential candidates came to be nominated by political parties, candidates for the office of elector pledged themselves in advance to vote for their party's candidates. Because the electoral college has not fulfilled its intended purpose, many proposals have been made to eliminate it. *See* United States (of America): *Administration and Social Conditions;* Electoral Commission.

ELECTORAL COMMISSION, in United States history, a commission created to settle the disputed presidential election of 1876. In this election Samuel J. Tilden, the Democratic candidate, received 184 uncontested electoral votes, and Rutherford B. Hayes, the Republican candidate, 163. The states of Florida, Louisiana, Oregon and South Carolina, with a total of 22 votes, each sent in two sets of electoral ballots, and from each of these states except Oregon one set gave the whole vote to Tilden and the other gave the whole vote to Hayes. From Oregon one set gave the three electoral votes of the state to Hayes; the other gave two votes to Hayes and one to Tilden.

The constitution leaves the manner of selecting electors to state law; the electoral ballots are sent to the president of the senate, who "shall, in the presence of the Senate and House of Representatives, open all certificates, and the votes shall then be counted." This provision left unanswered several questions of vital importance in 1876. Might congress or an officer of the senate go behind a state's certificate and review the acts of its certifying officials? Might it even examine into the choice of electors? And if it had these powers, might it delegate them to a commission? The fact, moreover, that in 1876 the Republicans controlled the senate and the Democrats controlled the house, made any harmonious settlement of these issues much more difficult to reach.

On Jan. 29, 1877, after weeks of acrimonious discussion in and outside of congress, including threats of civil war, congress created the electoral commission to pass upon the contested returns. Many more Democrats than Republicans voted for this measure. The decisions of the commission, which was given "the same powers, if any" possessed by congress in the premises, were to be final unless rejected by the two houses acting separately. Five Democratic and five Republican members of congress, five supreme court justices, including two of each party, and a fifth justice chosen by these four, composed the commission. This 15th member, so crucial to the outcome of the electoral-vote counting, was expected to be Associate Justice David Davis, an independent. He declined to serve, however, when at this juncture the Illinois legislature elected him to the senate. Thereupon, the four justices on the commission picked Justice Joseph P. Bradley, a Republican whose record on the court made him acceptable to the Democrats.

Although Hayes had polled a majority of the popular vote in Oregon, and in spite of wrongdoing by both parties in Louisiana and South Carolina was probably entitled also to their electoral votes, he almost certainly had been defeated by Tilden in Florida. Therefore, beyond much doubt, Tilden had won the election by an electoral vote of 188 to 181. Convinced that the Republican certification boards had fraudulently rejected many valid votes for Tilden, the Democratic leaders abandoned their customary adherence to states' rights and insisted that the electoral commission go behind the official returns to correct the injustice. With as much inconsistency, in view of their usual support of national power at the expense of states' rights, the Republicans argued that the sovereignty of a state made unconstitutional any probing back of its official returns.

Virtually accepting this latter contention, the eight Republicans on the commission overrode their seven Democratic colleagues and accorded the electoral votes of Florida and Louisiana to Hayes. On Feb. 23, by unanimously assigning South Carolina and Oregon also to his column, it awarded him the presidency by a margin of one electoral vote (185 to 184 for Tilden). After a week of ominous bluster, a joint session of congress confirmed the commission's count. By then it was March 2, or only 48 hours before inauguration day.

Both economic and political considerations help to explain why many leading Democrats, and notably those in the south, accepted the outcome without much protest. Although they failed to gain control of the White House, they had been promised by Pres. U. S. Grant or by politicians close to Hayes that the remaining troops would be withdrawn from the former Confederate states, that southerners would share in the federal patronage and, perhaps most gratifying of all, that the Republicans, unlike many northern Democrats, would support congressional appropriations for much-needed railroad construction and other internal improvements, especially in the lower Mississippi valley and Texas. But, quite apart from this so-called "bargain" or "compromise of 1877" whereby Tilden apparently was sacrificed "to make a southern holiday," the acceptance without violence of the decisions of the electoral commission, so soon after the Civil War, gave a welcome reassurance to the American people that self-government and domestic peace were not incompatible.

Bibliography.—C. Vann Woodward, *Reunion and Reaction: the Compromise of 1877 and the End of Reconstruction* (1951); Allan Nevins, *Abram S. Hewitt, with Some Account of Peter Cooper* (1935); Alexander C. Flick, *Samuel J. Tilden, a Study in Political Sagacity* (1939); Hamilton J. Eckenrode, *Rutherford B. Hayes, Statesman of Reunion* (1930); Paul L. Haworth, *The Hayes-Tilden Disputed Presidential Election of 1876*, new ed. (1927); and for the sources, many volumes of documents of the 44th congress, 2nd session. (W. T. Hs.)

ELECTORAL SYSTEMS. An electoral system is a means of choosing officers or representatives of an organization or making other decisions by the vote of its qualified members. An election is one of the means by which a society may organize itself and make formal decisions. Where voting is free, it is a system for making decisions regarding the power relations in a society, a method for securing political obedience with a minimum of sacrifice of the individual's freedom. The essence of a democratic election is freedom of choice. Under a dictatorship the individual may be asked to go through some of the formalities of casting a ballot, but this is a mock election, not a true one. He can merely ratify choices already made by the dictator or the ruling clique.

History.—Electoral systems had their beginnings in ancient times. In classic Greece most public officers were selected by lot but a few offices requiring special qualifications were filled by election. For the most part voting was by show of hands in a public meeting but in cases of ostracism and legal decisions voting was made secret by means of white and black balls of stone or metal and marked and unmarked shells. A Roman law of 139 B.C. provided that each voter should use a carved wooden ballot to indi-

cate his choices for candidates for the *comitia centuriata*.

In medieval times elections were used for various purposes, but the size of the electorate was extremely small. Indirect elections by means of electoral colleges were used to select chief executives in Germany and in some Italian cities. In 1562 a papal bull required the College of Cardinals to select the pope by a secret ballot.

In modern times elections have been associated with the system of representative government (*see* REPRESENTATION); but in some countries elections have included the selection of executive and administrative officials as well as legislative representatives. In the United States the president is chosen indirectly by a majority vote in the electoral college, but in practice the system almost amounts to a popular vote. U.S. state governors and many minor state and local administrative officials are elected by direct popular vote. A candidate for a representative position can stand on general principles, but a candidate for an administrative position, it is held by some critics, should have technical training and qualifications for the position. The inability of the average voter to make wise choices for administrative positions has led to the short ballot movement in the United States, which calls for the limiting of elective offices to policy-making posts.

Voting must be done in accordance with rules that are accepted in advance. Unanimous consent as a voting rule has been tried and found wanting. If a government cannot act until all members have agreed to a given course of action, stagnation results. Consent under a democratic electoral system means partial consent, a rule by the majority of those who vote. The winning of a majority is by peaceful means under a free election system—by persuasion, by electioneering, by trickery sometimes, but not by wholesale threats of violence. The reservation of certain minority rights at the same time guarantees that the majority will not misuse its powers.

The occasions on which the right to vote is to be exercised must be recurrent ones or the right becomes meaningless, for in the absence of a fixed term of office, the actual use of the election system may be indefinitely postponed. In countries where the parliamentary system is in operation and the chief executive, on the order of the cabinet or on his own initiative when the government in power has lost a vote of confidence, may call for a dissolution of one or both houses, general elections may be held before the regular term expires. Thus, in Great Britain and most of the commonwealth nations, the order for dissolution sets the time for the new elections.

The part that the electoral system plays in the machinery of government varies from country to country. The role of the electorate may be limited to a varying degree by the power of a hereditary second chamber as in England, by royal power in countries with a constitutional monarchy or by the power of the courts as in the United States. Ignorance, apathy and lack of training may keep some voters from making their influence felt. Various pressures, economic, social and political, are brought to bear upon the voters and they must learn to make wise choices or they will whittle away their freedoms.

Thus, of electoral systems in general it may be said that their real meaning depends upon their ultimate governmental effectiveness, their relationship to other political institutions and the social system within which they operate. In considering the subject, then, the examination should not be confined to the written text of the franchise acts alone, but must consider an electoral system broadly, as all those means whereby a person becomes a member of an elected assembly or is elected to fill some other government position.

Candidature.—The democratic development of the 19th and 20th centuries generally brought about (1) ease and freedom of candidature; and (2) an approximation of the age limit of candidature to that for the exercise of the vote.

As regards freedom of candidature, the tendency has been to avoid the interference of the state in the electors' choice of candidates, and to restrict the action of the state to providing that ineligibility shall be determined after the election by a tribunal likely to be impartial. Such tribunals are parliamentary as in the United States federal authority, Norway, the Netherlands and Belgium or extraparliamentary as in Great Britain and Australia. But statesmen have been unwilling to allow a perfect freedom of candidature and have sought to secure an element of electoral responsibility by requiring a candidate to be sponsored by a number of electors who sign a nomination form. The number thus required differs very much; *e.g.*, 10 in England and 100 in Belgium. Many countries also require candidates to deposit a sum of money that is forfeited if the candidate fails to receive a certain proportion of the votes.

Political Parties.—In the United States, as in Great Britain, the commonwealth and Europe, the electoral systems were established on the basis of the natural rights of the individual. The individual thus envisaged was regarded as an independent, self-sufficient, freely acting entity, without social or economic characteristics except those which originated in his self-interest and which were created by his private activity, unaided and unimpeded by the state. Thence naturally followed freedom of candidature and the universal franchise. But men do not act as independent atoms, and three-quarters of the way through the 19th century nothing was more striking than the enormous power and indispensable services exercised by political parties in organizing the enfranchised atoms; nor could one mistake the growing significance to men of the sense of state and its activities and institutions, or the grouping of men in vocations. The 19th-century state relied heavily upon political parties for the choice of candidates and even, in many countries, for other electoral services, like the distribution of ballot papers. But while the state was thus periodically created, dissolved and re-created by party activity, it is amazing that this process went on in a fit of statutory absent-mindedness.

The states of the United States were the first to suffer from this blithe unconsciousness. The large number of offices, including administrative and judicial, filled by election, the enormous territorial range of the country and the popular preoccupation with economic activity necessitated strong party organizations. Certain abuses arose in connection with party nominating conventions. Some conventions were held without due notice and in out-of-the-way places; they were packed with hooligans, party rolls were padded and ballot boxes stuffed. A reaction set in in the 1880s, and the states (California commenced as early as 1866) began to regulate the method of nomination. Legislation, at first permissive, became compulsory, and more and more offices were included. After 1912 some states added presidential preference primaries. All states save three or four established principles and methods for regulating the nomination of candidates through the party machinery.

It is necessary at the start to decide which parties are entitled to nominate candidates. Eligibility may be determined by the number of votes cast at the preceding election, or, more frequently, by a fixed proportion of the votes cast, ranging from 2% to 25% of the entire vote. It is also necessary to decide who is entitled to appear at the nominating primaries, and thus to define party membership. This may be done by a special enrollment in the party, by secret or open process, some time before the elections; by declaration asking for a primary ballot; by decision of the party officials (this in the southern states, where the problem of Negro franchise gave trouble); or by the Wisconsin open primary method, whereby the voter at the primary is given the ballots of all parties and may choose secretly which one he wishes to vote. Finally, it is necessary to regulate legally the time, place, method of voting and counting.

One question still remains: how can a citizen get his name placed on the nomination paper at the primary elections? This leads back to the ultimate problem. The methods laid down by the states vary. One method is that of petition which must be signed by a fixed number or an agreed percentage of the voters. This may be a very expensive method. In another the party committee nominates long before the election and dissentient elements may by petition present others. One other problem arises: what vote at the primary election constitutes an effective choice of a candidate? Some states require an absolute majority, secured in some

places by the second ballot and in others by the alternative vote system; others require at least 35% of the votes to be obtained for a nomination, and if the primary does not secure this a party convention is called.

Another peculiarity of the choice of candidates in the United States is that candidates are by law and custom required to be inhabitants of the state in which they run for office. Some states go further and demand that the candidate shall be resident of the district which he seeks to represent. Many U.S. observers believe that this restriction of choice is seriously detrimental to the quality of American legislatures and obstructs the development of party solidarity.

Democracy, Fascism and Communism.—No greater contrast could be presented to the developments in the United States than those in Fascist Italy, Nazi Germany and Communist U.S.S.R. During the period of Fascist rule in Italy (1922–43) and National Socialist hegemony in Germany (1933–45) and after 1917 in Russia the regimes of these states rested on a political basis which totally denied the individual's right to freedom. Fascism, for instance, started out with the whole nation as the unit of state life, and within that unity recognized the personality of corporations, economic and social, but only as integral elements of the state.

The individual had electoral significance in his proper corporation, and the corporation in the state, and it was positively denied that all individuals and all corporations are electorally equal and free. Under Fascism Italian legislation abolished the democratic electoral system. In May 1928, the corporative state having been created by the Fascist regime, a new system was set up. Various corporations were to nominate 900 candidates, the Fascist grand council to choose from these about 400 candidates, and these 400 candidates to be put before the electors for approval en bloc. (*See* also FASCISM.)

Though the policy of the U.S.S.R. was much different from the Fascist system, the electoral system is similarly converted to the use of a special organization working within a system which asserts the priority of the state over groups and individuals. The right to nominate candidates for the soviet is nominally given to Communist party organizations, professional organizations, military units, workshops and other vocational units. All this seems to favour a freedom of nomination so long as such nomination issues from a recognized "productive" group. But the Communist party is all-powerful, and is made so by various privileges and its capacity to extirpate opposition. There is only one nominee for each position to be filled. Candidature is thus totally unfree in practice.

In the countries where there is little or no legislature regulation of nominations, there are still local party caucuses which set their own terms for nomination—wealth, social status, electoral cleverness, even intellect. These caucuses have a great amount of power, though the law has not given it to them. Where party organization is strong, as in the Anglo-Saxon countries, the party is the deciding factor in the choice of candidates, though this does not mean the central authority of the party. Particularly in the Latin American countries, party organization is still weak, and small evanescent groups nominate. In the Germany of the Weimar republic and the Federal Republic of Germany, established after World War II, the peculiar nature of the system of proportional representation made the party machine dominant in the choice of candidates, for the party headquarters wishes to count upon safe seats for certain men and is able to offer safe seats to others out of its national fund of votes accumulated from the votes of unsuccessful candidates at the elections. It can be taken as fair generalization that, wherever the system of proportional representation operates with what is known as a list system, the party machine and central control within the party machine have the nominating power strongly in their hands.

Voters' Qualifications.—The trend in the 19th and 20th centuries has been toward the elimination of restrictions on voting, but there are still some suffrage restrictions with respect to property, residence, age, sex, education and race. The property qualification for voting, for example, has been reduced to a minimum in practically all countries. In the United States a few southern states clung to the poll tax requirement until ratification of the 24th amendment to the U.S. constitution in 1964. In some Latin American countries there are still occupational and income qualifications. In Europe and the U.S. there are occasional property requirements for voting on local affairs. Otherwise property ownership is no longer a measure of fitness to vote.

Residence requirements are defended on the ground that they promote stability and give the election authorities time to investigate qualifications. In Great Britain and France the requirement is six months' residence in the political jurisdiction concerned, and in the United States residence requirements range from six months to two years. Because of the ease and frequency with which voters may move from place to place, many persons are unjustly disfranchised by long residence requirements. Particularly for voting in federal elections, local residence requirements appear to have little value.

Closely related to the residence requirement is the rule that the voter must cast his vote in person. In an increasing number of countries this rule is mitigated by provisions for proxy voting or for absentee voting by mail. In Great Britain a voter who knows that he will not be able to go to the polls on election day may nominate someone in whom he has confidence to cast his vote for him. He thus votes by proxy. An alternative procedure for persons unable to vote in person because of illness, absence on business, or other reasons, is to obtain a ballot from election officials and send it in by mail, the so-called postal vote. In the United States the laws governing absentee voting vary from state to state and show considerable diversity. Nearly all other countries with a strong tradition of democratic voting and a high literacy rate have made provision for persons to cast their ballots *in absentia*. (*See* ABSENTEE VOTING.)

The most common minimum age for voting is 21 years, but 18 is the minimum in certain Latin American countries, the U.S.S.R., Yugoslavia, the Republic of Indonesia, Turkey and two states (Kentucky and Georgia) of the United States. Two other states, Alaska and Hawaii, have voting ages of 19 and 20 respectively. The minimum age for voting in the Netherlands is 23 and in Finland it is 24.

Women's suffrage (*q.v.*) had been widely adopted by the second half of the 20th century, but not by some Latin American countries, some middle eastern countries, a few Asiatic countries and in Switzerland. Countries with largely Muslim populations such as Libya, the Republic of the Sudan and Saudi Arabia held out against women voting, but Turkey gave women the right to vote in 1934. Indonesia also granted women the right to vote.

The trend has been away from literacy qualifications for voting but such qualifications remained in the electoral laws of Bolivia, Brazil, Ecuador, Peru, Chile, the Philippines and about one-third of the states of the U.S. Literacy tests are defended on the grounds that election campaigning depends upon written materials and election administration assumes familiarity with paper and pencil. In the Philippines the voters have to write out the names of the candidates whom they wish to support. Countries with high rates of illiteracy, such as India, have nevertheless conducted general elections by using tokens instead of paper ballots.

In countries where there is widespread belief in so-called "racial" differences, it has been difficult to apply the principle of equal suffrage. In the Republic of South Africa and in the southern states of the United States, Negroes have been kept from the polls. In South Africa the exclusion was made legal and complete. In the United States, the 15th amendment (1870) to the constitution forbids a legal colour ban, but intimidation and discrimination in the administration of the election laws (insofar as these involved literacy, residence and [until 1964] poll tax requirements) long prevented Negroes from voting in certain southern states. The trend in the United States, however, has been for an increasing number of Negroes to qualify as voters.

Suffrage is a precious right which has been vigorously sought by those to whom it has been denied. A disfranchised group is at a disadvantage. It cannot claim the same measure of public services, the same opportunities for public employment, the same chances of rising to power and the same self-respect that are claimed by those who vote.

Election Administration.—With the increasing size of the electorate, the administrative problems of conducting elections have become more and more complicated. The aim of election administration is to safeguard the rights of the voters, to serve the convenience of the voters, to keep the expense of holding elections to a minimum, and to prevent fraud. The devices to achieve these ends have been registration laws, ballot laws, laws governing the casting and counting of the votes, and corrupt practices acts.

When an election is to be held, registration machinery must be set up to determine who is entitled to vote under the law and to safeguard against voting by persons who are not qualified to vote. Registration procedures should impose the least possible burden upon the voters that is consistent with the achievement of these objectives. Proper identification of the voter is the key to a sound registration system. In some U.S. states and in the Philippines the signature of the voter is used as a means of identification. The voter signs his name at the time he registers and again when he applies to vote. The polling official compares the two signatures and establishes the identification. In India and parts of Africa, where there are so many illiterates, a fingerprint is used for the same purpose.

Time of Election.—The time of the year when elections are held, the day of the week, how early the booths are opened and how late shut are important points. In agricultural countries elections (where there are fixed terms of parliament) are sometimes fixed after the spring sowing or in autumn after the harvest. The day of the week is especially important in an industrial country, where work must proceed with regularity. Most European countries except Great Britain adopted a Sunday or a holiday as election day. In the United States some states have laws allowing voters to be absent from work for the time necessary to vote. In Great Britain arrangements were made for voting until 9 P.M.; in Germany and France, where polling does not fall on an ordinary workday, it ends at 6 P.M.

Process of Polling.—Around the process of polling great parliamentary fights were waged in the 19th century. The issue was to secure the freedom of the voter from outside influence, and honest counting of the votes. The first was secured by the proper organization of the whole electoral procedure; the institution of the secret ballot; and associating the representatives of the contending parties with the supervising officer of the proceedings. The official in charge of local election administration is usually a public official, either of the central authority such as the *préfet* in France, or of a local government authority such as the town clerk in Great Britain.

The Secret Ballot.—South Australia was the first state to introduce secrecy of the ballot (1858) and for that reason the secret ballot is referred to as the Australian ballot. The system spread to Europe and America to meet the growing public and parliamentary demand for protection of voters. The means for securing secrecy vary considerably.

Arrangements found in western Germany in the 1960s were typical of continental Europe. The voting urns were required to be four-cornered, of certain dimensions and closed, the only aperture being a small slit at the top. These urns were examined before the poll began and could not be opened until the count began. The voter indicated his choice by marking his ballot and placing it into an officially stamped envelope. The latter was given to him by an official, and no other envelope was legally valid. The envelope was made of opaque paper of a legally determined size. The voter placed his ballot into the envelope in a special stall or voting booth so arranged as to preserve secrecy. The envelope containing the ballot was then given by the voter to the returning officer, who took the name of the voter, verified his right to vote and then put the vote into the urn.

In Great Britain the secret ballot was finally introduced for all parliamentary and municipal elections by the Ballot act of 1872. In the United Kingdom, as in the commonwealth generally, the ballots were officially provided and issued from a counterfoiled volume, but no envelope was provided and the voter cast his ballot directly into the box.

Until 1913 the French system was full of defects; the candidates circulated the ballots even outside the polling sections, and the voter merely folded the paper and gave it to the presiding official to put into the ballot box. After 1913 the envelope system was adopted. After Oct. 1919 also, the state printed the ballots, though at the expense of the candidates, sending them to the voters by mail under an official stamp. No voter could be offered ballots on election day except from the returning officer's staff inside the polling station.

Before 1884 the general practice in the United States was either open voting or, where this rudimentary and clumsy process had been superseded, voting by ballot. After the presidential elections of 1884 the Australian ballot system was extensively adopted. By the early 1960s in all states except South Carolina ballots were printed at the public expense, distributed in the polling stations, officially numbered, secretly marked and folded, and were identifiable by reference to the counterfoil retained by the official. Increasing use was also made of voting machines (*q.v.*), which assure a greater degree of secrecy in voting and honesty in counting. In the United States controversy long raged around the question whether the form of the ballot should make it easy to vote a party ticket. In party column states the ballot is so arranged that the voter has only to mark it at one certain point to indicate his acceptance of all the candidates of any one party. Attempts to encourage the independence of voters were made by such plans as the Massachusetts plan, adopted in that state in 1888. All the names of candidates for all parties for each office are placed in alphabetical order under the title of the office. The only way to vote is to mark a cross opposite the name of each candidate desired. In many states the party signs are forbidden, especially for municipal offices.

In countries where large numbers of illiterates vote, special voting arrangements have to be made. In India voting is done by tokens and each candidate has his own ballot box in each voting booth which is identified by a symbol placed on the outside. Each voter puts the token in the box of the candidate he wishes to support. In Indonesia, where illiteracy is also widespread, the party list system is used and each party is identified on the ballot by its own symbol. The voter pierces the symbol of the party he wishes to support with a nail placed in the voting booth for that purpose.

Political parties are associated with the administration of the ballot in the better-organized democracies to an interesting but not surprising extent. In Great Britain agents of the candidates are allowed to be present at the polling station. They are sworn to secrecy and may not interfere with a voter. They are present at the count of votes and raise objections to spoiled or uncertain ballot papers. The clerks are appointed by the public authority. In the United States the parties have a larger share in the electoral proceedings. Besides the rights given by British law, American laws provide for the clerical work to be done by people chosen by the parties.

The Influencing of Voters.—It is of the essence of candidates' desires to represent others that they should make clear to their constituents their purposes and character. Indeed, the large number of constituents for each representative in the modern state compels the creation of a machine and methods to impress the voters. Representation is unthinkable, under modern social conditions, on any other basis. Candidates, however, have rarely conceived their task as one simply of enlightening the electors. The attitude of a John Stuart Mill or a Macaulay is a rarity. The candidate, and certainly his party followers and workers, as well as his agent, want victory, and this desire too often causes them to adopt undesirable tactics. Laws have therefore been made restraining injurious activity, and they can be broadly divided into two classes, those regulating expenditure at elections, and those defining and creating penalties for corrupt practices (*q.v.*).

In Great Britain the Corrupt and Illegal Practices act of 1883 codified and added to the piecemeal legislation of previous centuries, making a code of admirable strictness. The heart of the act was the placing of the onus on the election agent for any infringement of the provisions. Corrupt practices under this act included: (1) Bribery by gift, loan or promise of money or

money's worth to vote or abstain from voting; by offer or promise of a situation or employment to a voter or any one connected with him, by giving or paying money for the purpose of bribery, by gift or promise to a third person to procure a vote, or payment for loss of time, wages or traveling expenses to secure a vote. The consequences are the same whether bribery is committed before, during or after an election. (2) Treating, which means the provision or payment for any person, of meat, drink, entertainment or provision, at any time, in order to induce him or any other person to vote or abstain from voting—and such extends to the wives or relatives of voters. (3) Undue influence; *i.e.*, making use or threatening to make use of any force, violence or restraint, or inflicting or threatening to inflict any temporal or spiritual injury on any person in order to influence his vote, or by duress or fraud impeding the free exercise of the franchise by any man. (4) Personation; *i.e.*, applying for a ballot paper in the name of another person, whether alive or dead, voting twice at the same election, aiding or abetting personation, forging or counterfeiting a ballot paper. (5) Unauthorized expenditure; *i.e.*, expenditure which is not authorized in writing by the election agent. Illegal practices include paid conveyancing, advertising and hiring, without authority, committee rooms; voting without qualification; false statements made about candidates; disturbance of public meetings between the issue of the election writ and the return of the election; printing, publishing or posting any bill, placard or poster not bearing on its face the name and address of the printer and publisher; illegal proxy voting. Heavy fines and withdrawal of the right to vote or be a candidate were attached to these offenses. The expenses of the candidates were limited by the acts of 1883, 1918 and 1928, and after the passage of the Representation of the People act, 1948, stood at £450 plus 1½*d*, for each elector in borough constituencies and £450 plus 2*d*. for each elector in county constituencies. These provisions did not apply to the amounts that might be spent by national party organizations for general publicity during the elections. Each candidate was also allowed to spend a sum not exceeding £100 on his personal expenses, including travel and hotel expenses. Private vehicles used on behalf of a candidate at an election had to be registered with the returning officer. Their number was limited—in a county constituency to 1 for every 1,500 registered electors and in a borough constituency to 1 for every 2,500.

In the United States campaign expenditures have been subjected to close scrutiny in the 20th century and laws have been enacted to limit campaign expenditures.

The federal Corrupt Practices act of 1925 established a normal maximum of $2,500 for a candidate for the house of representatives and $10,000 for a candidate for the senate, though higher totals were permitted in the more populous states. The Hatch act of 1940 established two limitations. First, no national political committee might receive or disburse in any calendar year more than $3,000,000. This provision has had the effect of decentralizing campaign finance. The state committees have increased their share in the raising and spending of party funds. The second limitation of the Hatch act was that no single person, committee or corporation might contribute more than $5,000 toward the election of any person in any calendar year. The unrealistic character of the limitations has been recognized repeatedly by congressional committees which have urged the limits be substantially raised.

Besides money the party leaders had as election currency thousands of federal, state and local offices in their gifts, but these, after creation of the Civil Service commission in 1883, were gradually withdrawn from the spoils system. After 1890 many laws sought to regulate electoral activity, and they concentrated upon publicity of the source and destination of campaign funds, and securing reports of personal service.

Some states (*e.g.*, Alabama, Kentucky, Minnesota, New Jersey, Wisconsin and several others), required accounts and reports several days prior to the election. Others (Alabama, Minnesota, Massachusetts and Wisconsin), required that all political advertisements be signed and marked "paid advertisement," stating the price paid, the advertiser and the author. In the three latter states and Kansas the purchase of newspaper support was forbidden. One contribution of U.S. practice to electoral procedure was the introduction of the publicity pamphlet. Oregon in 1908 passed a law requiring all candidates to take at fixed rates from one to four pages of a pamphlet published and sent to all voters by the secretary of state. Persons desiring to oppose candidates could also take space for that purpose, providing due notice was given to the person attacked and subject to the law of libel. Several other states adopted this system, but it worked best in its native state.

French law on corrupt practices was codified in the decree of Feb. 1852, but little heed was taken of its provisions until the laws of July 29, 1913, and March 31, 1914, redefined the offenses and created severe penalties.

Delegate and Representative.—Once the candidate has succeeded in getting elected, is he to act upon instructions and primarily for the constituency or to use his discretion broadly for the welfare of the whole country? Formerly the generally accepted view was that the elected person is a representative and has no particular mandate. This theory was embodied in France in the constitution of Sept. 3, 1791 (sec. 3, art. 7); "The representatives chosen in the departments shall not be representatives of a particular department, but of the entire nation, and no one may give them any instructions (*mandat*)." The theory and even the words of the article were copied by most countries in their written constitutions.

In England much the same theory was enunciated by Edmund Burke in his speech to the electors of Bristol (1780), but with less abstract dogmatism. (*See* Representation.) The 19th century witnessed the triumph of the French view over new forms of political organization. The prime movers in political life became political parties, propagandist organizations and organized economic groups such as trade unions. They demanded an adherence not to their strict instructions but to their general principles, and it became a general though not a legal rule that the member who finds himself in plain disagreement with his supporting group must explain the grounds of his divergence.

In the Soviet Union, however, the mandate of the whole people is denied and the system of the definite and imperative mandate was adopted. By the constitution of 1936, art. 142, "It is the duty of every deputy to report to his electors on his work and on the work of his soviet. Every deputy may be recalled at any time upon decision of a majority of the electors in the manner established by law." These are high sounding words. In practice the power was employed as a method by which the Communist party could purge its own ranks.

About a dozen states in the United States have the system of the recall (*q.v.*) for controlling the activity of members both of the legislative and of the administration. It is actually used chiefly against local officials and arouses intense electoral interest.

(H. Fl.; H. F. Gl.)

Voting and Nonvoting.—While some groups have boycotted voting, groups generally have felt that the ballot is a weapon which they could wield to their own advantage. Most groups have felt that they might protect themselves better from hostile sections of the population or even from a majority of the population, that they might satisfy their desires to direct the policies of the state and that they might obtain their share of the material and emotional satisfactions that can result from political action if they go to the polls and vote.

One argument always presented prior to the extension of the suffrage was that the groups to be included in the electorate did not desire the vote and would not exercise the right if it were given to them. This was the plea made by conservatives in the United States, Great Britain, Belgium and Prussia prior to the adoption of important suffrage reforms.

What is the relationship of the size of the vote and the efficient functioning of the democratic process? Voting is a means for obtaining the consent of the governed. Other things being equal, the smaller the vote the more limited becomes the basis of consent. A low poll accompanied by widespread skepticism regarding the efficacy of the electoral process is a danger signal. It indicates that the masses do not have much faith in the democratic

process. Some maintain that the citizens who will not even go a short distance to the polls to mark their ballots may not be relied upon to put forth the strenuous efforts necessary to defend democratic institutions in troublesome times.

The plutocratic Athenian democracy had to pay citizens to attend the assembly meeting and vote—an indication of declining popular morale. According to Charles Seymour and D. Frary, *How the World Votes* (1918), the groups enfranchised by the French Revolution took slight interest in voting during the years before Napoleon. Participation in elections has not been a confirmed habit in several South American countries, among them Brazil. In the United States, the southern states, which are confronted with peculiar and burdensome problems, economic and political, offer striking examples of low participation.

The indifference of the Italians to the fate of parliamentary government prior to 1922 was one of the foundations of Fascism. The polls in Italy were low during the half century preceding the so-called march on Rome. Nonvoting in Italy was in part the result of the boycott of the Italian state by the Catholic Church, but it was also the product of a feeling of hopelessness on the part of the masses. If the Italian Çatholics had been rallied to the polls by a strong party such as the Belgian Catholic party, they might have stabilized the parliamentary system. When the ban on political action by Catholics was lifted by the Vatican the Populari party was founded in 1919, but it was not given a chance to prove its worth before Mussolini seized power.

Nonvoting is also important because it indicates the presence of inert persons who may be captivated by ruthless demagoguery. If the parties that wish to respect the democratic processes are not successful in winning the confidence of these indifferent citizens, then the antidemocratic parties are in a better position to bid for their support. Persons who feel that the existing regime has little to offer them may be won over by a group which makes wild promises regarding the future.

An examination of the situation in Germany in the 1930s shows that the nonvoters were of importance in Adolph Hitler's rise to power. During the 1920s interest in voting in Germany had been reasonably high, but it tended to decline. One-fourth of the eligible German voters failed to take part in the *Reichstag* election of 1928. The early 1930s brought larger polls and the chief beneficiary of the augmented interest was the Nazi party which appealed especially to those who felt frustrated under the Weimar regime. Hitler promised everything to everybody in true demagogic fashion. While the election of March 5, 1933, was not as free as the preceding elections had been, it was nevertheless more free than the following elections under Hitler. Prior to this election the Nazis were in violent clashes chiefly with the Communists. In absolute numbers the older parties did not decline greatly in this election. The vote for the Social Democratic party was practically the same as it had been in the two 1932 elections. The 5,000,000 votes gained by the Nazi party could be accounted for largely out of the increased poll of about 4,000,000. The glittering promises of Hitler and his aides awakened hopes in many persons who had shunned the electoral process before.

There are circumstances under which a low poll is not particularly alarming as far as the proper functioning of democracy is concerned. In fact, a low poll sometimes indicates that the electoral process is working very efficiently, for example, in the U.S. cities under the city-manager plan which have councils made up of leading citizens whose control of the city affairs is on the whole very acceptable. The election of the city council in these cities will not bring out the vote that a presidential election does. As long as the council proceeds along the traditional lines and does not raise upsetting issues, as long as there are no powerful interest groups which challenge the rule of the "leading" citizens, as long as the *status quo* is satisfactory, nonvoting may not be a particularly disturbing phenomenon. The efficient city-manager government may lull the less alert citizens to sleep and they may be unaware of their voting power, but this situation is certainly not so potentially upsetting as when nonvoting is accompanied by disillusionment. Electoral indifference becomes important in connection with the attitudes held toward the democratic process.

Voting or nonvoting may register insecurities or satisfactions depending upon the circumstances. The size of the vote under favourable conditions indicates something about the worthwhileness of voting. If there are too many elections and too many choices to exercise at each of them, many voters will ignore what they think are minor offices and minor issues. Long ballots and complicated procedures often mean the difference between voting and nonvoting.

The older English-speaking countries with democratic procedures possessed voting records inferior to the European democracies of the mid-20th century. Universal suffrage seemed to work remarkably well in Belgium, Switzerland and the Netherlands, as far as interest in voting was concerned. While the U.S. of the 1920s and 1930s fretted about the "vanishing voter" and the "millions of Americans who failed to vote," the smaller European democracies showed how the voters could be organized—in the case of Switzerland, even through the cataclysmic years of World War II. Contributing to the sustained interest in voting in these countries, it was felt by some critics, was the system of proportional representation, the convenient voting arrangements and, in the case of Belgium, an effective system of compulsory voting. Australia is the only English-speaking country employing compulsory voting, and the device has been just as effective there as it has been in Belgium. Money fines are assessed against qualified voters who fail to show up at the polls without proper justification such as sickness or absence on account of business.

Nonvoting is more prevalent among certain portions of the electorate than it is among others. The available figures on popular participation in elections in various democracies show that nonvoting is more common among women than among men, among the lower-income groups than among the upper, among the young voters than among the middle-aged, among the illiterate than among the highly educated and among the newer immigrant groups than among the natives. As far as women are concerned, they have steadily increased their voting efficiency. As immigrant groups are assimilated into their new political environment, they improve their voting records.

See VOTING BEHAVIOUR; *see* also references under "Electoral Systems" in the Index volume.

BIBLIOGRAPHY.—K. Braunias, *Das parlamentarische Wahlrecht* (1932); C. H. Titus, *Voting Behavior in the United States* (1935); H. Tingsten, *Political Behavior* (1937); H. F. Gosnell, *Democracy: the Threshold of Freedom* (1948); W. S. Woytinsky and E. S. Woytinsky, *World Commerce and Governments: Trends and Outlook* (1955); I. Jennings, *The Approach to Self-Government* (1956). (H. F. GL.)

ELECTORS (Ger. *Kurfürsten*), a body of German princes with whom rested the election of the German king from the 13th until the beginning of the 19th century. Their importance was the result of the peculiarities of political development which left their mark even on 20th-century Germany. In the early period of German history rulers succeeded to their position partly by heredity and partly by election. Primitive practice among all the Germanic peoples of western Europe, including the Anglo-Saxons, had emphasized the element of heredity. Even from the first, however, there had been some opening for election, for the principle of primogeniture was not observed and there might be several competing candidates, all of royal blood. One of these competing candidates would have to be recognized, and to this limited extent Teutonic kings may be termed elective from the very first. But whereas in the other nations of western Europe this element of election dwindled and the principle of heredity secured legal recognition, in medieval Germany the principle of heredity sank formally into the background, and legal recognition was finally given to the elective principle.

This difference between the German monarchy and the other monarchies of western Europe may be explained by various considerations. Not the least important of these is what seems a pure accident. Whereas in France the Capetian monarchs, during the 300 years that followed on the election of Hugh Capet in 987, always left a male heir, and often a male heir of full age, the German kings again and again, during the same period, either left a minor to succeed to the throne or left no issue at all.

Again the strength of tribal feeling in Germany made the mon-

archy into a prize, which was not the appanage of any single tribe but circulated, as it were, from Franconian to Saxon, from Saxon to Bavarian, from Bavarian to Franconian, from Franconian to Swabian; while the growing power of the nobility and its habit of erecting antikings to emphasize its opposition to the crown coalesced with and gave new force to the action of tribal feeling. In addition, the fact that the German kings were also Roman emperors consolidated the growing tendency toward the elective principle. The principle of heredity had never held any great sway under the ancient Roman empire (*see* under EMPEROR); and the medieval empire, instituted as it was by the papacy, came definitely under the influence of ecclesiastical prepossessions in favour of election.

Heredity might be tolerated in a mere matter of kingship; but the precious trust of imperial power could not be allowed to descend according to the accidents of family succession. Hence, after a period in the 10th and 11th centuries when the practice of hereditary succession seemed on the point of prevailing, a turning point came with the great struggle of empire and papacy known as the investiture contest. Pope Gregory VII revived the theory of electoral monarchy; and from the election of the antiking, Rudolph of Rheinfelden, in 1077, it was never again repressed, in spite of the efforts of the Hohenstaufen rulers to consolidate their dynasty.

The accessions of Conrad II, Lothair II, Conrad III and Frederick I were all marked by an element, more or less pronounced, of election. Particularly important was the election of Lothair in 1125, for by then there was visible an electoral body in process of formation. On this occasion ten princes were selected from the princes of the various duchies, to whose choice the rest promised to assent, and these ten selected three candidates, one of whom, Lothair, was finally chosen (apparently by the whole assembly) in a somewhat tumultuary fashion. In this case the electoral assembly seems to have been, in the last resort, the whole diet of all the princes. But a *de facto* pre-eminence in the act of election was already, during the 12th century, enjoyed by the three Rhenish archbishops—probably because of the part they afterward played at the coronation—and also by the dukes of the great duchies—possibly because of the part they also played, as vested for the time with the great offices of the household, at the coronation feast. Here without doubt is to be found the core of the future electoral body.

The next and decisive stage came with the conflicts between 1198 and 1212 and the chaos of the great interregnum (1250–73). The decay of the great duchies and the narrowing of the class of princes into a closed corporation, some of whose members were the equals of the old dukes in power, introduced difficulties and doubts into the practice of election which had been used in the 12th century. The contested elections of the interregnum of 1198–1212 brought these difficulties into strong relief. The famous bull of Innocent III (*Venerabilem,* 1202), in which he decided for Otto IV against Philip of Swabia on the ground that, though he had fewer votes than Philip, he had a majority of the votes of those principally concerned in the election, made it almost imperative that there should be some definition of these principal electors. The most famous attempt at such a definition is that of the *Sachsenspiegel* ("Saxon Mirror") which was followed, or combated, by many other writers in the first half of the 13th century. Eventually the contested election of 1257 brought light and definition. Here were seven potentates acting (the same seven whom the Golden Bull recognized in 1356); and these seven were described in an official letter to the pope as the seven princes having a vote in such elections. The doctrine thus enunciated was at once received; and by the date of the election of Rudolph of Habsburg (1273) the seven electors may be regarded as a definite body, with an acknowledged right; the pope having already acknowledged it in two bulls (1263).

But the definition of the body and the acknowledgment of the right were still imperfect. The composition of the electoral body was uncertain in two respects. The duke of Bavaria claimed as his proper right the electoral vote which had been assumed by the king of Bohemia; and the practice of partition in electoral families tended to raise further difficulties about the exercise of the vote. The Golden Bull of 1356 settled both these questions: Bohemia (of which Charles IV, the author of the Golden Bull, was himself the king) was assigned the electoral vote in preference to Bavaria; and a provision annexing the electoral vote to a definite territory, declaring that territory indivisible, and regulating its descent by the rule of primogeniture instead of partition, swept away the old difficulties which the custom of partition had raised. After 1356 the seven electors were regularly the three Rhenish archbishops, Mainz, Cologne and Trier, and four lay magnates, the palatine of the Rhine, the duke of Saxony, the margrave of Brandenburg and the king of Bohemia; the three former being vested with the three archchancellorships and the four latter with the four offices of the royal household.

The rights of the seven electors, in their collective capacity as an electoral college, were a matter of dispute with the papacy. The result of the election was in itself simply the creation of a German king. But from 962 the German king was also, after coronation by the pope, Roman emperor. Must the elected king be promoted inevitably to the imperial crown, or did such promotion depend on the discretion and subsequent action of the papacy? Boniface VIII pressed the latter view against Albert I in 1298, and John XXII expressed it in its harshest form when in 1324 he excommunicated Louis IV for using the title even of king without previous papal confirmation. This action led to a protest from the electors, whose right of election would have become practically meaningless if such assumptions had been tolerated. A meeting of the electors (*Kurverein*) at Rense in 1338 declared (and the declaration was reaffirmed by a diet at Frankfurt in the same year) that the act of election conveyed both kingship and empire without any need of papal assent. The doctrine thus positively affirmed at Rense was negatively reaffirmed in the Golden Bull, in which a significant silence was maintained in regard to papal rights. Its great achievement was to make majority decisions binding on the minority and thus to prevent double elections and take away the opportunity for papal interference. But the result was not a strengthening of the monarchy: the influence formerly exercised by the papacy was now concentrated in the hands of the electoral princes.

By the end of the 14th century the position of the electors, both individually and as a corporate body, had become definite and precise. Individually, they were distinguished from all other princes by the indivisibility of their territories and by the custom of primogeniture which secured that indivisibility; and they were still further distinguished by the fact that their person, like that of the emperor himself, was protected by the law of treason, while their territories were only subject to the jurisdiction of their own courts. Powerful as they were, however, in their individual capacity, the electors showed themselves no less powerful as a corporate body. As such a corporate body, they may be considered from three different points of view. They were an electoral body, choosing each successive emperor; they were one of the three colleges of the imperial diet (*see* DIET); and they were also an electoral union (*Kurfürstenverein*), acting as a separate and independent political organ even after the election and during the reign of the emperor.

It was in this last capacity that they had met at Rense in 1338; and in the same capacity they acted repeatedly during the 15th century. Again and again, from 1424 to 1530, attempts were made by the electoral union to erect a new central government, either composed of its members or acting under their influence and control, by the side of the emperor. There was one such attempt in 1424; another in 1453; and a third, the most ambitious of all, in which Bertold of Mainz was prominent, in 1500. But the opposition of the emperors combined with the forces of German disunion to shipwreck every attempt.

In the course of the 16th century a new right came to be exercised by the electors. As an electoral body (*i.e.,* in the first of the three capacities distinguished above) they claimed, at the election of Charles V in 1519 and at subsequent elections, to impose conditions on the elected monarch and to prescribe the terms on which he should exercise his office in the course of his reign. This

Wahlkapitulation (electoral capitulation), similar to the *Pacta Conventa* (agreement) which limited the elected kings of Poland, was left by the diet to the discretion of the electors, though after the treaty of Westphalia an attempt was made to turn the capitulation into a matter of legislative enactment by the diet. From this time onward the only fact of importance in the history of the electors is the change which took place in the composition of their body during the 17th and 18th centuries. From the Golden Bull to the early years of the 17th century the composition of the electoral body had remained unchanged. In 1623, however, in the course of the Thirty Years' War, the vote of the count palatine of the Rhine was transferred to the duke of Bavaria; and at the treaty of Westphalia the vote, with the office of imperial butler which it carried, was left to Bavaria, while an eighth vote, along with the new office of imperial treasurer, was created for the count palatine. In 1708 a ninth vote, along with the office of imperial standard-bearer, was created for Hanover. Finally, in 1778, the vote of Bavaria and the office of imperial butler returned to the counts palatine, as heirs of the duchy, on the extinction of the ducal line; the new vote created for the Palatinate in 1648, with the office of imperial treasurer, was transferred to Brunswick-Lüneburg (Hanover) in lieu of the vote which this house already held.

The perquisites and prerogatives were valued to the end; the constitutional importance of the electoral college, on the other hand, had declined. The fact that the house of Habsburg had established a *de facto* claim to the imperial crown made the electors' basic right nugatory; and the efforts of Mazarin and Louis XIV, after 1658, to stir them into anti-Austrian action failed. The rise of Prussia in the Seven Years' War brought a new constellation of forces in Germany, in which the electors counted for little. The French Revolutionary Wars and secularization of the ecclesiastical principalities (1803) destroyed the old order, in which they had risen to prominence. In 1806, on the dissolution of the Holy Roman empire, the electors ceased to exist. *See* also EMPIRE.

BIBLIOGRAPHY.—M. Krammer, *Quellen zur Geschichte der deutschen Königswahl und des Kurfürstenkollegs* (1918), *Das Kurfürstenkolleg von seinen Anfängen bis . . . 1338* (1913); H. Mitteis, *Die deutsche Königswahl. Ihre Rechtsgrundlagen bis zur Goldenen Bulle* (1938); C. C. Bayley, *The Formation of the German College of Electors* (1950). *See* also J. Bryce, *Holy Roman Empire;* A. Meister, *Deutsche Verfassungsgeschichte,* 3rd ed., pp. 150–163 (1922); and R. Schröder, *Lehrbuch der deutschen Rechtsgeschichte,* pp. 471–481, 819–820 (1919).
(E. B.; G. BH.)

ELECTRA ("bright one"), the name of three Greek legendary figures.

1. Electra, daughter of the god Oceanus and sister of Styx, river of the underworld, was the wife of Thaumas, monstrous son of the earth (Gaea) by her son Pontus, the sea. By Thaumas, Electra was mother of the rainbow (Iris) and of the Harpies, who came to be stylized as female birdlike monsters, snatchers and defilers of food.

2. One of the Pleiades (*q.v.*), the second Electra was mother by Zeus of Dardanus, the ancestor of the Trojan royal family. Some accounts made her mother of the Cabiri, who protected sailors, or of Iasion, lover of the goddess Demeter. Because of Electra's grief at the destruction of Troy, the star of the Pleiades that represented her lost its brilliancy; or she was changed into a comet, its tail representing her streaming hair.

3. Electra, the daughter of Agamemnon and Clytaemnestra, saves the life of her young brother Orestes by sending him away when their father is murdered; when he is grown up and returns, she helps him to slay their mother and their mother's lover. Electra then marries Orestes' friend Pylades. The plays of Sophocles, Aeschylus and Euripides (*qq.v.*) vary this theme in detail. Many other versions, some very complicated, exist.

BIBLIOGRAPHY.—Aeschylus, *Choephoroe;* Sophocles and Euripides, *Electra;* Euripides, *Orestes. See* especially the introduction to Sophocles' *Electra* by R. C. Jebb for an account of the handling of the theme in ancient and modern literature. *See* also H. J. Rose, *A Handbook of Greek Mythology* (1933).
(H. W. PA.)

ELECTRICAL INSTRUMENTS: *see* INSTRUMENTS, ELECTRICAL MEASURING.

ELECTRIC (HOUSEHOLD) APPLIANCES: *see* HOME EQUIPMENT.

ELECTRIC CHARGE. An electrified body is said to have acquired an electric charge (*see* ELECTRICITY; *The Particles of Electricity: Protons*). The unit of electrical quantity is the unit charge (*see* ELECTRON: *The Electron as an Elementary Charge*).

ELECTRIC EEL (*Electrophorus electricus*), a South American fresh-water fish that looks like an eel; it has a greatly elongated body, almost round in cross section. It is, however, not closely related to a true eel (*q.v.*) but rather belongs to the largest order, the Cypriniformes or Ostariophysi, of fresh-water fishes—

BY COURTESY OF NEW YORK ZOOLOGICAL SOCIETY

ELECTRIC EEL (ELECTROPHORUS ELECTRICUS)

the order that includes the minnows, carps, catfishes, etc. The electric eel belongs to the family Gymnotidae, which also includes the knifefishes of home aquarists.

Electric eels commonly exceed a length of three feet, and are said to reach eight to nine feet, and the thickness of a man's thigh. The pelvic fins are lacking and the dorsal and caudal fins are rudimentary. The anal fin runs the whole length of the body, forward to the anus, which is under the throat.

The tail region, which constitutes about four-fifths of the entire length of the fish, encloses the powerful electric organ, which consists of longitudinal columns, each a battery, composed of an immense number of electric plates. The alternating electric current passes from the electrically positive tail end of the organ to the negative front end, and the maximum shock is given when the head and tail are in contact with well-separated points on the victim's body. The severity of the shock depends of course on several factors including the size and the state of health of the fish. In some cases, for the very brief duration of the shock, more than 500 volts at about 2 amperes (that is, 1,000 watts) have been measured.

The chief use of the electric shock seems to be to immobilize prey. The shock is powerful enough to stun a man and perhaps to kill a horse, and can be discharged at will (the electric organs are innervated by numerous branches of the spinal nerves). Like other electric fishes, such as the electric rays (*Torpedo,* etc.) of warm and temperate seas, the electric stargazers (*Astroscopus,* etc.) of warm American shores and the electric catfish (*Malapterurus electricus*) of the Nile, the electric eels are rather sluggish creatures. The electric powers have been evolved independently in each group, but, in all, the electric organs are modifications of muscle tissue.
(C. L. Hs.)

ELECTRICITY. The question "What is electricity?" was called unanswerable for several decades after electricity came into common use. Now it is answered with known facts. Electricity, of the kind which flows through wires, is composed of little particles called electrons, much lighter than atoms and all exactly alike, even in wires of different metals. More definitely, an electron weighs only $\frac{1}{1,837}$th as much as an atom of the lightest chemical element, hydrogen.

Following are the main sections and divisions of the article:

I. The Particles of Electricity
  1. Electrons
  2. Protons
  3. Ions
  4. Forces Driving Electrons in Wires
  5. Neutrons and Atomic Nuclei
  6. Ephemeral Particles
II. Amperes and Volts
  1. Electrostatic Unit of Charge
  2. Some Electromagnetic Forces
  3. The Current Balance
  4. The Accuracy of the Fundamental Laws

5. The Ampere
6. Ammeters
7. Why Do Like Currents Attract?
8. Special Relativity
9. The Conventional "Direction of the Current"

A. Electric Power
1. The Watt
2. Kilowatts and Kilowatt-Hours
3. Power and Electrostatic Forces
4. Potential Difference and the Volt
5. Transmission of Power
6. Nonelectrostatic "Voltages"

III. Direct-Current Circuits
1. Alternating and Direct Currents
2. Direct-Current Charges
3. Electron "Bookkeeping"
4. Some Evidence About Electrons

A. Ohm's Law
1. Resistance
2. Heat Produced by a Current
3. Resistivity
4. Other Terms and Units

B. Batteries
1. The Ordinary Storage Battery
2. Why Batteries Are Series
3. Why Long Power Lines Have High "Voltages"
4. Battery-and-Generator Circuits
5. Some Other Battery Cells
6. Electroplating
7. Faraday's Law of Electrolysis
8. The Algebra of Battery Circuits

IV. Electrostatics

A. Charges Without Currents
1. Test Charges
2. Low-Current, High-Voltage Apparatus

B. Coulomb's Law
1. Coulomb's Law in Coulombs
2. The Restriction to "Free Space"

C. Some Corollaries of Coulomb's Law
1. The Potential-Difference Law
2. Charges on Surfaces
3. The Accuracy of Coulomb's Law
4. Electrostatic Shielding and the Conservation of Charge
5. Point Charges, Straight Wires and Flat Plates

D. Measuring a Force on One Electron
1. Best Value for the Electronic Charge

E. Condensers (Capacitors)
1. Condensers as Analogues of Springs
2. Capacitance
3. Insulating Materials in Condensers
4. Polarization Charges
5. The Specific Inductive Capacity
6. Mechanisms of Dielectric Polarization

F. Electrostatic Fields
1. Special Cases
2. Lines of Force
3. Some Uses of Lines of Force
4. Lines and Field Strengths
5. Electrostatic Fields With Direct Currents

G. Radio Tubes
1. The Diode
2. The Coolidge X-Ray Tube
3. Space Charges
4. The Triode
5. Other Tubes

V. Magnetic Fields of Direct Currents

A. Exploring the Magnetic Field
1. The Exploring Coil
2. The Conventional Direction of the Magnetic Field
3. The Direction of the Earth's Field
4. A Magnetic "Line of Force"
5. The Field of the Circular Coil

B. Some Other Magnetic Fields
1. Lines of Force Around a Straight Wire
2. The Solenoid
3. Magnetic Lines, Endless but Not Closed

C. Flux Densities
1. Measuring the Flux Density
2. More About the Force on a Single Wire
3. Lines and Flux Densities
4. Biot and Savart's Law
5. Some Other Flux Densities
6. The Two Laws as Partners

D. Masses of Electrons and Ions
1. Cathode Rays
2. Positive Rays
3. Measuring the Speeds and Masses
4. Masses of Isotopes

VI. Magnetic Effects of Iron

A. Amperian Circuits
1. Oersted and Ampère
2. Barnett's Discovery
3. Amperian Currents

B. Co-operation of Currents
1. Contrasting Iron With Glass
2. Solenoid and Plunger
3. Motors
4. Relative Strengths

VII. Induced Electromotive Forces
1. Induction and Flux Density
2. A Question of Units
3. Lenz's Law
4. Induction With Iron

A. Magnetization of Iron
1. Mapping a Field With Iron Filings
2. The "Magnetic Field Strength"

B. Transformers and Alternating Currents
1. Alternating Currents
2. Step-Up and Step-Down Transformers
3. Primary and Secondary Currents
4. Lagging Currents and Power Factors
5. Reactance and Impedance
6. Eddy Currents
7. Self-Inductance
8. Resonance and Tuning

C. Moving Wires, Moving Magnets and Their Relativity
1. The "Flip Coil"
2. The Lorentz Force
3. Pros and Cons of Moving Lines
4. Induced Electric Fields
5. The Relativity of Moving Coils

D. Magnetic Flux
1. Defining Magnetic Flux
2. Flux Changes and Electromotive Forces
3. Electromotive Forces in Motors
4. Generators
5. Rotating Loops
6. Rotation, Relative to What?
7. Why Specify Wire?
8. What Is Basic About Flux?

VIII. Electromagnetic Waves

A. Maxwell's Theory
1. Open-Circuit Currents
2. Maxwell's "Displacement Currents"
3. The Fields Are Set Free
4. Maxwell's Plane Waves
5. Comparing the Waves With Light

B. Hertz's Experiments
1. Hertz's Oscillator
2. Hertzian Waves

C. Wireless
1. Wireless Telegraphy
2. Wireless Telephony and Broadcasting
3. Television
4. Radar

D. Waves and Quanta
1. The Photoelectric Effect
2. Quanta

I. THE PARTICLES OF ELECTRICITY

1. Electrons.—One of the most important known facts about electricity is that which tells how many electrons flow past any point in a wire carrying any specified current. To specify the current one may use the common unit, the ampere. This unit will be defined exactly later, but for the present it may be imagined roughly as the current carried by a medium-sized incandescent lamp. With a current of exactly one ampere, the number of electrons passing any point in one second is 6,242,000,000,000,000,000,[1] or, as physicists say, 6.2420×10^{18}. Further studies may raise or lower this number, but not by more than a few units in its fifth digit.

The number of electrons in the wire, likewise, can be stated accurately. If the wire is made of copper, for example, each atom contains 29 electrons, and there are 1.385×10^{24} atoms per cubic inch; and corresponding figures can be given for any metal.

Most of these electrons take no part in carrying the current, but stay in the atoms where they are, while one or two electrons from each atom move freely through the metal as carriers, or free elec-

[1]This constant, as well as a number of others in this article, is taken from *Fundamental Constants of Physics* by E. R. Cohen, K. M. Crowe and J. W. M. DuMond. Copyright 1957 by Interscience Publishers, Inc., New York.

trons. Even so, however, there are so many carriers that they do not have to flow very fast to carry any ordinary current. As individuals they move fast, to be sure, but at random; the systematic drift which constitutes a current is very slow, usually much less than an inch per second.

This slowness may be surprising in view of the speed of telegraphic signals, for which electric currents were early noted. Nevertheless it is not unparalleled.

The flow of electrons may be compared with the flow of air in a tube such as a simple speaking tube of the type often used for calling an apartment from the vestibule. When sound goes through this tube, the air at any point moves forth and back, changing its direction many times per second; but it never moves fast, even though the sound waves and their messages make 760 m.p.h.

In the same way, electrons in a wire never need to flow fast, even though the waves in their flow make telegraphic speeds. (*See* ELECTRICITY, CONDUCTION OF; ELECTRON.)

Repulsions Between Electrons.—Another of the basic laws about an electron relates to the force it exerts on any other electron. This force varies, both with the distance between the electrons and with their states of motion. Restricting ourselves at first to electrons at rest, and to distances longer than the diameter of an atom, the fundamental law about this force may be stated as follows: (1) every static electron repels every other one; (2) the strength of this repulsion varies in inverse proportion to the square of the distance between them; (3) and at a distance of exactly one inch this repulsion is 0.000,000,000,000,000,000,000,000,080,38 lb., or 8.038×10^{-26} lb., with no chance of error in the last digit here. This force is far too small to measure, or even to detect; but by combining the repulsions of vast numbers of electrons it is easy to produce measurable electrostatic forces.

2. Protons.—Indeed, if electrons were the only particles of electricity, the forces would be disastrously strong. With the numbers given above for the electrons in copper, a little arithmetic shows that there must be about 4×10^{21} electrons in a cube only 0.1 in. on each edge. A similar group, an inch away, would repel this group with a force 16×10^{42} times as strong as the force between single electrons, and this would be about 600,000,000,000,000 tons. Clearly, if such forces were unopposed, any piece of copper large enough to be seen would be torn to pieces. Since there must be opposing forces, copper must contain particles of some other kind.

These other particles are of a type which do not flow through a wire, for reasons to be explained later, and are called protons. Electrostatically, a proton exactly balances an electron, in the sense that a proton attracts an electron exactly as strongly as another electron at the same distance repels it. Very exactly, indeed: the accuracy with which electricity obeys simple laws is one of its most remarkable characteristics; and this relation between electrons and protons is probably the most exactly known of all physical laws.

Protons, like electrons, repel each other, and the law of their repulsion is exactly like that of electrons.

Charged Bodies.—Combining all these forces, therefore, if each of two bodies contains exactly as many electrons as protons, and the electrons are distributed through the body essentially like the protons, there is no resultant electrostatic force between them. On the other hand, if each body contains more electrons than protons, again similarly distributed, there is a resultant repulsion; likewise if each contains an excess of protons. But if one contains an excess of protons and the other an excess of electrons, there is an attraction. Such excesses of electrons or protons are called electric charges, and these bodies are called charged, or electrified, bodies.

In practice, these resultant electrostatic forces are usually very weak. If we try to make them as strong as possible, with apparatus in air (not in oil or in very high vacuum), a limit is set by the air's breaking down and discharging the bodies through an electric spark.

In air, pieces of paper can be picked up by electrostatic attractions, but coins cannot. Such weak forces are far indeed from the 3×10^{14} tons calculated for the electrons in those 0.1-in. copper cubes.

Therefore, the excesses or deficiencies of electrons on bodies of visible size in air are extremely minute in comparison with the numbers of electrons in these bodies.

Contrast Between Protons and Electrons.—Although the proton electrostatically is an exact counterpart of the electron, mechanically it has no such complementary relation. A mechanical difference is suggested by the fact noted above that protons do not flow when a wire carries a current.

The simplest relation between a proton and an electron is found in the atom of ordinary hydrogen (not heavy hydrogen), which consists of just these two particles. The electron, as noted already, is only $\frac{1}{1,837}$th as heavy as this atom, so all the rest of the weight of the atom is that of the proton.

This contrast results in the proton's staying almost at rest, while the electron moves rapidly in various ways around it. Evidence on these motions appears whenever lone hydrogen atoms, not in chemical combination, are made to emit light, as they can be in a tube like a neon sign, containing nothing but rarefied hydrogen. This light comes almost entirely from motions of the electrons, so by spectrum analysis and other tests it reveals all the essential features of the motions. Electricity, though popularly considered invisible, is all that is visible here.

And not only here. Substances made of polyatomic molecules emit light with spectra much more complicated than that of monatomic hydrogen. Many of these spectra contain lines which result from vibrations of atoms, as well as lines which are caused by electrons. But vibrations of atoms show only in the infrared; that is, beyond the red end of the part of the spectrum that we can see. Electrons, far from being invisible, are practically the only things that are visible.

Nevertheless, it is true that electric charges are invisible. A body looks no different when charged with excess electrons than when uncharged or even when deficient in electrons. But as noted above, the numbers of electrons in such excesses or deficiencies are very minute in comparison with the vast numbers of electrons in all bodies. Conversely, the invisibility of charges is simply a confirmation of this rule of comparative minuteness. The fact remains that practically all we ever see of anything is its electrons.

In any element but hydrogen, each atom contains more than one proton. A copper atom, for example, contains 29. Indeed, this is why there are 29 electrons per atom in a copper wire; if each atom contained only 28 electrons with its 29 protons, or if each had 30, any piece of copper of visible size would be torn apart by electrostatic repulsions.

This statement implies that the number of protons in the copper atom is fixed at 29, even if the number of electrons changes. This implication is correct, although the explanation of this fixity must be postponed; and it is in line with the fact that electrons can flow through copper, thus changing their numbers in some atoms, while protons cannot.

3. Ions.—Chemical action can change the electron numbers in many atoms systematically. If copper, for example, is converted from its metallic form to the salt cupric chloride, $CuCl_2$, each copper atom keeps its 29 protons but it loses an electron to each of the two chlorine atoms, leaving only 27. Meanwhile each chlorine atom, which contains 17 protons, gains an electron, raising the electron number of each to 18. The three atoms together contained 63 protons and 63 electrons, and they still do; so there is no excess of particles of either kind to tear a piece of the salt apart. On the contrary, the charges in the individual atoms cause each copper atom to attract all chlorine neighbours (not merely two) and vice versa. Thus electrostatic attractions are the chief source of mechanical strength in the crystals of many of the simpler salts.

One of the best examples of this, easier to picture than cupric chloride, is common salt or sodium chloride, NaCl. Here each atom has an excess of only one particle (sodium having 11 protons and, in salt, only 10 electrons) and each atom of either element has six nearest neighbours, all of the other element.

Currents in Molten Salts and Water Solutions.—When a salt is heated to the melting point, the thermal agitation is increased to a point where the electrostatic forces and any others (and there are others in many salts, such as silicates) are unable to keep atoms from exchanging places and wandering around. Charged atoms are then called ions (Greek for "wanderers"). Their wanderings can be given preferred directions by the use of electrically charged metal plates, called electrodes, as shown in fig. 1.

A flow of ions of either kind, or of both in opposite directions, is equivalent to a flow of electrons; but it introduces complications not found in wires. If an end of a copper wire is connected to an iron wire, electrons flowing through the copper can flow right on through the iron, because they are just as much at home in one as in the other; but when the chlorine ions, flowing through this molten sodium chloride, come to electrode P, they cannot flow on through it. Within it the current can be carried only by its free electrons. If electrons are to flow steadily through the surface of electrode P, therefore, they must be carried to this surface as the excess electrons in chlorine ions, and away from it as free electrons within the metal. That is, at the surface the electrons must take leave of the chlorine ions; so these ions, losing their charges, become electrically neutral atoms and join themselves into pairs as molecules of chlorine gas. This gas, therefore, accumulates on the surface of electrode P.

There is an "if" here: if electrons are to flow steadily through this surface. Some of them may not flow through it, but only from it, starting from surface atoms of the metal. Such atoms, left deficient in electrons, become ions and dissolve off into the molten salt. The more reactive the metal is, the more the surface reaction takes this form. With any metal, however, there is at least one reaction at this surface, whereas there is none at a junction between two wires.

At the surface of electrode E there is another reaction. Within electrode E, its free electrons are flowing toward the surface, to meet sodium ions attracted toward it. These ions, being deficient in electrons, capture them at the surface and accumulate as neutral metallic sodium. This is, in fact, one of the easiest ways to reduce this highly reactive element to metallic form.

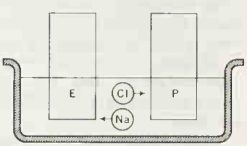


FIG. 1.—FLOW OF IONS IN MOLTEN SODIUM CHLORIDE

Electrode E and chlorine ion Cl have excess electrons; electrode P and sodium ion Na are deficient in electrons (arrows show the directions in which the ions are driven by electrostatic forces)

From the Greek verb *luein,* "to loosen," the name given to such conduction as by this molten salt is "electrolysis," the molten salt is an electrolyte, and the current electrolyzes it. The whole assembly shown in fig. 1 is called an electrolytic cell.

Water solutions of such salts also are electrolytes, and so are solutions in liquid ammonia and some other solvents. In them, the reactions are often essentially like those in molten salts, but with further complications. With sodium chloride in water, for example, and iron electrodes, at electrode P the reaction products include chlorine gas and ferrous chloride, as one might guess from what has been said above; but at electrode E, since metallic sodium would react instantly with water, the products are hydrogen gas and sodium hydroxide.

Altogether electrolysis runs into many complications; therefore further discussion will be postponed till after that of other fundamentals.

Currents in Gases.—Another sort of conduction involving ions occurs in gases; for example, through neon in a neon sign or through the air in an electric spark. These currents obey laws very different from those of electrolysis.

Gases are not habitually ionized like electrolytes. Cool air is an excellent insulator. In a spark, however, the air is a fairly good conductor because the high temperature causes violent impacts between molecules, and these impacts drive electrons out of some of the molecules, making them into ions. Cool neon, likewise, is a good insulator until violent impacts ionize it.

In both cases, a few ions are present before the current starts. Even these few are the result of something locally violent, such as a cosmic ray or a stray atom of some radioactive element. Ordinarily there are too few of them to conduct more than a bare trace of current. To start a spark, they must be made to increase their number greatly. Each of these few ions must be driven to exceptionally violent impacts. The spark starts, therefore, only when the electrostatic field is very much stronger than it will need to be later, when the air has become hot.

In the neon sign there is less difference of this sort. Conditions for different cases of conduction in gases vary greatly. Historically the study of currents in gases has been important; it led, for example, to the discovery of X-rays and of the electron. Practically, also, it is important to certain persons: to designers of neon tubes and fluorescent lamps, for example, and to power-line engineers, who have to design against sparking and even against quieter losses of power. For present purposes, however, these details are not so important as the broader principles of electricity.

The points about gas ions deserving attention here are these: first, the points already stated; and second, that, whether a gas ion is a molecule deprived of an electron or a molecule to which a free electron has attached itself, when the ion reaches an electrode and is neutralized, it does not stay there. In short, gas ions are quite different in their habits from ions in electrolytes.

4. Forces Driving Electrons in Wires.—Just as ions are driven through a molten salt by electrostatic repulsions and attractions, so are electrons in wires. Sometimes, to be sure, these electrons are subject also to other forces, such as electromagnetic forces acting in the moving wires of a motor or a generator; but with or without such forces, electrostatic forces always are present and fundamentally important.

Their importance arises from the fact that it is through these forces that power is transmitted. When a lamp filament glows, the power to keep it hot is delivered to it by the electrostatic forces on its electrons; that is, by repulsion on them by one end of the filament, which has an excess of electrons, or attraction by the other end, which has a deficiency of them, or both.

It may seem surprising that these forces can deliver enough power to keep a lamp hot, especially in view of the statement made above, that electrostatic attractions in air can pick up a piece of paper but not a coin. The key to this paradox is that the number of electrons flowing through the lamp in every second is enormous, in comparison with the little excess or deficiency which the paper or coin can hold in air.

If the charges on the ends of the filament were not maintained by continual pumping of electrons against electrostatic forces at the source of power (generator, battery or whatever) the current through the lamp would draw off the excess electrons and fill up the deficiency in less than a millionth of a second. These charges do not contain any great reserve of energy, but they are of basic importance in understanding how the lamp, or any other electrical appliance, really works.

5. Neutrons and Atomic Nuclei.—Other fundamental questions include the much-postponed one of why protons do not flow through a wire, and the related question of why the number of protons in an atom cannot be changed by chemical reactions or any ordinary electrical phenomena.

The difference in weight, and therefore also in inertia, between the proton and the electron is suggestive on these points but is not the whole cause. The rest of the cause is the existence of another particle, the neutron.

The Neutron.—This particle, as its name implies, is electrically neutral. That is, it exerts no electrostatic force on either a proton or an electron. Nevertheless, at extremely short distances (of the order of 10^{-13}, or a tenth of a millionth of a millionth, of an inch) a neutron does exert forces of other types on a proton or another neutron. These forces bind protons and neutrons together into groups, and these groups are atomic nuclei.

The Nucleus.—Every atom has just one nucleus, and all its protons are therein, although its diameter, roughly 10^{-13} in., is only about $\frac{1}{100,000}$th of the diameter of the vaguely defined cloud

of electrons around it which is the chemically and optically active part of the atom. Moreover, the forces exerted on these protons by the neutrons are so strong that the amount of energy required to remove a proton from any ordinary nucleus is millions of times the energy released in any chemical reaction. This is why the number of protons in an atom is never changed by a chemical reaction, nor by any electrical apparatus less violent than a cyclotron.

This explains also why protons do not flow through a wire. The wire is solid; that is, its atoms are not wandering around, as they would if it were melted; so its protons, being tied into the nuclei of the atoms, can neither wander nor flow.

Numbers of Neutrons.—A few more points, almost off the subject of electricity, should be added here. These relate to numbers of neutrons. In almost every nucleus, the number of neutrons is at least equal to the number of protons. Heavy hydrogen (or deuterium), for example, has 1 proton and 1 neutron; ordinary helium, 2 and 2; ordinary oxygen, 8 and 8; ordinary sodium, 11 protons and 12 neutrons. A nucleus of chlorine has 17 protons; most chlorine nuclei have 18 neutrons, but some have 20. Uranium has 92 protons and usually 146 neutrons, though in the famous U^{235} it has only 143 (or 235–92).

The distinguishing feature of any element is its number of protons, because this determines the number of electrons each atom will hold, as already explained; thus indirectly but certainly the number of protons determines its chemical properties, regardless of the number of neutrons. Different varieties of any one element, with different numbers of neutrons, are called isotopes. The weight of a neutron is very near that of a proton (just barely greater), so the sum of the numbers is practically the atomic weight of the isotope. Hence isotopes are designated by superscripts for these approximate weights; *e.g.*, U^{235}, U^{238}, Cl^{35}, Cl^{37}.

A tendency, illustrated by the examples given above, is for the excess of the neutron number over the proton number to be greater, even on a percentage basis, in heavy elements than in light. This is connected with the electrostatic repulsion between protons. The more protons there are, the harder it is to hold each one of them in the nucleus, and the more extra neutrons are needed.

Finally, if there are too many protons, especially with only barely enough neutrons to hold them, the nucleus may eventually split into two nearly equal pieces, which drive each other away electrostatically with great violence. U^{235}, with its 92 protons and 143 neutrons, is more liable to this catastrophe than U^{238}, with 92 and 146; and so is plutonium, Pu^{239}, with 93 and 146. This is how U^{235} and Pu^{239} make possible the atomic bomb. (*See* Atom; Neutron; Nucleus.)

6. Ephemeral Particles.—In addition to the particles described above, there are some of other types which may be dismissed briefly but not without recognition of their existence. The reason for dismissing them is that they do not take part in any phenomena of the general class to be discussed in this article; and the reason for recognizing them is that some of them are particles of electricity and therefore could not logically be refused recognition.

Positrons.—A positron illustrates these ideas well. It is a particle exactly like a proton in its charge, but like an electron in its mass. It may appear suddenly, along with an electron, when matter is subjected to X-rays which have been produced in an X-ray tube operated at more than 1,000,000 v. Both these particles, when they appear, are moving at extremely high speeds in different directions. Both lose speed as a result of forces exerted on them by electrons in atoms near which they pass. The electron thus settles down into what one may call, rather loosely, a normal way of life for an electron. The positron cannot do this. Instead, it soon comes dangerously near an electron in some atom; then both it and this electron vanish in a flash of X-rays.

How soon the positron vanishes depends on many circumstances, and especially on how many atoms there are in the neighbourhood. If it is in a solid body, such as a piece of wire, its life is very much less than a millionth of a second.

Because of this shortness of life, there are never many positrons present at any one time in such a body, even if the radiation that produces them is very intense. There are not enough of them to have any appreciable effect on such phenomena as conduction of current in the wire.

Moreover, their production is not one of the phenomena of the type under discussion in this article. The 1,000,000 v. mentioned above really is required for their production (more exactly, 1,021,950 v., according to the much-quoted equation $E=mc^2$). This number of volts is not used in ordinary electrical work; even when 250,000 v. is used in power transmission at long distances, it is not used in a way to produce X-rays. This way of producing positrons, therefore, is not a part of ordinary experience with electricity.

Neither is any other way of producing them. Unstable atomic nuclei of certain kinds, for example, emit positrons, apparently in each case by converting a proton into a neutron and a positron and ejecting the positron. These nuclei, however, exist only as a result of treatment of matter at least as violent as the production of 1,000,000-v. X-rays. And these positrons are annihilated by electrons as soon as any others; so they do not become abundant. Even though positrons can be produced in more than one way, therefore, we may dismiss them now because they are too rare to exert any appreciable influence on the phenomena to be discussed; and we may rely on their rarity because it is caused by the shortness of their lives.

Mesons.—There are other particles about as ephemeral as positrons, that are called mesons, from the Greek word *mesos*, middle, because they have masses intermediate between those of electrons and protons. There are several kinds of mesons. They have different masses, but the charge of a meson is always either zero or exactly the charge of an electron or a proton. Mesons will probably be important in the theory of forces within nuclei when there is a good theory of such forces. Outside the nuclei, however, we may dismiss mesons along with positrons.

Neutrinos.—Finally there should be some recognition of the neutrino, this name being Italian for "baby neutron." This is not a particle of electricity, because it has no charge. Also it seems to have no mass, unless perhaps when it is going at the speed of light. Its existence has been postulated for the sole purpose of balancing the equations of mechanics in calculations on the breakups of unstable nuclei that emit positrons or electrons. It has to carry away energy, linear momentum and angular momentum. Once it is gone it never does anything to any other particle, so far as we know. So it can be dismissed from further consideration in this article, though it and the positron are discussed in the articles Nucleus and Particles, Elementary.

II. AMPERES AND VOLTS

Among the technical terms most often encountered in descriptions of electrical appliances are the words "ampere" and "volt." From their context one can usually infer that the ampere and volt are units of measurement; and they are. Sometimes one might also infer that they are related like the foot and the inch; but they are not. On the contrary, they are as independent as the foot and the horsepower, so they must be defined separately.

The ampere has already been described roughly as the current carried by a medium-sized incandescent lamp, and it was stated that a current of exactly one ampere carries 6.2420×10^{18} electrons past any point in every second.

This statement was made as a part of the description of the electron above, so to take it now as a definition of the ampere seems like reasoning in a circle. On the other hand, it illustrates a very important point. It is like the statement that the flow of the Mississippi river past New Orleans is 7.8×10^{11} cu.yd. of water per year. In this, the cubic yard and the year are of equal importance; in the ampere, their places can be taken by the electron and the second. Units of flow, of either water or electricity, are composite units, made up of the more fundamental units of quantity and time. Corresponding to the cubic yard, therefore—though not measuring volume—we can use as a unit of quantity of electricity, either the electron or the ampere-second, 6.2420×10^{18} electrons.

In a way this definition of the ampere-second as 6.2420×10^{18} electrons is ideal, because the electron is a natural unit. The im-

perial standard yard, a bronze bar in terms of which the yard was defined legally in 1878, is very slowly warping and changing its length; but 6.2420×10^{18} electrons will be the same quantity of electricity throughout the ages.

Ideal as this definition is on a long-time basis, for the immediate, practical work of exact measurements with the techniques of the 1960s it is not the best. In this respect, also, electricity is like water; the number of molecules in a cubic yard of water is known as accurately as this number of electrons; but any given volume of water can be measured more exactly by reference to the bronze bar than by any test of its number of molecules.

In another aspect, however, the practical units for water and electricity are notably different. The changes in that bronze bar were very small, even in several decades, and its successor, made of platinum-iridium alloy, is even less changeable; but if we were to try to preserve a standard quantity of electricity, as we preserve these bars, we should have no success. Even with the best of insulating materials, leakage would ruin such a standard within a day. Units of quantity of electricity, therefore, have to be defined in a radically different way.

1. Electrostatic Unit of Charge.—Among such units, the one simplest in definition is called the electrostatic unit of charge, because it is defined as the quantity giving unit strength to a certain electrostatic repulsion, of the type described above under *Protons*.

The dimensions of the system used in its definition are specified in centimetres, and the force unit is a force easily obtained electrostatically with centimetre-sized bodies in air. For calculating charges on such bodies, therefore, this unit is convenient not only because of its direct relation to forces which may furnish data for the calculation but also because the charges can be expressed in terms of this unit with numbers having few digits.

These charges, however, are not so often calculated as are quantities of electricity passing through appliances such as lamps and motors. These quantities, as noted under *Forces Driving Electrons in Wires*, above, are relatively enormous. For them the ampere-second is a convenient unit; and one ampere-second equals 2,997,930,000 electrostatic units. For calculating such quantities, the electrostatic unit would be obviously inconvenient.

The electrostatic unit is so small, indeed, as to suggest that it may be the best unit for subatomic calculations. Nevertheless, for these it seems too big; an electron is only 0.000,000,000,480,286 of an electrostatic unit. Comparing this number with the one in the last paragraph, it appears that the electrostatic unit is about halfway from the practical to the subatomic. Going halfway does no good; we can shorten the latter number by calling the electron 4.80286×10^{-10} electrostatic unit, but we can just as easily call it 1.60206×10^{-19} amp.-sec. and use the more familiar unit for subatomic as well as for practical calculations.

The exact specifications of the electrostatic unit therefore will be left to the section *Electrostatics*, below, and we shall pass on now to the specifications of the ampere.

2. Some Electromagnetic Forces.—The ampere also is defined in terms of a force, but it is a force of a different kind, called electromagnetic and not yet described in this article. The electrostatic force was described guardedly as a force between electrons or protons at rest, or static, and it was noted that there would be different forces if they were in motion. It is this difference that is used in defining the ampere. To see how, we may consider first two long, straight, parallel wires, hung loosely, so they will show the presence of any forces by appropriate movements.

From what was said under *Protons*, above, about the weakness of the electrostatic forces obtainable with ordinary household voltages, it is evident that one cannot expect household lamp cords to be appreciably moved by such forces. Nevertheless, it is easy to get static charges, strong enough to move small wires, from laboratory equipment such as the "static machine" (described below under *Low-Current, High-Voltage Apparatus*). If both wires are connected to the same terminal of such a machine they are given like charges (both excesses of electrons, or both deficiencies), so they swing apart. On the other hand, if one wire is connected to one terminal of the machine, and the other wire to the other terminal, the machine draws electrons out of one wire and drives electrons into the other. Then the charges are unlike, and the wires swing together.

These tests can be made, and indeed made best, with only one end of each wire connected to any other conductor. Therefore no electrons need flow all the way through either wire, nor even part way except during the process of charging the wires. After charging, in fact, even the one connection to each wire can be removed. This leaves the charges trapped, unable to escape from the wires but still able to exert the electrostatic forces just described.

To show the electromagnetic forces mentioned at the beginning of this section, the best connections are quite different. Static charges should be made negligible by using only low voltages, and the electrons should flow as fast as they can without overheating the wires. So each wire must be connected to some other conductor at each end, and a strong current must flow through it. The simplest connection of this sort is the one shown schematically in fig. 2. As soon as the connections shown in this figure are completed, and the electrons begin to flow, the wires swing apart. This swing shows that *oppositely directed currents cause repulsion.*

To test currents having the same direction, the U-turn and one of the connections to the dry cell must be replaced with detours of wire. These detours should be made long, so that they can be far enough from the parallel wires to keep them from exerting appreciable forces. Then it is found that the wires swing together, showing that *similarly directed currents cause attraction.*

The contrast between these rules for currents and the corresponding rules for charges is often expressed in the sentence "Like charges repel, but like currents attract." The word "like" as used here has one meaning for charges and another for currents. Neither meaning would be self-evident unless the rules had been stated previously in less epigrammatic forms; so this quotation is not explanatory, but mnemonic.

The contrast well deserves mnemonic aid. Many of the laws of electrostatics and electromagnetism may be contrasted like this pair; and many, though not all, of these contrasts may be connected by fairly direct logic with this one. So this contrast is both typical and fundamental.

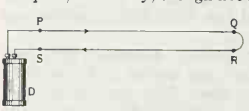


FIG. 2.—OPPOSITELY DIRECTED CURRENTS IN PARALLEL WIRE SEGMENTS PQ AND RS CONNECTED TO DRY CELL D

The electromagnetic forces between parallel wires, and their contrast with the electrostatic, were discovered in 1820 by André Marie Ampère (1775–1836). He followed up his discovery with a series of brilliant researches on wires in other shapes and geometric relations. In 1825 he collected his results in a system of mathematical laws. These laws are not nearly so simple as those of electrostatic forces. They cannot be, because currents have directions, and different parts of a circuit may have many different directions, whereas static charges have no such complications. It was a great achievement, therefore, to find laws covering the variety of cases presented by circuits. Nevertheless, all forces due to steady currents in wires, in the absence of iron or any other magnetic substance, are calculable through Ampère's laws. It is, therefore, especially appropriate that the unit of current defined in terms of such forces should perpetuate his name.

3. The Current Balance.—Any definition in terms of such a force naturally implies some provision for measuring the force, to make sure it has the strength specified in the definition. A part of the circuit, for example, may be a circular coil, lying horizontally on one pan of a good chemist's balance. If the force acting on this coil is directly up or down, it will be measurable as a change in the apparent weight of the coil when the current is turned on or off. To make the force act directly up or down, the wire exerting the force may be another coil just like the one on the balance but held fixed directly above or below it. The whole apparatus is called a current balance, or Kelvin balance, since the best-known early current balance of this type is credited to Lord Kelvin (1824–1907).

The best current balances involve extensions of the plan just

described, such as the use of more than one pair of coils; and they involve refinements and precautions, especially precautions against appreciable electrostatic forces. The essential principles, nevertheless, are just those of the parallel wires. The connections are in series (that is, such that every electron flows through all the wires and there is only one current strength to be defined), and its strength is exactly one ampere if the force registered by the balance has exactly the specified value.

4. The Accuracy of the Fundamental Laws.—The manner of specifying this force depends on certain extraordinary characteristics of Ampère's laws. These laws say nothing about the materials composing the wires, except in the negative way of excluding magnetic materials from them and their neighbourhood. Some other basic laws of electricity, such as the laws of electrostatic forces stated above, are likewise independent of the properties of specific materials; so are Newton's laws of motion, his law of gravitation and a few other laws.

These laws are in a class apart from most others, in regard to accuracy. More specifically, so long as we keep within reasonable bounds, they are so far more accurate than other laws involved with them in almost any problem that we may treat them as if their accuracy were mathematically perfect.

"Reasonable bounds" for Newton's second law of motion refer to speed. Speeds of missiles, or even of planets, are within reasonable bounds; but when Sir Joseph John Thomson (1856–1940) in 1897 found some electrons in vacuum tubes to be going a few tenths as fast as light, he began to question Newton's law. Later Hendrik Antoon Lorentz (1853–1928) found the more general law of motion, which is practically Newton's at speeds "within bounds" but which becomes infinitely different as the speed approaches that of light.

For Newton's law of gravitation things must not be too big or too heavy; the corrections to this law, made in 1915 by Albert Einstein, are very small for anything in the solar system, and only at much greater distances or with much heavier bodies do the corrections become relatively great.

For Ampère's laws, the laws of electrostatics and other laws in the group we call classical electromagnetic theory, the "reasonable bounds" are in the other direction. Atomic sizes and the vibration frequencies of light are out of bounds; but for any wires we can see, even with a microscope, the laws of classical electromagnetic theory may be used as if their accuracy was strictly perfect.

5. The Ampere.—This fact gives us an infinite variety of equally accurate ways of defining the ampere. If geometric specifications are written for any two current balances, and the force for one ampere is specified for one of them, Ampère's laws enable us to specify it for the other one with no fear of disagreement between them.

Therefore the standard definition of the ampere, adopted internationally, does not have to refer to the current balance which is most accurate in practical use. Instead, it refers to the current balance with the simplest geometry, even though this balance would be completely impracticable.

For shape, this is of course a pair of infinitely long, straight, parallel wires. For size, the separation distance simplest to specify (in the metre-kilogram-second system) is of course just one metre. The movable piece on which the force would have to be measured is a piece just one metre long. We disregard the fact that with only one ampere the force acting on this wire would be so small that we could never measure it in such big, heavy apparatus. There is no intention of even building this apparatus. It is described only as a simple prototype from which to calculate the force to be specified for any practical current balance.

The only quantity still to be specified is the force. This is given a reasonably round-number value in a unit called the newton. One newton is defined as a force which, if it could be allowed to act, without interference from other forces, on the carefully guarded piece of platinum-iridium alloy called the international prototype kilogram, would make that body pick up speed at a rate of exactly one metre per second in every second. Needless to say, no such risky treatment of that body is ever permitted. But the force unit called the kilogram is defined as the weight of that body in a place where any body falling in a vacuum (and all bodies fall alike in vacuum) would pick up exactly 9.806,65 m. per second in every second. (Such places occur near latitude 45°, north or south.) So one newton is $\frac{1}{9.806,65}$ kg. of force. One pound is 0.453,592.427.7 kg.; and these figures make the newton about 0.2248 lb. of force.

Returning with this unit to the specification of the ampere, the ampere is defined as the current in each of those infinitely long, parallel, straight wires, just one metre apart, when the force on a piece of either of them, just one metre long, is just 2×10^{-7} newton.

This is only about half a ten-millionth of a pound, so it is much too small for practical measurement with such big wires. But one of Ampère's laws is that if all dimensions of any system of wires are reduced in the same ratio, but the currents are unchanged, the forces are unchanged. Thus if the separation of the straight wires is reduced to one centimetre, the 2×10^{-7} newton will appear on a movable part of either wire, only one centimetre long.

Moreover, if their currents are increased to ten amperes the force is increased by a factor of ten because of each wire, or 100 altogether. This gives nearly five millionths of a pound, on a wire light enough to be measurably affected by it. So a small straight-wire current balance can be used, and it makes a good lecture demonstration; nevertheless the many wires in Lord Kelvin's coils co-operate to give much stronger forces, so his current balance is much more accurate.

6. Ammeters.—Everyday measurements of currents are made not with current balances but with more convenient instruments, called ammeters in abbreviation for amperemeters. Ammeters are of many types and sizes, ranging from the little, inaccurate ones on the instrument boards of cars to big laboratory-standard instruments. Common features of most of them include pointers and graduated scales, springs restraining the pointers, and means of deflecting the pointers more or less, against the springs, by electromagnetic forces.

There is great variety in the means of applying these forces, but most of them have one common characteristic. This is the absence of any attempt to make them calculable. Some involve iron, and its electromagnetic properties are notoriously difficult to handle mathematically. Others are theoretically covered by Ampère's laws, but are not geometrically simple enough for practicable calculations. Fortunately, however, these calculations are unnecessary. If an ammeter and a calculable current balance are connected in series (like the two parallel wires of fig. 2) they must carry the same current. The scale of the ammeter, therefore, can be calibrated.

The use of an ammeter, like its calibration, depends on the equality of currents in series. To measure the current in any conductor, such as a lamp or a motor, an ammeter is connected in series with it.

Of course modern accuracy of manufacture makes ammeters of any one model almost exactly alike. Only when special accuracy is needed, therefore, is it necessary to calibrate an ammeter individually. For most purposes it is sufficient to calibrate a few instruments of each model. In one way or another, however, the reading of any ammeter depends for its meaning on the parallel wires which define the ampere.

7. Why Do Like Currents Attract?—These parallel wires are of fundamental importance, not only in the definition of the ampere, but also, as noted under *Some Electromagnetic Forces*, above, in the contrast between the directions of electromagnetic forces and those of corresponding electrostatic forces. This contrast is, as noted, a prototype for a general system of such contrasts. As such, it implies a fundamental relation between the two types of forces. Indeed, there is such a relation.

This relation shows best with straight, parallel wires. With curved wires (Kelvin's coils, for example) it is obscured by various complications, especially in the electromagnetic forces; but the forces between long, straight, parallel wires are only simple attractions or repulsions. For them, therefore, the relation be-

tween electrostatic and electromagnetic forces is simply a ratio of strengths.

For the electromagnetic force with any specified number of amperes (the same in both wires) the strength can be computed directly from the definition of the ampere. The electrostatic force most closely related to this electromagnetic force is naturally the repulsion exerted by the electrons concerned in the specified currents. These are not all the electrons in the wires, but only the electrons free to flow. Even so, the repulsion between them (*i.e.*, the force which would become apparent if all protons and all other electrons could be removed) is of unimaginable strength. It is completely out of the order of magnitude of the electromagnetic attraction. Their ratio, even with the current as strong as it could be, without melting the wires, is a number with more than 20 digits.

Such a number might well seem hard to relate in any simple way to anything else. But another big ratio is right at hand. This is a ratio between two speeds. Of these speeds, the low one is the speed of flow of the free electrons when they carry the specified current. The high one is the speed made by radio waves or by light in empty space.

In terms of these speeds, the fundamental relation between corresponding electrostatic and electromagnetic forces in long, straight, parallel wires is this: *the ratio between the forces is the square of the ratio between these speeds.*

Algebraically, these quantities may be arranged so as to exhibit this relation as follows:

$$\frac{\text{attraction due to flow}}{\text{repulsion between free electrons alone}} = \left(\frac{\text{speed of flow}}{\text{speed of light}}\right)^2$$

This law is one of the most fundamental of all that group of laws called classical electromagnetic theory, which seem to have perfect accuracy.

To test its accuracy may look difficult. The speed of flow seems too uncertain. We do not know just how many of the electrons in any piece of wire are free to flow. Each ampere means 6.2420×10^{18} electrons per second, passing any fixed point; so the more free electrons there are, the less speed each electron has to make, for the specified number of amperes. The ratio of their average speed to the speed of light, therefore, is uncertain.

However, so is the ratio of the forces. The more electrons we must consider, the stronger is their electrostatic repulsion; but the attraction due to their flow is fixed by the specified number of amperes. Altogether, as may be proved fairly easily, any change in our estimate of the number of free electrons has exactly the same effect on both sides of this equation.

To test it, actually we need not attempt to isolate the free electrons. An equivalent test is to charge each wire with a known fraction of an ampere-second of excess electrons and measure the force. Naturally there are many other equivalent tests, with conductors of many shapes. Altogether, therefore, this law has been tested many thousands of times, often with great accuracy, and found good. We shall see more of it under *Electrostatics*, below.

Meanwhile let us note another aspect of this law, related to the question "Why do like currents attract?" To answer such a "why" is to bring the question under some broader law or theory. This question about currents comes under the theory called special relativity.

8. Special Relativity.—The term "relativity" has often connoted a degree of abstractness or complexity almost beyond human understanding; and it must be admitted that much of which is called general relativity gives ground for this connotation. Special relativity is relatively simple. It does not ask for curvature of space, nor attempt to relate gravitation to electricity. Moreover, while many parts of general relativity remained in doubt, special relativity was confirmed by so many tests that it became an accepted group of principles.

Special relativity is enough for present purposes, because it covers all steady motions along straight lines—such as the steady flow of electrons in the straight wires. Neglecting the random motions of individual electrons, let us imagine the free electrons as well-drilled soldiers on parade, as all moving at the same speed. The protons and bound electrons, now, can be imagined as spectators, standing quite stationary. Or at least, so it would seem to the "spectators." Now for relativity: It says that the "soldiers," if their motion is perfectly straight and steady, have just as good a right as the "spectators" to call themselves stationary.

Special relativity means exactly what it says. Many of the deductions from it depend on this word "exactly." One of these deductions is about the speed of light in vacuum. The spectators can measure this speed with apparatus standing on the ground, and the soldiers with exactly similar apparatus on a truck rolling along with them. Special relativity demands that the two measurements give the same speed, exactly.

At first sight this seems impossible. At a second stage of examination it seems not quite impossible, yet apparently very unreasonable. It is possible only if clocks, rulers and other measuring instruments behave very queerly when they are moving. At a third stage of examination this queer behaviour becomes reasonable. It is according to a set of equations called the Lorentz transformations, which Lorentz deduced from his prerelativity theory of electrons. Assuming that all matter is composed primarily of electricity, he showed that we could expect clocks and rulers, and everything else, to behave, at least to a very good approximation, according to these equations.

In 1905 Einstein introduced special relativity through his postulate of strict equality in the rights of such groups of observers as the soldiers and the spectators to call themselves stationary. This postulate promoted the Lorentz transformations from the rank of very good approximations to that of laws to be treated as strictly exact.

One of the most surprising points about the Lorentz transformations is that each group of observers says every object carried along by the other group is shorter, along the line of motion, than the carrying group thinks it is. This apparent shortening is called the Lorentz contraction. Because of this rule the free electrons say the protons are a bit crowded; and the protons say the free electrons are not quite so far apart as they think they are.

As to how far apart the electrons think they are, there is no reason to assume this to be the same as if they were at rest. They are free, so their spacings are variable. Actually it makes no difference for present purposes what we assume about them, within reasonable limits. But for a definite mental picture we may note that the wires really cannot be infinitely long, but must be parts of closed circuits, and the process of starting the current neither creates nor destroys any electrons. So the best assumption is that the protons, which are not disturbed in this process, say the number of electrons in each inch of each wire is exactly equal to the number of protons, regardless of the current. Then the protons say the electrons are just as far apart as if there were no current.

In this case, the electrons say they are slightly farther apart, while, at the same time, they say the protons are a bit crowded. For two reasons, therefore, if the wires are carrying "like currents" the electrons say each wire has an excess of protons. Because of this, each free electron in either wire says it is attracted electrostatically by the other wire.

A proton, seeing no charge, cannot agree that this force is electrostatic; yet he cannot deny its existence. So he calls it by another name, "electromagnetic." In short, electromagnetic forces must exist if we are to have relativity.

This conclusion would hold even if we had made some other assumption about the spacings of the electrons. Moreover, there is an infinite variety of cases, other than this one of straight wires, which can be treated as evidence along this line; and they all lead to this conclusion. The case of straight wires is unique only in being the simplest.

This answer to the question "Why do like currents attract?" has been kept purely qualitative here; but it can be made quantitative. By-passing the mathematics, we can see that the speed of flow must enter into the calculation. So must the speed of light, because it was to make the soldiers and spectators agree on the speed

of light that they had to disagree on their spacings. In short, the theory of special relativity actually does demand the equation:

$$\frac{\text{attraction due to flow}}{\text{repulsion between free electrons alone}} = \left(\frac{\text{speed of flow}}{\text{speed of light}}\right)^2$$

9. The Conventional "Direction of the Current."—In connection with the definition of the ampere it was noted that if the electrons in two parallel wires flow oppositely there is repulsion. This relation of force to direction made it necessary to have some way to assign a conventional direction to any current, long before anyone knew in which direction anything really flowed. Even before Ampère had discovered these forces, Benjamin Franklin (1706–90) and other investigators of electrostatic forces had found a need for algebraic signs for charges, to use in equations about their forces.

For a long time it had been assumed that charges of either kind could flow in metals; but Franklin pointed out the fact that this assumption was unnecessary. His one-fluid theory was essentially just like the description given here, of electrons flowing in metals, while protons have fixed positions, except for the fact that he did not know which kind of electricity was the fluid. Finally, when pressed for an answer, he made a tentative guess. Despite his statement that the guess was tentative, his contemporaries adopted it and conventionalized it. Unfortunately it was wrong: when a wire is said to be carrying a current from left to right, the electrons in it are really flowing from right to left.

Strangely enough, this false convention caused no serious inconvenience for many decades, and it still causes none in most of the branches of electrical engineering dealing with motors, generators, power-transmission systems and electric lighting. In such things, electrons simply circulate around closed metallic circuits, remaining within the metal; so if all currents are reversed, all outward manifestations of force, power, heat and light are unchanged.

With vacuum tubes, on the other hand, the situation is quite different. In a tube of almost any type used in radio and other communication systems, there is at least one electrode which can be heated, to enable electrons to escape from it into the vacuum, while other electrodes, kept relatively cool, can only receive electrons from the vacuum and cannot emit them. Through such a tube, electrons can flow from a hot electrode to a cold one, but not the other way. In the wires connecting the tube to other things, therefore, the real directions of flow must be recognized, as shown in fig. 3.

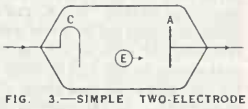


FIG. 3.—SIMPLE TWO-ELECTRODE VACUUM TUBE: (C) CATHODE; (A) ANODE; (E) ELECTRON WITHIN THE VACUUM

As shown there, in the wire connected to the hot electrode, electrons can flow only toward the tube, whereas in the other wire they can flow only away. Consequently, when diagrams based on the false convention show "currents" opposite to these flows, the convention is an endless source of confusion to students. Even to an expert, the need for thinking always the opposite of what is said, and guarding against wrong algebraic signs, is a perennial waste of mental effort.

Much of this waste, though admittedly not all, can be prevented by dropping the term "current" and speaking of "electron flow" whenever there is any consideration of its direction. This procedure indeed comes near to eliminating the trouble entirely. What remains of it is a set of + and − signs, some of them on electrical measuring instruments and the rest in equations referring to electrostatic charges.

These signs are present because of the natural guess that if a "current" is running "into" a body, this body is receiving something, just as a basin receives water when a stream runs into it. So a body is said to be gaining a positive electrostatic charge, when what it is really doing is losing electrons. The conventional + signs, therefore, go with deficiencies of electrons, while the − signs go with excesses.

A. Electric Power

Almost all uses of electricity are also uses of power. This power may be a thousand kilowatts for an electric locomotive; it may be a hundred watts for a lamp; or it may be only the very minute fraction of one watt which a radio receiver picks out of "the air." Whether it is measured in kilowatts or in microwatts, or is simply read in watts from the trade-mark on a lamp, the power and its basic unit the watt are central features in almost any question about a use of electricity.

1. The Watt.—Nevertheless the watt is not really an electrical unit. Power was used in mechanical engineering, long before electricity seemed to have any possibility of use. Any unit of power therefore belongs to the science of mechanics; so the watt is appropriately named for James Watt (1736–1819), the inventor of the first good steam engine. The definition of the watt naturally is based on the technical definition of power. This, in turn, is based on work. These terms "power" and "work" obviously were suggested by their nontechnical meanings, related to muscular effort; but they are given technical restrictions which make them more definite in meaning, as well as more useful.

Gravity does work on any body which moves downward. The amount of work is found by multiplying the force of gravity (the weight of the body) by the height through which it descends. Force times distance, this is, if the descent is vertical; but if the descent is oblique, only its vertical component is used in computing the work. In general, as in this case, the work done by any force, acting on any moving body, is a product of force by distance, in which the distance is the geometric projection of the motion, on a line in the direction of the force.

This restriction to a geometric projection, exemplified by height, is an essential part of some very important principles to be discussed later. For defining the watt, however, the essential features of work are the units of force and distance, which appear in this product. If a 200-lb. man goes from one floor to another ten feet lower, the work done by gravity is called 2,000 ft.-lb. This compound unit, the foot-pound, is a prototype for many units formed and defined mathematically.

Returning to power, it is defined as the rate at which work is being done. The power exerted by gravity on the 200-lb. man, for example, depends on his rate of descent. If his ten-foot descent is accomplished in five seconds the power is 2,000 ft.-lb. divided by five seconds, or 400 ft.-lb. per second. Any unit of power like the foot-pound per second is a compound unit, whose definition includes division as well as multiplication.

The watt is the unit of this type most suitable for use in relation to electricity. Since the ampere is defined in terms of the newton and the metre, these are the best units of force and distance for use in the watt. The time unit to use with them is suggested by the definition of the newton, which involves the second. Therefore the watt is defined as one newton-metre per second.

To emphasize the strictly mechanical character of the watt, we may express that 200-lb. man's weight approximately as 900 newtons, and the ten feet as three metres. The power released by gravity is therefore

$$P = \frac{900 \text{ newtons} \times 3 \text{ m.}}{5 \text{ sec.}} = 540 \frac{\text{newton-metres}}{\text{second}}$$

The name "watt" is simply an abbreviation for this last fraction. That is,

$$1 \text{ w.} = 1 \frac{\text{newton-metre}}{\text{second}}$$

so this power is 540 w.

This algebraic form of the definition of the watt has the advantage of showing at a glance how the power depends on the force, distance or time. The positions of the newton and metre, above the fraction line, go with the fact that if either the force or the distance is increased, other things remaining the same, the power is increased. The position of the second, below the line, goes with the fact that more time, for the same force and distance, means less power. Most electrical units, like this purely mechanical one, are fractions formed out of previously defined units. So in each of them, the positions of the component units in the fraction are useful memoranda.

2. Kilowatts and Kilowatt-Hours.—Among the facts called to mind by this fraction defining the watt is the fact that when we pay an electric light bill we do not pay for a number of kilowatts, even though we may carelessly say we do. This is similar to a sign on the highway saying "Speed 25 miles"; 25 mi. is not a speed. The common misuse of "kilowatt" is like this, but in reverse. The kilowatt, being 1,000 newton-metres per second, is not like the mile but like the mile per hour; it is a unit of the rate of doing work, not of the total work done. What we pay for, each month, is the total work done. If four 100-w. lamps are kept turned on for 200 hours altogether, the work is

$$4 \times 100 \text{ w.} \times 200 \text{ hr.} = 80{,}000 \text{ w.hr.} = 80 \text{ kw.hr.}$$

It is the kilowatt-hour, therefore, rather than the kilowatt, in which this work is billed; and the kilowatt-hour is

$$1{,}000 \text{ w.} \times 1 \text{ hr.} = 1{,}000 \frac{\text{newton-m.}}{\text{second}} \times 3{,}600 \text{ sec.} = 3{,}600{,}000 \text{ newton-m.}$$

By human standards this is a prodigious amount of work to do for a few cents. One newton is about 0.22 lb., and one metre is about 3.3 ft., so the kilowatt-hour is easily found to be about 2,600,000 ft.-lb. This is roughly the work done by 30 men climbing the Washington monument via the staircase—regardless of whether they do it in an hour or in only ten minutes.

3. Power and Electrostatic Forces.—With the watt, the kilowatt and the kilowatt-hour defined in purely mechanical terms, we may now apply them to questions about electricity. The watt, as already noted, is mentioned in the trade-mark on a lamp; and along with it is another unit, the volt. If the trade-mark includes "100 w. 120 v." this part of it means that the lamp will take 100 w. when connected to a pair of wires for which one may say "120 v." The "100 w." refers to power exerted on the electrons flowing through the filament of the lamp. This power is exerted by electrostatic forces, described above, under *Forces Driving Electrons in Wires*. These forces are what must comply with the specification "120 v."

We move a lamp from one fixture to another in the same house, or even into another house, usually without raising the question whether the electrostatic forces in the new location will comply with this specification. Usually, indeed almost always, they do. But if the question is raised, this compliance becomes evidence that there must be some very general law, compelling the charges on the wires to distribute themselves so as to produce the specified forces.

This law, actually, is the basic law of repulsions and attractions between electrons and protons, stated near the beginning of this article. These forces are very much like forces of gravity. Between any two small particles, a force of either type acts directly along the line from one particle to the other. Its strength depends only on the distance between them, and it is not affected by the presence or absence of any third particle. These similarities to gravity enable us to carry many laws of gravity over into electrostatics. Among these are laws about work and power.

If water flows from a millpond to a pool below the dam, the work done on it by gravity is the same, regardless of whether it goes through the mill, turning most of this power to good account, or merely becomes the proverbially useless, though slightly heated, water over the dam. Carrying this law over to electrostatics, we may apply it to electrons flowing through different lamps or motors. Starting on one terminal of the big double-pole switch in the feeder lines of a house (usually in a box on the cellar wall), one electron may find its way through a lamp to the other switch terminal, and another electron may find its way through a motor. These ways may look very different. Nevertheless, the work done by electrostatic forces on one of these electrons is exactly equal to the work done on the other.

This law is especially important because it is *sine qua non* for most of the convenience of operation of household electrical appliances. All such appliances, therefore, are connected so as to bring them under this law. This type of connection, called parallel, is shown in fig. 4.

If now the numbers of electrons passing through the lamp and the motor in a second are equal, the rates of doing work—that is, the amounts of power—are also equal. If (as is more likely) five times as many electrons go through the motor as through the lamp, they receive five times as much power from the electrostatic forces.

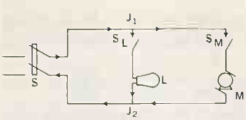


FIG. 4.—PARALLEL CIRCUIT CONTAINING LAMP L AND MOTOR M

(S) double-pole main switch; (S_L and S_M) single-pole switches controlling lamp and motor. With all switches closed, electrons flow as indicated by arrows; any given electron takes one of two paths when it reaches junction J_1 and the streams reunite at junction J_2

In short, when electrons flow from any one point (such as one terminal of that big switch) to another point (such as its other terminal), the power delivered to them by electrostatic forces is proportional to the current, and the amount of power is dependent neither on the paths of the electrons nor on the type of conductor through which they flow.

4. Potential Difference and the Volt.—Any law of proportionality, such as this, can be expressed in a statement that the value of some specially defined fraction stays constant. In this case, the fraction is power divided by current. The name of this fraction is "potential difference."

If this name seems unreasonable without a previous definition of potential, we may recall the analogy between electrostatic forces and gravity, and for gravity recall the importance of the height difference between the millpond and the pool below the dam. If we have no simpler way to measure this height difference, we could calculate it from such a fraction, $\frac{\text{power}}{\text{current}}$, for any stream of water. Potential difference therefore is analogous to height difference.

In such a calculation of the height difference we are not concerned with the height of either the millpond or the lower pool, relative to sea level; but if we have many heights whose differences must be calculated, a fixed base, or "zero height," may be useful. Then the term "height," without "difference," acquires a definite meaning. Analogously, some point in an electrical system may be called "zero potential," and potentials of other points may be based on it.

Returning to the definition of potential difference as $\frac{\text{power}}{\text{current}}$, and the motive for this definition in the constancy of this fraction, it may be well to illustrate this constancy with typical data. For this purpose the lamp marked "100 w. 120 v." may be compared with a slightly larger one marked "150 w. 120 v." and a still larger one marked "200 w. 120 v." Suppose each of these lamps, alone, is connected in series with an ammeter, to measure its current, and the ends of the series are connected to a pair of wires conforming to the specification "120 v.," as yet undefined. Then the ammeters will read about 0.83 amp., 1.25 amp. and 1.67 amp., respectively. So the fraction which stays constant may be calculated as

$$\frac{100 \text{ w.}}{0.83 \text{ amp.}}, \quad \frac{150 \text{ w.}}{1.25 \text{ amp.}} \quad \text{or} \quad \frac{200 \text{ w.}}{1.67 \text{ amp.}}$$

With reasonable allowance for instrumental errors these values for the fraction are equal. Each is practically

$$120 \frac{\text{watts}}{\text{ampere}}$$

so this is what is meant by "120 v."

The word "volt," therefore, like the word "watt," is simply a short name for a compound unit, defined as a fraction; that is, by definition,

$$1 \text{ v.} = 1 \frac{\text{watt}}{\text{ampere}}$$

Its derivation is from the name of Alessandro Volta (1745–1827), who invented the battery, described by him in 1800, because this was the first device for maintaining steady potential differences between terminals of conductors.

5. Transmission of Power.—Electrons are subject to forces of several common types other than electrostatic. Among these are the chemical forces in batteries and electromagnetic forces in generators, transformers and even motors. Indeed a battery, a gener-

ator or a transformer is built and used primarily as a sort of electrical pump, to drive electrons against electrostatic forces, somewhat as a water pump drives water against gravity.

These devices differ from the water pump, however, both in their principles of operation and in their basic purposes. Water may be needed in the place to which it is pumped; but electrons are driven through a battery or a generator, from one of its terminals to the other, only for the purpose of letting them take some external detour—through a lamp or motor or something else—back to their starting point. Of course water also could be used in this way, transmitting power from a pump to a mill. So again, the analogy between electrostatic forces and gravity is fairly close.

Gravity makes such a use of water possible because it repays in the mill the work done against it in the pump. (Friction robs us of some of this work, to be sure, but this loss is beside the point here.)

Gravity can accept and repay work in this way, because it is present in both machines.

In exactly the same way, the electrostatic field—*i.e.*, the region where electrostatic forces can be found—can accept and repay work done on electrons, because it covers the whole apparatus: generator, transmission line and motor, lamp or other receiver of electric power. In short, the transmission of power by electricity is accomplished by driving electrons against electrostatic forces in one place and letting these forces drive electrons in another place.

6. Nonelectrostatic "Voltages."—This characteristic of the electrostatic field, covering the whole system, distinguishes electrostatic forces from all others. Chemical forces, acting only within a battery, cannot obey any such law as that of equal work for all paths between its terminals. Chemical forces can do work only on electrons whose paths take them through the battery; electromagnetic, only on paths through generators, motors or transformers; and so on. When we speak of the potential difference between two points, we cannot be referring to any forces but electrostatic.

Other forces, nevertheless, often exert power in proportion to current, provided the current takes a specified path. So the fraction $\frac{\text{power}}{\text{current}}$ is useful for them also, but for them it must be given another name. The name given is unfortunately a misleading one, electromotive force, for a fraction that is not a force, however much it may be related to forces. Electromotive forces cannot be measured in force units, such as pounds or newtons; but like potential differences they can be measured in volts.

Because of this, the colloquialism "voltage" is often used to cover both potential differences and electromotive forces. In using it, however, care must be taken not to confuse these quantities.

An automobile battery, for example, usually has an electromotive force of about six volts. That is, when current flows through it, in obedience to its chemical forces, they exert power at the rate of 6 w. per ampere; for example, 60 w. for headlights taking ten amperes. A little of this power is lost in driving ions through the liquid of the battery. If this loss is 2 w., only 58 of the 60 w. are available for work against electrostatic forces, so the charges adjust themselves accordingly. That is, while the electromotive force within the battery is 6 v., the potential difference between its terminals is only 5.8 v. Thus these "voltages" differ in strength as well as in nature. For this concrete reason, as well as for the less demonstrable but really more important objective of clarity, these quantities should be distinguished.

The common abbreviation for electromotive force is emf and that for potential difference is p.d.

III. DIRECT-CURRENT CIRCUITS

The whole duty of a generator or a battery, as stated above under *Transmission of Power*, is to drive electrons against electrostatic forces, so that elsewhere in the same field the electrostatic forces can do the driving and deliver the power. Water could be used like electrons; but there is one important difference, although it is more a difference in degree than in kind. If water runs from a pump to a mill through a canal, and the pump stops, the mill need not stop till it has drained the canal. Likewise, electric lights need not go out when a switch is turned off, till they have drained the wires which have excess electrons; but they do go out very quickly. Wires, evidently, differ from canals in being drained very quickly by the appliances they feed. Even larger electrical systems have this characteristic; a lightning flash goes out, "quick as lightning."

To maintain a steady current in a lamp, therefore, we must treat the wires almost as if they had no capacity as reservoirs for electrons at all. The wire which feeds electrons to the lamp must have its supply of electrons replenished continuously by the source of power; and the wire which receives electrons from the lamp must likewise have them taken away. In short, the lamp and the source of power must be connected in a closed circuit, as shown in fig. 5.

Circuits are of many types, some amazingly complicated. In many, notably those of radio systems, where charges do change considerably in fractions of a millionth of a second, the currents involved in draining and rebuilding the charges are important;

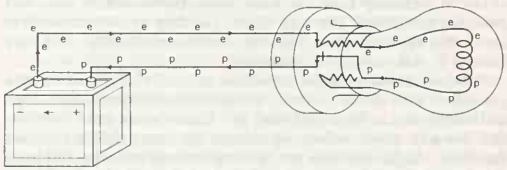


FIG. 5.—FLOW OF ELECTRONS BETWEEN BATTERY AND LAMP. WIRES MARKED e ARE CHARGED WITH EXCESS ELECTRONS; THOSE MARKED p ARE DEFICIENT IN ELECTRONS

but circulatory currents like that of fig. 5 are more universally important, and constitute practically the whole current in any power or lighting circuit.

1. Alternating and Direct Currents.—Most household and industrial circuits contain not batteries but transformers. For the present purposes this is a minor difference, since the transformer is connected to the rest of the circuit like the battery of fig. 5. Moreover fig. 5 is not purely academic, because batteries are used in isolated systems such as those of automobiles and many farmhouses. One essential difference, however, should be noted. A battery pushes its electrons in the same direction all the time, but a transformer cannot do this for more than a small fraction of a second. Consequently transformer circuits must carry alternating currents; that is, currents whose directions alternate many times in every second.

These alternations complicate the principles governing the currents. In fact, the complications were a serious obstacle to the use of alternating currents in the early days of electrical engineering, even though they not only permitted the use of transformers but also simplified the construction of motors and generators. For purposes of explanation, therefore, it seems best to consider first the laws of currents whose directions of circulation do not alternate; that is, unidirectional currents. Among these, it is well to restrict first consideration to currents whose strengths do not change appreciably in any moderate time, such as a few seconds. Currents thus doubly restricted are called direct currents.

Currents in battery circuits, such as that of fig. 5, are typical of direct currents.

2. Direct-Current Charges.—Obviously the steady flow of a direct current indicates steadiness of its driving forces. These include the chemical forces in any battery, of course; also, and just as surely, they include all electrostatic forces. Therefore, when a current can properly be called direct, all charges associated with it must also be practically unchanging. That is, charges associated with direct currents are "static."

In starting the current, to be sure, the charges had to be established. Electrons had to accumulate on some wires and to leave others deficient. But once the current is steady, these changes belong to history.

3. Electron "Bookkeeping."—Under these conditions all electrons entering a wire at one end must be compensated by others leaving it somewhere else. If they cannot escape except at its other end, and no more can enter anywhere, the number of electrons passing any point on the wire in a second must be the number entering the wire. This rule holds, regardless of differences in thickness or material, along the wire. Where the wire is thickest, the electrons flow slowest—still water runs deep—but they can all be accounted for.

Remembering that the term "current" refers to number per second (each ampere being 6.2420×10^{18} electrons per second), this rule takes the form that the current is uniform all along the wire.

Even a lamp filament, thinner than the lamp cord wires and made of tungsten, which is relatively hard going for electrons, makes no break in this uniformity. A lamp does not "burn electricity."

Instead, what the lamp does is to convert energy (*i.e.*, ability to do work) from a form associated with electricity into the random motion of molecules which we call heat. The spillway of a dam does the same with energy associated with water. So the electrons returning from the lamp to the power station have lost potential usefulness, like this water; but they are not themselves lost. We, as "consumers," do not consume electricity. We only borrow it, and we repay it promptly.

Another corollary of the steadiness of D.C. charges, or charges associated with direct currents, relates to junctions of wires, such as those shown in fig. 4 as J_1 and J_2. Once the currents have settled down to steady values, no charges are accumulating at either junction. At J_1, electrons are arriving through one wire and leaving through two. So the sum of the currents in the "leaving" wires must equal the current in the "arriving" wire; and a corresponding, but reversed, rule holds at the other junction.

These laws about the uniformity of current all along each insulated wire and the addition of currents at each junction are among the most useful of all laws about electricity. The former is used by engineers and physicists so often, practically subconsciously, that it is anonymous. The latter, about junctions, is called Kirchhoff's first law, after Gustav Robert Kirchhoff (1824–87). His second law, incidentally, is a mathematical equivalent of the law that the potential difference between any two points has the same value for all paths between them.

4. Some Evidence About Electrons.—These laws about currents have been presented deductively, as if believed because they can be deduced from the principle that a current is a flow of indestructible electrons. Historically, they were known long before electrons were recognized. For experimental tests, ammeters can be inserted in the wires shown in fig. 4. With an ammeter at the position of each arrowhead there, the test is easy. Actually, therefore, we believe these laws not because we believe in electrons but because numerous tests of this sort are summarized in these laws.

Conversely, if we did not believe these laws, we could not believe in electrons. Nevertheless, these laws alone would not compel us to believe in electrons. They might have some other explanation. Our belief in electrons is not based on tests of any one type. Instead, it is based on a great variety of types of tests, all permitting the belief but none compelling it. It might well be said, "We believe in electrons because it would be much too strange a coincidence if all these experiments worked without them."

This curiously backhanded reasoning has characterized the growth of most knowledge of electricity, as well as of molecules, atoms and their structural parts. Theories are proposed, perhaps with enthusiasm but more often with apologetic hedging, to the effect that the theory may serve as a limited analogy to the unknown reality, if it turns out not to be a true description. Later, if any experiment shows results unquestionably inconsistent with the theory, the apology is accepted. If, on the other hand, more and more experiments are found to permit a belief in the theory and none to deny it, the apology is forgotten and the theory becomes a set of principles.

Principles thus established often form a basis for explanation of phenomena much briefer and clearer than any logical structure based directly on their permissive experiments.

A. Ohm's Law

Next in order of importance here, after Kirchhoff's laws, is a law discovered in 1827 by Georg Simon Ohm (1787–1854), which relates the strength of a current in a wire not containing any electromotive force to the potential difference between its ends.

With any given potential difference the strength of such a current depends on the nature of the metal composing the wire. Ohm's law, therefore, is not in the special class of laws independent of materials, discussed above under *Accuracy of the Fundamental Laws*. As noted there, we treat many of these laws as if their accuracy were strictly perfect; and among these laws are Kirchhoff's, because they belong to the group mentioned there as classical electromagnetic theory.

Ohm's law, while not a member of this group, is like these laws in being very simple; it says the current is proportional to the potential difference. This is so much simpler than any corresponding law for water that, when Ohm announced it, it seemed literally too good to be true. So it was not believed. Ohm was considered unreliable because of it, and was treated so badly that he resigned his professorship and lived in obscurity and poverty for several years. Nevertheless he was right; and, fortunately, while still alive he received recognition, such as the Copley medal of the Royal society, awarded 14 years after his great discovery.

The simplicity of Ohm's law, while real, is also risky. Its risk is a temptation to oversimplification. One must not forget the restriction to wires not containing any electromotive forces. Moreover, the nature of the material must not be ignored. It changes with temperature. An incandescent lamp filament, for example, seems in simple tests to disobey Ohm's law very badly; but this is because changes in current can make such extreme changes in temperature, if given any appreciable time. On the other hand, if the changes occur fast enough, and the measurements are made before the temperature changes appreciably (that is, in less than a thousandth of a second), Ohm's law is found to be valid.

Finally, it must be recognized that Ohm's law does not apply to all materials. It does apply well to most true metals, such as copper, aluminum, iron, tungsten, brass, bronze and many others. It even applies to many salt solutions and other electrolytes, when enough care is taken to be sure there are no electromotive forces. But many nonmetallic solids are non-Ohmic; even some substances such as silicon, which appear offhand to be metals, are in this class.

Ohm's law must be used with at least reasonable care. When so used, but only then, it is almost as useful as if it did belong in the group of fundamental laws.

1. Resistance.—Since Ohm's law is a law of proportionality, it may be expressed algebraically as a statement that a specially defined fraction stays constant. This fraction is $\frac{\text{potential difference}}{\text{current}}$. Of course its constancy is restricted to one wire: in each Ohmic wire the current varies in proportion to the potential difference, and the fraction keeps a constant value; but for another wire the fraction has another value. Higher values mean higher potential differences for any given current, of course, or more tendency of the wire to resist the flow of electrons. Consequently this fraction $\frac{\text{potential difference}}{\text{current}}$ is technically called the resistance of the wire.

Algebraically, therefore, we write, for each Ohmic wire, with V for potential difference, I for current and R for resistance,

$$R = \frac{V}{I} = \text{constant}$$

To illustrate this we may consider again the lamp marked "100 w. 120 v." At 120 v. its current is 0.83 amp., so its resistance is

$$R = \frac{120 \text{ v.}}{0.83 \text{ amp.}} = 144 \frac{\text{volts}}{\text{ampere}}$$

Appropriately, this is called 144 ohms; so by definition,

$$1 \text{ ohm} = 1 \frac{\text{volt}}{\text{ampere}}$$

Lest there be any confusion, it may be well to recall the definition of the potential difference V, as applied to this lamp, namely

$$V = \frac{\text{power}}{\text{current}} = \frac{100 \text{ w.}}{0.83 \text{ amp.}} = 120 \text{ v.}$$

Each of these fractions, $\frac{\text{power}}{\text{current}}$ and $\frac{\text{potential difference}}{\text{current}}$, has its own condition for constancy. $\frac{\text{Power}}{\text{current}}$ keeps a constant value for a given pair of power-line terminals (with different currents carried by different users of power); $\frac{\text{potential difference}}{\text{current}}$ keeps a constant value for a given wire (with different currents caused by different potential differences).

2. Heat Produced by a Current.—The production of heat is perhaps the most familiar of all the possible effects of electric currents, because of its occurrence in the filaments of incandescent lamps and because of the fire hazards of overloaded electric wires. If a householder plugs a big heater into a lamp fixture, burning out the fuses, and then replaces them with fuses of higher rating, disaster may result. Incidentally the fuses also depend on this heating effect: each of them contains a short bit of wire made of some readily fusible metal (hence the name) which will melt if overloaded; but this bit of wire is enclosed in a fireproof container.

None of these wires is subject to any chemical or electromagnetic forces to help or hinder the electrostatic forces in driving its electrons. In these wires, therefore, the electrostatic forces are balanced only by the forces analogous to friction, which oppose the passage of electrons among the atoms composing the wires. So the power delivered electrostatically to these electrons is all converted into heat.

Algebraically, this power is $P = VI$. Moreover these wires fulfil the conditions for applying Ohm's law, so $V = RI$. Combining these equations, we find $P = RI^2$.

That is, the power P, or the heat produced by it, is proportional to the square of the current. Doubling the current, for example, quadruples the heat.

3. Resistivity.—Since a lamp filament and the wires of the lamp cord are all in series, they all carry the same current; but obviously the heat produced in each inch of the filament is far greater than in each inch of either wire of the cord. The same current means the same I in $P = RI^2$, so the greater heat shows proportionally greater resistance. This difference is attributable chiefly to the comparative thinness of the filament, but partly also to a difference between tungsten and copper.

Looking first at the thinness, wires of the same metal and same length have resistances inversely proportional to the areas of their cross sections. From this one may infer readily that the electrons flow just as fast, well within a wire, as near its surface. This inference is contrary to a popular belief, which seems to be based on two facts. One of these is that static charges usually accumulate entirely on surfaces; the other, that radio-frequency currents run stronger near the surfaces than well within the wires. Nevertheless, direct currents really do flow uniformly all through their wires; the low-frequency (60-cycle) alternating currents used in lamps flow practically as uniformly; and radio-frequency currents are so strongly affected by nonstatic forces that they pay little attention to the low resistances of ordinary wires. Regardless of popular belief, therefore, inverse proportionality between resistance and cross-sectional area is the law.

For different lengths, with the same cross section and the same metal, the equation $P = RI^2$ shows readily that the resistances are directly proportional to the lengths.

Combining these proportionality laws, and finding a fraction that stays constant and so deserves a name, the fraction is

$$\frac{\text{resistance} \times \text{cross-sectional area}}{\text{length}}$$

Its name is resistivity, and its constancy extends over all wires of any one metal at any one temperature.

Expressing all dimensions in metres and resistances in ohms, the units in this fraction cancel down to the product ohm × metre. Copper, for example, has about 1.7×10^{-8} ohm-metre at room temperature. So if one wants to know the resistance of a copper wire ten metres long and one millimetre in diameter, he needs only to multiply the resistivity of copper by ten metres and divide by the area:

$$\frac{1.7 \times 10^{-8} \text{ ohm-metres} \times 10 \text{ m.}}{\pi (\frac{1}{2} \, 10^{-3})^2 \text{ m.}^2} = (1.7 \times 10^{-8}) \times 10 \times \left(\frac{4}{\pi} 10^6\right) \text{ohm} = 0.2 \text{ ohm}$$

Low resistivities are especially desirable in motors and other devices in which many wires must be packed into small spaces and the production of heat must be kept down. In this respect only one metal is better than copper. This is silver (gold not being quite so good). Its resistivity is less by only about 6%, and this is not enough, usually, to be worth the cost of silver.

Aluminum, with a room-temperature resistivity of about 2.8×10^{-8} ohm-metre, is also among the best of the common metals. For equal weights, regardless of size, it is even better than copper, because it is only about one-third as heavy.

Tungsten has a resistivity about 2.5 times that of copper; zinc, 3.3; iron, 6; lead, 12. Brass, although made of copper and zinc, has a resistivity about four times that of copper; and, in general, alloys have resistivities higher than any reasonably weighted averages of those of their constituents.

With rising temperature the resistivities of most elementary metals increase slightly faster than in proportion to the absolute temperature (centigrade plus 273°, or Fahrenheit plus 460°, approximately). Resistivities of alloys also increase, but at notably lower rates and not even approximately under any such simple rule.

4. Other Terms and Units.—Some other terms often used in relation to Ohm's law should be defined here. These are as follows:

$$\text{conductance} = \frac{1}{\text{resistance}} = \frac{I}{V}$$

$$1 \text{ mho} = \frac{1}{\text{ohm}} = 1 \frac{\text{ampere}}{\text{volt}}$$

and

$$\text{conductivity} = \frac{1}{\text{resistivity}}$$

For example, the conductance of that ten-metre copper wire, whose resistance was 0.2 ohm, is five mhos; and the conductivity of copper is about $6 \times 10^7 \frac{\text{mho}}{\text{metre}}$.

The ending *-ance* (as in conductance and resistance) is used for names of constants whose constancy is restricted to individual things, such as this piece of wire. Backward spelling of the names of units for these constant fractions gives the names of units for their reciprocals. The ending *-ivity* is used for properties of special substances, such as copper, which are common to all things made of these substances.

Incidentally, all the other electrical units of the M.K.S. system are compound units, like the volt and ohm, defined in terms of the metre, kilogram, second and ampere, either directly or in an equivalent form through the newton. Each of these units is named for an outstanding research worker, prominent in the development of that part of the science in which the unit is most used.

One other such unit already has been mentioned under its compound name. This is the ampere-second, whose equivalent, 6.2420×10^{18} electrons, was discussed under *Amperes and Volts*, above. It is called the coulomb, after Charles Augustin Coulomb (1736–1806), who devised the first reliable method for measuring quantities of electricity.

The definitions of the volt, ohm and coulomb given here are typical and illustrative; to avoid confusion, the definitions of other units will be postponed until a discussion of them is needed.

B. Batteries

The simplest source of power for direct currents is a battery.

Batteries are chemical apparatus with currents and thus are related to the chemical effects already discussed briefly under *Ions: Currents in Molten Salts and Water Solutions*, above. Batteries, in fact, always contain water solutions—even when they are sealed up, watertight, and are called "dry batteries."

In the discussion of currents in molten salts it was stated that molten sodium chloride could be electrolyzed; that is, decomposed by the passage of current. The products of this electrolysis, metallic sodium and gaseous chlorine, are both highly reactive. If brought together they recombine into salt, violently, with obvious conversion of chemical energy into heat. To separate them, therefore, requires work against chemical forces. For a view of these forces, and of the forces which do the work against them, we may extend the theory which was introduced with fig. 1.

As noted there, at electrode E sodium ions are reduced to atoms of metallic sodium by the addition of electrons, which are transferred to them from the metal of electrode E. At the other electrode, P, chlorine ions are robbed of electrons, which are transferred from them to the metal of this electrode. In short, at each electrode, electrons are transferred from one substance to another, and these transfers change ions into atoms of chemically different substances. So the work done against chemical forces in such a case as this is primarily work done in transferring electrons, and the chemical forces with which we are concerned are forces acting on these electrons.

These forces are exerted partly by other electrons and partly by the nuclei of the atoms; thus, entirely by other electric particles. They can justly be called electric forces, therefore; but they differ greatly from the electrostatic forces described near the beginning of this article. None of these electrons is static. On the contrary, the effects of their motions are important. Moreover the subatomic distances between them invoke laws (classified under quantum mechanics) very different from the laws of long-range forces. The effects of these laws depend so much on the types of organization of the electrons in these atoms, and they vary so much from one chemical element to another, that it is convenient to call these forces chemical forces.

Now let us consider the forces which do the work against the chemical forces in such a case as the electrolysis of sodium chloride. These forces are electrostatic. At the surface of each electrode, wherever it touches the electrolyte, there are charges, forming what is called a double layer: excess electrons in surface atoms of one substance, and electron-deficient surface atoms in the other substance.

If this double layer were not present, the chemical forces would build it up instantly, by enabling atoms of one kind to take electrons out of atoms of the other kind. This building up is limited only by the electrostatic forces which the double layer exerts on the electrons within it, tending to return them to the atoms which have been robbed. These electrostatic forces usually strike a balance with the chemical forces, such as to make electrons just as likely to go one way as the other.

When the salt is being electrolyzed, obviously more electrons are going against the chemical forces than with them. But it takes only a slight increase in the electrostatic forces to disturb the balance enough for this, and this increase is obtained readily with electrons forced through the cell by an external source of electric power. These electrostatic forces, like any other forces rightly called electrostatic, can be described in terms of potential differences. Therefore there is a potential difference between any point within either electrode and a point within the electrolyte.

Since the chemical reactions occurring at the electrodes are different, so are the chemical forces, and so also are the electrostatic forces within the double layers. Consequently the potential differences between the electrolyte and the two electrodes are not the same, and a point within one electrode is not at the same potential as a point within the other. For most purposes we are not concerned with the potential difference between the electrolyte and either electrode. Instead, usually, we need the potential difference which specifies the work done electrostatically on an electron driven all the way through the cell. This is the potential difference between the two electrodes.

1. The Ordinary Storage Battery.—Charging a battery, as already noted, is not putting an electrostatic charge on anything; or more exactly, not on anything larger than an ion. Instead, it is electrolysis. In principle it is fundamentally like the electrolysis of sodium chloride, although with different materials for practical reasons. Using the battery, or discharging it, is simply allowing its chemical forces to drive the electrons the other way.

Like a poor rule, the sodium chloride cell will not work both ways; so it is a poor battery. Its chlorine atoms, when robbed of their excess electrons, will not stay put. They react. If the metal of electrode P refuses to react with them, they react with each other, forming pairs, molecules of gaseous chlorine. Bubbles of this gas grow on the surface of the electrode, break away, rise through the molten salt and are gone. To make a good storage battery, we must have products of electrolysis which stay put, both in position and in readiness for a reversal of the eletrolytic reaction.

Electrolytic reactions with such products are rare. Moreover, both the charging reactions and the discharging must yield only stable products. And if any product belongs specifically at one electrode, rather than the other, it must stay there. This last requirement puts a premium on insoluble products, mechanically strong enough to stick to the electrodes; yet their formation must not blanket the electrodes with insulating materials and stop the flow of electrons. Altogether these requirements are very stringent. Materials meeting them are not quite worth their weight in gold; but even in an automobile they are worth the weight of another proverbially heavy element, lead.

Lead is the essential element in all four products of reaction, one at each electrode, charging and discharging, in each of the six electrolytic cells which compose the typical 12-v. automobile battery of the 1960s.

In 1859 Gaston Planté (1834–89) discovered the ability of lead to make a good storage battery. Many further researches greatly improved it in such practical aspects as mechanical strength (which lead requires but does not readily give) and methods of manufacture. Other researches were inspired by the hope of discovering reactions as good as Planté's in lighter materials.

The great talents and resources of Thomas A. Edison (1847–1931) were attracted by this prospect and led to the Edison cell, to be described later. But U.S. automobiles of the 1960s still carried lead.

Lead, in metallic form, is one product of the charging reaction, so it corresponds roughly to the sodium in the electrolysis of sodium chloride. Corresponding roughly to the chlorine is lead peroxide, PbO_2. The discharging reactions produce, at both electrodes, lead sulphate, $PbSO_4$. Fortunately, this differs from most sulphates in being insoluble.

The liquid here is dilute sulphuric acid; *i.e.*, distilled water and H_2SO_4.

Now for the electrolytic reactions. Returning for a moment to the simpler case of sodium chloride, the behaviour of individual particles at electrode E can be summarized thus:

$$\text{electron} + \text{Na ion} = \text{Na atom}$$

This electron, of course, is one of those electrostatically forced out of electrode E; the sodium atom is one more electrically neutral atom adding itself to the accumulating mass of metal.

Corresponding to this in the battery is a reaction involving two electrons from the electrode, thus:

$$2\ \text{electrons} + PbSO_4 = \text{Pb atom} + SO_4\ \text{ion}$$

The sulphuric ion dissolves off into the dilute acid (which is strongly ionized already) while the lead atom simply stays about where it was. The lead sulphate on this electrode, therefore, seems to change gradually into metallic lead. (Naturally this metal is in a very porous form; so it is called "spongy" lead, to distinguish it from the stronger, more compact lead alloy composing the frame which supports it.)

At electrode P in the sodium chloride cell, the reaction is:

$$\text{Cl ion} = \text{Cl atom} + \text{electron}$$

This electron enters electrode P and flows away through it; the neutral chlorine atom reacts as described above.

On the corresponding electrode in the battery,

$PbSO_4 + 2\ H_2O = PbO_2 + H_2SO_4 + 2$ H ions $+ 2$ electrons

These two electrons enter the electrode and flow away through it; the molecule of sulphuric acid and the two hydrogen ions dissolve off into the dilute acid; and the molecule of lead peroxide, being insoluble, stays about where the molecule of lead sulphate was. So the mass of lead sulphate on this electrode seems to change gradually into lead peroxide. More exactly, we cannot of course expect each molecule of lead sulphate on this electrode to be attacked by a pair of water molecules, acting simultaneously in such good team work as to carry out this reaction at a single blow. Trimolecular reactions generally go by stages, and this one is no exception. Details of these stages would be out of place here, however, but they will be found in the article BATTERY.

Returning to these over-all reactions, and checking up on the statement that, despite the name "charging," they do not electrostatically charge anything bigger than an ion, we find four checks, as follows.

1. The reaction starting with two electrons and $PbSO_4$ yields a doubly charged ion; so it neither creates nor destroys any charge.

2. The reaction starting with $PbSO_4$ and $2H_2O$, both uncharged, yields two H ions (lone protons, till they attach themselves to water molecules) and two electrons; and these balance electrostatically, so this reaction neither creates nor destroys any charge.

3. The sulphate ion from reaction 1, with its two-electron charge, is added to the dilute acid, but so are the two protons from reaction 2; so the acid, as a whole, neither gains nor loses any charge.

4. Reaction 1 takes two electrons away from the metallic part of the system (*i.e.*, everything but the acid) but reaction 2 gives two electrons to it; so it, as a whole, neither gains nor loses any charge.

But there is a weak point in this last check. Suppose the circuit is broken outside the battery, so that the electrons given to one electrode cannot make their way to the other. Then each electrode must be gaining or losing a charge. This cannot go on for long. If the circuit is broken, check 4 is soon changed from an endorsement to a veto.

Therefore, and in further confirmation of the theory: When the circuit is opened, so that the flow of electrons is stopped, the reactions also are stopped. This law holds, both for the charging reactions described by reading these equations as written, and for the discharging, or using, reactions described by reading these equations backward.

2. Why Batteries Are Series.—Properly speaking, the term "battery" should not be used to denote a single electrolytic cell. Rather, like a battery of artillery, an electric battery is a group of similar units working together. Three cells constitute the six-volt automobile battery. Many automobiles, trucks and small airplanes have 6 cells; farm lighting systems often have 16; and so on.

Back of any such number is the matter of electromotive force. To drive the starter of a car, for example, its battery must deliver power. Moreover, the power required by the starter must be delivered at the starter, not wasted in heating the wires which lead to it from the battery. Of course some power will be wasted in these wires, but this waste must not be too great. As noted previously, if the current in such a wire is doubled, this waste is quadrupled. Conversely, if by any change of design the current can be halved, the waste will be quartered.

A change from one cell to two in series will do this, and three in series will do even better. If two cells are in series, an electron passing through them is worked upon by chemical forces in each cell, so twice as much work is done on it as in only one cell. Consequently, for any required amount of work only half as many electrons is needed; thus for any required power, the current may be halved.

Extending this rule, the more cells there are in series, the less current is needed. Why stop at three, then? For a truck we do not; more power is required for starting its bigger engine, so six are used. Additional accessories—air conditioners, power-operated windows, etc.—have led to the six-cell automobile battery. The farm lighting system, with its 16 cells, does not require still more power; but its wires are longer (often from the barn to the house) so there is more need for reducing the requirement of current.

On the other hand, the more cells there are in a battery, the more it costs for separate containers of acid, and there are other complications. So the number chosen for any type of use is a compromise.

3. Why Long Power Lines Have High "Voltages."—Similar compromises occur in other systems for the transmission of power. The long lines held up by tall steel towers often have potential differences of 110,000 v. between their wires, or even more. The longer a power line is, the more need there is for using a small current and therefore a high potential difference.

Thus the line from Hoover dam to Los Angeles, 270 mi. long, runs at 275,000 v.; 110,000 is a common voltage for lines of such lengths as 100 mi.; and 13,000 is common for distribution over 10 or 20 mi. Even this 13,000, however, is far too high for household use, or for motors of a few horsepower in a factory with nonexpert handling. For such uses, short lines run at many voltages in the low thousands or in the hundreds, and power is fed to houses at 110 to 120 v.

Such a variety of potential differences is made practicable only by the use of transformers, therefore also only with alternating currents, on which more will be said later. But the basic principles are those noted above: if the current in a wire is halved, its power loss is quartered; but to halve the current and still transmit as much power, the potential difference must be doubled.

4. Battery-and-Generator Circuits.—Taking the automobile system to illustrate still further the general principles of circuits, let us consider its generator in relation to its battery. A generator alone would be insufficient: it is electrically dead when the engine is not running, so it could not drive the starter, nor even keep the parking lights glowing. A battery alone would do better; but it would have to be charged occasionally with current from some source outside the car. The generator charges the battery and does much or all of the other electrical work when the engine is running, not too slowly; and when the engine stops, the battery takes over. The essentials of the circuit which make this possible are shown in fig. 6. This figure does not show all the details of the system but only such essentials as are pertinent to the purpose of illustrating general principles.

The engine is started by closing the starter switch SS (with generator switch GS open), thereby letting the battery drive a very strong current around the circuit through the starter. This current does not show on ammeter A because the connections are designed to protect this ammeter from having to carry any such heavy current. (If the big switch SS is controlled by a little push button, ammeter A may show the push button current, which is relatively weak.) As soon as the engine gets running well, however, and the generator becomes strong enough to drive current backward through the battery, an automatic device closes switch GS and the pointer of the ammeter A swings toward the end of its scale marked "charge."

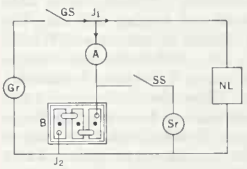


FIG. 6.—BATTERY CONNECTION IN AN AUTOMOBILE

(B) battery; (Sr) starter; (SS) starter switch; (A) dashboard ammeter; (Gr) generator; (GS) generator switch; (NL) normal load including the ignition system, lights and electrical devices other than the starter. The stream of electrons divides at junction J_1 and reunites at junction J_2

Under these conditions the generator is driving electrons not only through the battery but also through the ignition system, which is a part of the "normal load" shown in the figure as NL. So the stream of electrons divides at junction J_1, as shown by the arrowheads, and reunites at J_2. At each junction, of course, it obeys Kirchhoff's first law, stated above under *Electron "Bookkeeping"*: The sum of the currents in the wires by which electrons leave J_1 must equal the current in the wire by which they arrive at J_1; and a similar, but reversed, rule holds at J_2.

If the engine is now slowed, weakening the generator, the pointer swings back toward zero. If it is slowed enough, the pointer may

swing beyond zero, toward the end of the scale marked "discharge." If it is only a little way into discharge, electrons may be flowing to J_1 from both the battery and the generator; then the sum of these streams is the flow to the ignition system. Kirchhoff's first law still holds; only the quantities to be added are changed.

On further slowing of the engine the discharge current is increased and perhaps it may divide at J_1, some electrons from the battery going to the ignition system and some going to the generator, backward from its viewpoint. This is obviously a waste of the battery's energy, so the automatic switch GS soon opens to stop it. Under any of these conditions, if the lights are turned on, increasing the current through NL (which includes the lights as well as the ignition system), the ammeter will show less charge or more discharge. The current through the generator also will change; but Kirchhoff's first law is always obeyed.

5. Some Other Battery Cells.—Almost any electrolytic cell with electrodes of different materials can act as a battery cell, even though, as noted above, most cells cannot be recharged by merely running current backward through them. The electrolyte need not always be acid. In the Edison storage cell it is a solution of caustic potash, the most alkaline of common chemicals. It may even be neutral: sea water is well known to boatmen for its ability to rust a steel rudder exceptionally fast on a boat with a bronze propeller. Steel and bronze in sea water do not make a really good battery cell; but they work well enough—or from the boatman's viewpoint, badly enough—to drive electrons from the steel rudder through the boat to the bronze propeller. This loss of electrons by iron atoms enables them either to dissolve off as ions or to capture hydroxyl ions (OH with an excess electron) from the water, much as Planté's spongy lead captures sulphate ions. But the ions which dissolve wander away from the electrode, so they destroy any hope one might have for making a storage battery out of iron, bronze and sea water. Moreover, those which capture hydroxyl ions undergo subsequent changes (to be described later) which further ruin any such hope.

Meanwhile on the other electrode, the bronze propeller, electrons coming from the boat are given to hydrogen ions in the water, making them into neutral atoms, which combine into molecules of hydrogen gas. The escape of this gas completes the ruin of such a cell as a storage cell.

Nevertheless a cell much like it is very useful. This is the "dry cell," invented by Georges Leclanché (1839–82) about 1865 and found in many portable radio receivers, hearing aids, etc. For sea water the dry cell substitutes a slush of sal ammoniac (NH_4Cl) and its solution in water; for iron, zinc; and for bronze, carbon. The slush, being saltier than sea water, is a better conductor; and zinc is more reactive than iron. The greatest advantage of this cell, however, is that in it the formation of hydrogen bubbles is prevented by a layer of manganese dioxide around the carbon electrode. This readily gives up some of its oxygen atoms, which combine with the newly formed hydrogen atoms (technically called nascent hydrogen) to form water.

The cell is called "dry" because it is sealed over, watertight, so that its outside can be dry. This cell, like the boat, is no storage cell. The solution of its zinc atoms as ions, alone, would prevent recharging it. So would its method of removing the hydrogen, which is as irreversible as the escape of bubbles. Nevertheless, dry cells are very useful in any portable apparatus in which the absence of any "This side up!" is more important than capability for recharging.

Finally, returning to storage cells, the Edison cell, like the boat, drives electrons from an iron electrode. But here the iron cannot dissolve off, because the only ions with excess electrons in its electrolyte, caustic potash (KOH), are hydroxyl, OH, and with these the iron atoms form an insoluble compound, ferrous hydroxide, $Fe(OH)_2$. This is not rust. Its colour is green, which is characteristic of ferrous iron; that is, iron with each atom deprived of two electrons. In rust each atom has lost three electrons (either to OH or to oxygen without the H) and iron in this state is called ferric iron.

Ferrous iron, as an impurity in magnesium, accounts for the colour of many greenish rocks; and prolonged weathering often lets the free oxygen of the air rust such rocks to reddish brown. In the Edison cell, such weathering of its more porous and vulnerable ferrous hydroxide is prevented by an airtight cover, which keeps fresh air from dissolving into its electrolyte. So its ferrous hydroxide stays ready for reduction to metallic iron by a backward, or charging, current.

On the other electrode hydroxyl ions are given to the electrolyte while the battery is driving current forward, and are deposited from the electrolyte while charging. This is made possible by changes in hydroxides of nickel, which are held on this electrode, between the nickelic $Ni(OH)_3$ and the nickelous $Ni(OH)_2$. At both electrodes, therefore, the only ions in the electrolyte which react are the hydroxyl ions. Its other type of ion, potassium, is the commonest ion which is certain to stay in solution and never deposit any unwanted material on either electrode.

Cells of several other types are used for special purposes; many others are obsolete; and innumerable types could be made and described. The above examples, however, cover all their main principles.

6. Electroplating.—Another use of electrolytic cells is in electroplating. In details electroplating is often complicated, but its general principle is simple.

For example, to plate a brass object with copper, one has only to make it one of the electrodes in an electrolytic cell containing cupric sulphate, $CuSO_4$, and send current through the cell in the right direction to drive cupric ions toward the object to be plated. To avoid contaminating the solution with any other sulphate the other electrode should be made of copper.

At the very first, of course, this cell is a battery cell; but almost instantly, as the surface of the brass becomes coated with copper, it becomes like the copper electrode, so the electromotive force of the cell drops to zero. It remains zero, then, as continued current thickens the layer of copper plate.

Thus far all is simple; but complications arise when one considers effects of dirt or other inhomogeneities of the surface to be plated; also of irregularities of shape which tend to attract more ions to its humps than to its hollows. These complications and others require special techniques for good electroplating. Silver, for example, can be deposited readily from its nitrate, $AgNO_3$, but when so deposited it forms a layer of crystals, beautifully glittering but so loosely attached as to be useless as silver plate. The double cyanide of potassium and silver $KAg(CN)_2$, however, deposits a smooth, usable silver plate. Details of this sort are endless; for more of them and for their explanations, *see* ELECTROPLATING.

7. Faraday's Law of Electrolysis.—In any electrolytic cell, either a battery cell or one used for electroplating, the law governing the amount of any metal dissolved off an electrode or deposited on it is exactly what one might expect from the nature of ions, as described above. For example, the amounts of copper deposited from $CuSO_4$ on different objects being plated are proportional to the numbers of electrons conducted through these objects to their surfaces. So the amounts deposited by any given current, running for different lengths of time, are proportional to these times. Likewise the amounts deposited by different currents in a given time are proportional to the currents. And since each cupric ion in $CuSO_4$ is an atom deprived of two electrons, one atom of copper is deposited for every two electrons coming through the object to its surface.

For silver, either in $AgNO_3$ or $KAg(CN)_2$, the relation is not this 1 for 2, but 1 for 1, because each silver ion is an atom deprived of only one electron. For aluminum, which is produced commercially by electrolysis of a molten salt, it is 1 to 3. In all cases the ratio is a simple one of this sort.

This law was discovered in 1833 by Michael Faraday (1791–1867) and is often called Faraday's law, though he discovered so many important laws that it is better to call this one Faraday's Law of Electrolysis.

Strictly speaking, Faraday himself did not formulate his law in terms of electrons, nor did he apparently see what seems so clear in retrospect—that it is excellent evidence for the existence of some such particles. H. L. F. von Helmholtz (1821–94) seems

to have been the first to see this clearly, in 1881, 48 years after Faraday's discovery.

Meanwhile the law had been used so often that it had acquired a conventional form. This is in terms of a quantity of electricity appropriately called the faraday, for which modern measurements give the value 96,521.9 coulombs, or ampere-seconds. The faraday may be defined as the quantity of electricity which will deposit a coat of silver weighing exactly as many grams as there are units in its atomic weight. For copper from $CuSO_4$, in which each copper atom has lost two electrons, it takes two faradays to deposit a coat weighing as many grams as there are units in the atomic weight of copper; for aluminum, likewise, three faradays. For ferrous iron, it takes two; for ferric iron, three; and so on. Thus if the quantity of electricity passing through any cell is expressed in faradays, the weight of any element this quantity of electricity deposits can be computed easily.

Incidentally, the faraday may be called 26.81 amp.-hr. So, if the storage battery in a car is rated at 100 amp.-hr., this is nearly four faradays. Since each atom in its spongy lead must give up two electrons in its conversion to $PbSO_4$, and since the atomic weight of lead is nearly 208, 208 grams of this lead in each cell would be converted by two faradays. So the 100-amp.-hr. rating implies nearly twice 208, or about 400, grams of spongy lead in each cell. This is nearly a pound.

Since one electron is 1.60206×10^{-19} coulomb, it is easy to show that one faraday is $\frac{9.65219 \times 10^4}{1.60206 \times 10^{-19}} = 6.0249 \times 10^{23}$ electrons. So 6.0249×10^{23}, called Avogadro's number after Amedeo Avogadro, conte di Quaregna (1776–1856), is the number of atoms of any element in a mass whose weight in grams equals the atomic weight of this element.

8. The Algebra of Battery Circuits.—As noted under *Why Batteries Are Series*, above, the work done on any electron is proportional to the number of cells in the series. From the definition of electromotive force (see *Nonelectrostatic "Voltages,"* above) it is evident that the electromotive force E of a battery of n cells in series, each with emf E_1, is nE_1.

Typical values for E_1 are those of the Edison cell, about 1.2 v., and Leclanché's dry cell, 1.5 v. Planté's lead storage cell has an exceptionally high E_1, almost two volts. To light ordinary 120-v. household lamps with a battery (as in a stand-by emergency system) about 63 lead cells would suffice; but 100 Edison cells would be needed.

Farm lighting systems with 16 lead cells need special low-voltage lamps (which would burn out instantly if given 120 v.), and automobile headlights have very short filaments, obviously different from those of household lamps.

For a battery to be well suited to its purpose, most of the work done by its chemical forces should be made available for external use, and only a small fraction of this work should be wasted by resistance within the battery. That is, calling its current I and the potential difference between its terminals V, most of the chemically developed power EI should become the power VI delivered by the electrostatic field (as explained under *Transmission of Power*, above) to whatever is connected externally between its terminals. So V should almost equal E.

The power wasted by internal resistance, as heat within the battery, is like the RI^2 in an Ohmic conductor with resistance R, explained under *Heat Produced by a Current*, above. So, since we shall need the letter R later for an external conductor, we may call this internal waste rI^2 and write

$$VI = EI - rI^2$$

and so

$$V = E - rI$$

The internal resistance r is nr_1, where r_1 is the internal resistance of any one of the n cells in the series. And, as may be inferred from what has been said about V and E, the voltage loss rI is usually small in comparison with E or V.

An exception to this last rule occurs in starting the engine of a car. The starter current is much stronger than any current the battery could maintain for long without overheating. So it drops V quite noticeably. This is why the headlights go dim if the starter is used while they are on.

If now the external circuit consists only of a single Ohmic resistor, as in fig. 5, and its resistance is called R, we may write Ohm's law for it as

$$V = RI$$

This I and the I which occurs in $E - rI$ are the same because there is no division of current at either junction of the external resistor to the battery. Moreover, the V's are the same because each is the potential difference between these two junctions. So

$$RI = E - rI$$

and from this we find

$$E = (R + r)\, I$$

This equation enables us to calculate I whenever we know E, R and r, without worrying about the variable and unknown V. Then V can be found if wanted, either as $E - rI$ or as RI.

With any given R, such as the resistance of a lamp or that of a set of tube filaments in a portable radio receiver, to keep rI small in comparison with E we must evidently make r small in comparison with R. To make r small, we must make r_1, in each cell, still smaller. Since most of r_1 is the resistance of the electrolyte, it may be diminished by any shortening of the distance an ion must travel through the electrolyte (like shortening of an Ohmic wire) or by any increase of cross-section area. A good battery, therefore, has its plates very close together, with large surface area.

How close, and how large, depend on how small r_1 must be made; and this depends on I. For the tube filaments in the radio receiver, I is very much smaller than for the headlights of a car. So a battery with plates plenty large enough in area for the portable radio would be very inadequate for the car—as one might expect without the aid of algebra. Conversely, this illustrates and qualitatively checks the algebra.

In deriving these equations from which to find I and V in terms of the more constant and therefore more readily known E, r and R, we have used both Kirchhoff's laws without naming them. His first law, about currents at junctions, said here only that the currents in two lines, the battery and lamp, joined at either battery terminal, were equal; but in other cases we have seen (under *Battery-and-Generator Circuits*) it has related three currents, at a junction of three wires.

Kirchhoff's second law was applied for the simple battery and lamp to the simple two-wire junctions at the battery terminals, equating the two expressions, RI and $E - rI$, for the potential difference V between them. It also could be applied to the two three-wire junctions in the battery-and-generator system. When so applied it would equate three algebraic expressions, all for the same potential difference. Thus Kirchhoff's laws would give us more equations for the more complicated system than for the simple system; but we should need more for more unknown quantities.

This idea of more equations for more unknowns can be extended indefinitely, still relying on Kirchhoff's laws. Often an expression for the potential difference between two junctions is far more complicated than the simple RI or $E - rI$ used above; but this is a detail. Kirchhoff's laws enable us to calculate the currents and potential differences whenever all such expressions can be set up, even in very complicated mazes of wiring.

IV. ELECTROSTATICS

The importance of potential differences in all direct-current calculations is due to the law that the same potential difference applies to all conductors leading from any one fixed point to any other. This law, first introduced under *Power and Electrostatic Forces*, above, was described there as a logical consequence of the fundamental law of repulsion or attraction between electric charges.

This law about potential differences is not by any means the only important corollary to the force law. Electric charges are so nearly omnipresent in electrical appliances that many laws

about them are important. Moreover, the force law itself is mathematically simple enough to invite research, and many of the results of such research possessed an unusual degree of the quality aptly described by mathematicians as elegance. This led to vast extensions of such research, producing a highly developed mathematical theory of electrostatics.

A. Charges Without Currents

The law about potential differences holds whenever all charges are static, in the sense that the number of excess electrons or protons in any fixed region does not change with time, even though electrons may be flowing steadily through the charged conductors. Many other laws of electrostatics hold likewise in the presence of steady currents. Indeed they hold to a very good approximation with alternating currents of low frequencies, such as those of power and lighting circuits. Some other laws of electrostatics, however, including some with important bearings on the nature of electricity, hold only when all electricity is staying at rest.

In this respect, as in many others, electricity is like water. The surface of water is level when the water is at rest, though usually not level when it is flowing, even if the flow is steady. Likewise when the electrons in a piece of metal all stay at rest, we can safely infer that there is no electrostatic force anywhere within the metal to move them. Here is an obvious simplification of the forces, as compared with those in the presence of current. In terms of potential differences, when all electricity stays at rest there is no potential difference between any two points in the same piece of metal. Such a piece of metal is called equipotential.

Outside the metal, there may still be electrostatic forces. Certainly there are if there is another piece of metal near by, at a different potential, as for example when the two pieces are the terminals of a battery on open circuit. These external electrostatic forces have much to tell about electricity.

1. Test Charges.—For an experimental study of these forces a small charged body is needed, called a test charge, which can be moved from place to place in the neighbourhood of the charged pieces of metal; a good means of estimating the force acting on the test charge in any place also is needed. Moreover, the number of excess electrons or protons in the test charge must not change during the tests.

These requirements are reasonably well met by any small charged body, hanging by a silk thread. The thread cannot conduct electrons to or from the body; and the body hangs obliquely if any horizontal electrostatic force acts on it. So it can serve for a qualitative study if not for actual measurement.

Even the smallest body of visible size, however, is too heavy to show in this way the forces obtainable with an ordinary six-volt battery. This fact may be partly responsible for a common but mistaken impression, that "static electricity" is a different material from "current electricity." For this sort of study of the forces voltages far higher than six are needed. For example (albeit an unusual one), a battery of 20,160 Planté cells, each one a chemical test tube containing acid and electrodes, was built by John Trowbridge at Harvard university shortly before 1900 for measurements of the voltages required for the production of X-rays. With this battery, giving more than 40,000 v., electrostatic forces could be studied easily with a cork, a match or any such light object, hung by a thread from one end of a glass rod.

Traditionally, test charges carried in this way are pith balls, or pieces of dried willow pith, not necessarily round. A good pith ball, $\frac{1}{4}$ in. in diameter, shows forces well with about 10,000 v.

These voltages seem forbidding; with Trowbridge's battery, they were. If any two metal objects connected to its terminals were brought too close together, a bright blue spark, changing instantly to a big, writhing flame, showed clearly one of the relations of current electricity to static. Ordinarily, therefore, the study of static charges is made with apparatus capable of producing only relatively weak currents, although at high voltages, as described under the next heading.

Meanwhile it may be noted that with the ordinary household potential difference, 120 v., a limited demonstration of the existence of such forces can be given with an apparatus called a gold-leaf electroscope, invented by Abraham Bennet in the year 1787. Its simplest form is shown in fig. 7.

If the plate P, shown in this figure, is connected to one of the wires and the case C to the other, rod R and leaf L have charges of the same kind, so they repel each other. Of course rod R cannot move visibly, but the charges given by 120 v. swing the leaf to an angle of a few degrees from the vertical. For this experiment one may use alternating voltages just as well as direct. So with a small transformer, for example a 1-to-5, giving 600 v., the forces may be strengthened enough to swing the leaf out by 30° to 60°.

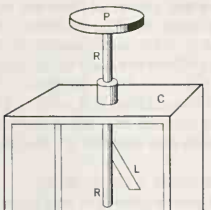


FIG. 7.—GOLD-LEAF ELECTROSCOPE, SHOWING METAL PARTS (*See* TEXT)

The electroscope gives a good check—as good as the spark with Trowbridge's battery—on the oft-repeated principle that the forces driving electrons through such a wire as a lamp filament are electrostatic. If its plate P and case C are connected to the terminals of a lamp, its leaf shows an electrostatic force only when the lamp glows.

2. Low-Current, High-Voltage Apparatus.—Historically, electrostatic forces were observed more than 2,000 years before the existence of any battery or transformer. Theophrastus, in 321 B.C., and probably Thales about 600 B.C., knew that pieces of amber, when rubbed with other dry substances, acquired an ability to attract feathers and bits of straw. In modern parlance, they acquired potentials of several thousand volts. So, after 1,800 years, the Greek word *elektron*, for amber, was used by Queen Elizabeth I's court physician William Gilbert (1544–1603) as a basis for his neo-Latin term for this force, *vis electrica;* and the word nearly regained its original form in the term "electron."

Appropriately enough, it is an excess of electrons, rather than a deficiency, that is so readily acquired by amber and by other resins. So this type of charge was named resinous electricity by C. F. de Cisternay du Fay (1698–1739), about 1733, when he discovered that there were two types. The deficiency type, being characteristic of glass, was named vitreous electricity. The unlucky guess on algebraic signs for charges was made later by Benjamin Franklin.

Many substances have strong preferences like these. Others, such as woolen cloth and cat's fur, which are good for rubbing either resin or glass, are easily overruled in any preferences of their own and charged with whatever is left to them. Always, as proved by Faraday in an experiment to be described later, the algebraic sum of the charges produced is exactly zero; if it were not zero, we could not explain charges this way.

Neither charge, for that matter, is big by modern standards. A piece of amber as big as the end segment of a man's finger may hold as many as 100,000,000,000 excess electrons, and it may acquire them in less time than a second; but at that, the current charging it is usually less than a ten-millionth of an ampere. Although it is a high-voltage generator, it is indeed of a low-current type.

Moreover, frictional charging is quite inefficient, in the sense that the electrical energy of the charges is only a tiny fraction of the mechanical work done in the charging process. This process seems to consist essentially of two parts, neither of them actually being friction. The first is the establishment of contact between the surfaces, one of which robs the other of electrons (as described above under *Batteries,* in terms of a "double layer" of charges). There the surfaces were those of conductors; here they are of insulators, such as amber and wool. When a piece of amber robs a woolen cloth of electrons to form a double layer between them, the electrons cannot flow freely in either material; so the double layer is formed only where the cloth actually touches the amber. Under a microscope the cloth looks rough, and with enough mag-

nification the amber also would look rough. So their surfaces actually touch only over a very small fraction of their area. To form a double layer over any large fraction of it, different parts of the surfaces must be used at different times. That is, the amber must be rubbed with the cloth; hence, there is friction. The energy of the charges, however, is not derived from the work done against friction. Instead it comes from work done after the rubbing. This work is done when the cloth is pulled away from the amber, dragging its excess protons away, against the electrostatic force of attraction exerted by the excess electrons on the amber. This dragging away is the second of the two essential parts of the charging process, mentioned above; the friction is only incidental, even though it is what requires most of the work.

For charging a little piece of amber, perhaps as a parlor trick, this waste of work is unimportant, but beyond this it is tiresome. Consequently, although friction machines containing spinning globes, usually of glass, rubbed by various stationary objects, were used for electrical studies throughout the 18th century, they are now only museum pieces. They were abandoned in favour of such devices as the Wimshurst machine, also called an influence machine or static machine. This is based on a device invented in 1775 by Volta, who called his device an electrophorus, or carrier of electricity. Its essential parts are a plate of some good insulating material such as glass, wax or sulphur, not too smooth, lying on a non-insulating base, to keep its lower surface at ground potential, and a metal plate carried by an insulating handle. Its operator first charges the upper surface of the insulating plate by friction. Then he lays the metal plate on top of it and touches the metal with his finger. This contact permits electrons to flow through him in obedience to attraction or repulsion by the charge on the insulating plate, giving the metal a charge opposite in sign to that of the insulator. (The charge is usually much stronger than the one which would have been given by their contact potential difference alone, if the insulating plate had not been charged before the metal plate came near it.) This type of charging is called electrostatic influence, or sometimes electrostatic induction, though the latter term is contraindicated by another technical use of the word "induction." Then the operator takes his finger off the metal plate and uses its insulating handle to carry it and its charge away and give the charge to some other piece of metal. The charge on the insulator does not escape except from the few high spots where the metal actually touches it, so its influence can be used repeatedly.

This is a slow, laborious process. But actual contact of the metal plate with the insulating plate is unnecessary, as an equivalent process is to carry metal plates past the insulating plate on a big, rotating glass disk, with a little, stationary wire brush playing the part of the operator's finger. The Wimshurst machine does this in duplicate, having two insulating plates oppositely charged; and it increases the charge of each of them by diversion of some charge from the metal plates influenced by the other one. The only limit to the increase is one imposed by leakage over the insulators. Mechanical details are nonessential here and are indeed variable.

Machines of this sort, commonly used for lecture experiments on electrostatics, usually have glass disks about two feet in diameter, which are spun at about 100 r.p.m. by a belt from a hand wheel. Such a machine can produce a few tenths of a milliampere (0.001 amp.) at about 50,000 v. A similar machine with four disks about six feet in diameter can run a fairly powerful X-ray tube taking several milliamperes at about 100 kv.—too much power for hand operation. One milliampere × 100 kv. is equivalent to 1 amp. × 100 v., or 100 w., the power taken by a medium-sized lamp. Moreover, the watt is one newton-metre per second, the newton is about 0.22 lb. of force and the metre is about 3.3 ft.; so 100 w. is about 72 ft.-lb. per second. Lifting a 72-lb. weight one foot every second soon becomes tiresome; therefore a big Wimshurst machine requires a motor to drive it. Even with a hand-operated lecture-table machine it is easy to feel the increase in driving effort as the charges are built up, because force is required to pull each moving plate away from the stationary one under whose influence it has been charged.

Wimshurst machines are quite unlike batteries in their relation of terminal potential difference to current. Indeed, since they carry charges mechanically in discrete pieces, they are essentially analogous to piston pumps.

Recalling what was said about the relation of so-called "current electricity" to "static" above, another aspect of this relation is shown by the Wimshurst machine. A current of 0.1 milliampere is easily measured by a sensitive ammeter (often called a milliammeter or a microammeter), which is deflected like other ammeters by electromagnetic forces. When such an instrument measures the output of a Wimshurst machine, it is measuring as current electricity the little static charges delivered by the moving plates.

Another static machine, which can give potentials far too high for lecture-table experiments but good for very penetrating X-rays, is the Van de Graaff generator. Externally this is an insulating tube, several feet tall and a foot or two in diameter, set up as a column to support a still wider, hollow metal sphere. Within the sphere is a pulley, driven by a belt made of silk or rubber, which in turn is driven by another pulley in the base of the column. Near the lower pulley, electric charge is sprayed onto the ascending part of the belt from a metal comb with sharp points, maintained at a potential high enough to cause corona, or glow discharge, in the air near them. The charge thus acquired by the belt is carried up into the sphere, where it comes near enough to another metal comb to cause corona to it and thus deliver the charge to the sphere. By a special arrangement of resistors connecting this comb to the sphere, potential differences are produced, such that another comb within the sphere can be made to spray the descending side of the belt with charge of the other sign, which is taken off at the lower pulley. Machines of this type, enclosed in tanks of compressed air to suppress undesired corona currents, have produced potential differences greater than 4,000,000 v.

Cyclotrons, to be sure, are rated in terms of voltages often well over 4,000,000. However, a cyclotron does not contain any parts at such a high potential difference. Instead, it gives energy to gas ions (not ions in an electrolyte, but charged atoms in a highly rarefied gas) by repeated application of a relatively low potential difference.

Some other types of apparatus, to be described later, also surpass the Van de Graaff in energy given to an ion in a gas, or to an electron in a practically perfect vacuum; but they do this with electric fields of essentially nonstatic types, to which the concept of potential difference does not apply.

Finally, recalling Trowbridge's 40,000-v. battery and its use for X-ray measurements, it is evident that any source of power suitable for an X-ray tube gives a potential difference quite ample for electrostatic experiments. Most modern X-ray systems involve transformers, with alternating currents and potentials; but by various devices, also to be described later, most of them allow only currents in one direction to reach the X-ray tube. The only disadvantage in such a system for electrostatic experiments is that it is too powerful for safety unless the experimenter protects himself by connecting his apparatus to this system only through very high-resistance resistors. However, these are simple: ordinary glass laboratory tubing, several feet long, full of alcohol, has many millions of ohms. With them one may safely demonstrate the laws of electrostatic forces and also reinforce the relations between static charges and currents with the fact that the charges he is using have come from a most essentially electromagnetic device, the transformer.

B. Coulomb's Law

In the section *Electrons* it was stated that every static electron repels every other one; also that the strength of this repulsion varies inversely as the square of the distance between them (provided this distance is longer than the diameter of an atom); also that at one inch the force is 8.038×10^{-26} lb. In the next section, *Protons*, it was stated that static protons repel each other according to the same law, and that a proton and an electron attract each other likewise. Consequently all charged bodies repel or attract each other, though we cannot expect the inverse-square law to apply to whole bodies accurately unless their own

dimensions are all very small in comparison with the distance between them.

Historically, this law for charged bodies was not discovered by any deduction like this from laws about pairs of electrons or protons. On the contrary it was first found by deduction from a less simple law (to be stated later), which is regarded as one of its corollaries. The deduction was made in 1766 by Joseph Priestley (1733–1804). Nevertheless Priestley's fame rests less on this than on his discovery of oxygen, and his law came to bear the name of Coulomb, who made a series of direct experimental tests of forces between small bodies in 1785.

Because of the weakness of electrostatic forces between charged bodies in air, Coulomb had to invent delicate apparatus for these tests. With it, however, he proved three important parts of his law. One of these was that if the charged bodies really could be reduced to negligible size—to point charges, as we call such idealized bodies—their static forces would indeed be accurately proportional to the inverse squares of their distances.

A second part of his law was that if the charge on either body was changed (for example, halved by sharing with a similar body which was then removed) the force was changed in proportion to this charge.

The third part of Coulomb's law, proved mathematically from the second, was that the force exerted by any point charge on any other one is unaffected by the presence or absence of any third charge.

Calling the quantities of electricity in two point charges Q_1 and Q_2, and the distance between them R, the three laws just stated may be summed up in one law: the force between Q_1 and Q_2 is proportional to $\frac{Q_1Q_2}{R^2}$, regardless of any Q_3.

Coulomb could not measure his quantities of electricity in coulombs. The coulomb is defined only as an ampere-second, and Ampère did not demonstrate how to measure currents until after Coulomb had died. But Coulomb's law provided another way to define a unit of charge. This unit, called the electrostatic unit of charge, is the value of either one of two equal point charges which repel each other with a force of one dyne (or 10^{-5} newton) when placed one centimetre apart.

Expressing the values of any point charges Q_1 and Q_2 in electrostatic units, their distance in centimetres and the force in dynes, the force is not only proportional to $\frac{Q_1Q_2}{R^2}$, but also numerically equal to this fraction.

1. Coulomb's Law in Coulombs.—For most purposes, as already stated in the section *Electrostatic Unit of Charge*, this unit is much too small for convenience. So an expression for Coulomb's law in coulombs is needed. Since the coulomb is an ampere-second, we must go back to the definition of the ampere (the current in each of two infinitely long, parallel wires, which would make them attract each other with a force of just 2×10^{-7} newton on any part of either wire, of length equal to their distance apart). So this 2×10^{-7} newton can be expected to enter into any expression for Coulomb's law in coulombs.

As stated above under *Why Do Like Currents Attract?*, the ratio of their attraction to the electrostatic repulsion that would be found if the current-carrying electrons had no protons to balance them is the square of the ratio of their speed of flow to the speed of light. And under the section, *Special Relativity*, it was shown that this value for the ratio was no chance coincidence, but an essential part of the fundamental law of electrostatic forces. So when we express Coulomb's law in coulombs, defined through forces between currents, we can expect the square of the speed of light, as well as the 2×10^{-7} newton, to come into it. This expectation is verified.

Along with the definition of the ampere in terms of this force there was a statement of how one could calculate such forces for currents of other strengths, but this statement was not put into the form of an equation. Conventionally, the equation for it is written as

$$F_m = -\frac{\mu_0 I_1 I_2 l}{2\pi r}$$

The minus sign here goes with a conventional definition of repulsions as positive and attractions as negative. l is the length of the piece of wire for which F_m is calculated, and r is the distance between the wires. μ_0 is a constant, whose value is set by definition at exactly

$$\mu_0 = 4\pi\times10^{-7}\frac{\text{newton}}{\text{ampere}^2}$$

One may well wonder what right we have to set any constant of an equation like this by definition. To explain, we may note that if we make $l=r$, as in the definition of the ampere, and likewise $I_2=I_1=1$ amp., the defined value of μ_0 makes this equation give $F_m=2\times10^{-7}$ newton. So this definition of μ_0, when used with the equation for F_m, is only an alternative form of the definition of the ampere.

One may still wonder why we put the 4π into this definition and 2π into the equation for F_m, rather than a simple 2 into the definition and no number at all into F_m. Strictly speaking, the 4π and 2π are unnecessary, and formerly they were not there. But without the 4π in μ_0, a factor 4π appeared explicitly in so many equations related to this one that it was a nuisance. So it paid well to simplify these other equations or "rationalize" them, by hiding their 4π in μ_0, even though it required a 2π to appear in this equation. For reasons more historical than logical, this constant μ_0 is called the permeability of free space. The best that can be said for this name here is that the equation would not give the whole electromagnetic force on the piece of wire if the space between or near the wires was not magnetically free space. Magnetically free space here means space free from iron or other magnetizable material. Nevertheless, even in the presence of iron this equation would still give the part of the force on either wire that is due directly to the other wire, rather than to the iron.

For like currents of equal strength I, this equation reduces to

$$F_m = -\frac{\mu_0 I^2 l}{2\pi r}$$

If now the wires are alike, each containing free electrons whose speed of flow is v and whose charge per unit length is q $\left(\text{measurable in } \frac{\text{coulombs}}{\text{metre}}\right)$, we have $I=vq$ and

$$F_m = -\frac{\mu_0 v^2 q^2 l}{2\pi r}$$

These wires come under the requirement of relativity, that the electrostatic repulsion between these free electrons, with no protons, would be $\left(\frac{c}{v}\right)^2$ times this attraction. So this electrostatic force must be

$$F_s = +\frac{\mu_0 c^2 q^2 l}{2\pi r}$$

Of course neither wire is a point charge, like those of Coulomb's law. But its current-carrying electrons can be compared to an array of point charges, strung on it like beads; and one can show mathematically that for a pair of such arrays Coulomb's law would lead to a formula,

$$F_s \text{ is proportional to } \frac{q^2 l}{r}$$

Comparing this with the F_s just found from relativity, it may be shown that for point charges the force is

$$F = \frac{\mu_0 c^2 Q_1 Q_2}{4\pi R^2}$$

This is called Coulomb's law, even though Coulomb had no way to find its constant factor, $\frac{\mu_0 c^2}{4\pi}$. Since $c=299{,}793{,}000$ m. per second, or nearly 3×10^8 metre per second, this is

$$\frac{\mu_0 c^2}{4\pi} = \text{nearly } 9\times10^9 \frac{\text{newton-metre}^2}{\text{coulomb}^2}$$

There are more units in this fraction than in any other in this article; but when Q_1 and Q_2 are expressed as numbers of coulombs, and R as a number of metres, the units obviously cancel down to give F in newtons.

With this cancellation the numeric 9×10^9 tells us readily why static charges on small bodies (such as a piece of amber) are very

small fractions of a coulomb. If Q_1 and Q_2 could be each a whole coulomb, with R one metre, there would be nothing but these units to cancel, and no arithmetic to be done, so we should find $F = 9 \times 10^9$ newtons. This is almost 10^9 kg., or 1,000,000 tons. Small bodies charged with whole coulombs would be as destructive as bombs.

Conversely, to keep the forces between small charged bodies so weak that they can lift only such things as bits of paper, but not coins, we must have F of the order of 0.01 g., or 10^{-4} newton, when R is only about a centimetre. With these values, if $Q_1 = Q_2$, each Q is only about 10^{-9} coulomb, or a millimicrocoulomb.

Even so, this is a vast number of electrons. Since a whole coulomb is 6.2420×10^{18} electrons, this little charge is about 6,000,000,000 electrons.

Conversely, one electron is only 1.60206×10^{-19} coulomb. With Q_1 and Q_2 each at this value, Coulomb's law in the algebraic form just given tells us that the force between two static electrons an inch apart is 3.575×10^{-25} newton; this is 8.038×10^{-26} lb.

2. The Restriction to "Free Space."—Conventionally, Coulomb's law is written either as above or in an equivalent form,

$$F = \frac{Q_1 Q_2}{4\pi\epsilon_0 R^2}, \quad \text{with} \quad \epsilon_0 = \frac{1}{\mu_0 c^2}$$

This ϵ_0 is called the permittivity of free space—a name, like that of μ_0, more historical than logical. But again we can say for it that if the space is not "free" the equation does not tell the whole story. That is, the equation does not give the resultant force that acts on either charge, though it does give the component of this force that is attributable directly to the other charge.

In this case the "freedom" is not simply from magnetizable matter but from matter of any kind. Iron is no worse here than brass, which is nonmagnetic. The trouble with any matter here is that it is subject to displacement of electrons or ions by electrostatic influence. Metals and electrolytes are the worst causes of such trouble, but even insulators cause serious complications.

Formerly the forces occurring in insulating liquids, such as most oils, were thought to be of great theoretical significance. They seemed to reveal important facts about the nature of electrostatic forces. Now, on the contrary, we know that such forces are not entirely electrostatic, after all. They contain a large admixture of hydrostatic forces, like buoyant forces on floating bodies. These are caused by crowding of the oil molecules against the charged solid bodies, under electrostatic attractions on submolecular charges displaced within these molecules. Altogether, therefore, forces on bodies immersed in oil are rather complicated, and they are not of enough importance, either theoretical or practical, to deserve detailed analysis here.

In conclusion, however, we should note well that the equations written above, both for point charges and for wires, tell the whole story only when the space near these bodies is "free."

C. Some Corollaries of Coulomb's Law

Almost every law of electrostatics can be worded as a corollary of Coulomb's law. For many laws, such wording would have to include special conditions; but some of the most important laws can be deduced almost unconditionally from Coulomb's.

1. The Potential-Difference Law.—Among the latter is the law stated before, under *Power and Electrostatic Forces*, that the work done by such forces on an electron going from one point A to another point B is the same for all paths from A to B. This law was described there as *sine qua non* for most of the convenience of operation of household appliances. Next it was used as the basis for the definition of potential difference. So every mention of a potential difference is further testimony to its importance.

To deduce it from Coulomb's law we need not use any algebra, unless it is algebra to say that the lack of any symbol for the direction of the line between Q_1 and Q_2 in the equation shows that the strength of the force on either charge does not depend on this direction. This aspect of the law will now be applied to fig. 8.

Suppose an electron Q_1 makes several trips from point A to point B here, by different paths, with electron Q_2 staying always

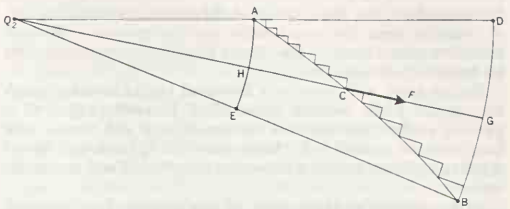


FIG. 8.—ALTERNATIVE PATHS FROM A TO B FOR AN ELECTRON Q_1 IN THE PRESENCE OF ANOTHER ELECTRON Q_2. WHEN Q_1 IS AT C THE FORCE F IS AS SHOWN BY ARROW CF (*See* TEXT)

where it is shown. If Q_1 follows the curved path ACB, the force F is always obliquely forward, doing work on it, but the amount of work done may not be obvious. If Q_1 goes from A straight to D, however, and then along the arc DGB, all the work is done on the line AD and none on the arc DGB. Likewise if Q_1 goes first along the arc AHE and then along line EB, all the work is done along EB. Moreover, since the strength of the force on Q_1 does not depend on the direction of Q_1 from Q_2, the same amount of work is done on either of these simple paths, ADGB and AHEB, from A to B.

Extending this reasoning, the same work is done again if Q_1 follows the zigzag path ACB, consisting of short radial lines and arcs.

This path is like a flight of stairs, so it may be related at once to the case used under *The Watt*, above, to illustrate the definition of work, namely the work done by gravity on a man going downstairs. If he walks smoothly, his centre of gravity can follow a line related to the stairs essentially as the smooth curve ACB is related to the zigzag in fig. 8. Then the definition illustrated by that man makes the work done on Q_1, along the smooth curve ACB, or any other path from A to B, exactly the same as along ADB or AEB.

If now there is another electron, Q_3, at some fixed position other than that of Q_2, we may recall another aspect of Coulomb's law, namely that the force exerted on Q_1 by Q_2 is independent of the presence or absence of Q_3. Therefore the amounts of work done on Q_1 by Q_2 and Q_3 are independent of each other. Since each of these amounts is independent of the path taken by Q_1 from A to B, their sum is independent of the path.

So also, therefore, is the work done on Q_1 by the forces due to any system of electrostatic charges, no matter how complicated the system may be. This makes the law, that the same potential difference applies to all paths from A to B, a corollary of Coulomb's law.

2. Charges on Surfaces.—Although we did not have to use the R^2 in Coulomb's law explicitly in deriving the potential-difference law, we should have to use this R^2 in deriving almost any other of the corollaries of Coulomb's law. Moreover such uses of this R^2 usually involve mathematics higher than algebra.

Among the most important of these is a law that if any conductor is charged, but is not carrying any current, all its charge is on its surface.

This law was not discovered as a corollary to Coulomb's but by experiment. Franklin found it first, in the case of a metal cup, which showed no appreciable charge on its inner surface, except near its edge, even when its outer surface was very strongly charged. Franklin asked Priestley to investigate this difference more thoroughly; and this law was the one used by Priestley in his deduction of what we call Coulomb's law.

3. The Accuracy of Coulomb's Law.—Priestley's deduction actually gave more accurate tests of Coulomb's law than Coulomb's own method of testing it. Henry Cavendish (1731–1810) enclosed a sphere, coated with tin foil, within a pair of similarly coated hemispheres which could be removed; he connected them all together with a wire, then charged the system, then pulled the wire away with a silk thread (which could not conduct any charge) and finally removed the hemispheres and tested the inner sphere for charge. If the law of force ought to have had an exponent of R appreciably greater than 2, he would have found a charge on the

inner sphere, of the same sign as on the outer; if appreciably less, of opposite sign. He found none, and his tests were sensitive enough to assure him that the correct value for this exponent must lie between 1.98 and 2.02.

Similar tests, with sensitivity increased by the technical progress of 100 years, enabled James Clerk Maxwell (1831–79) in 1878 to confine this exponent between 1.999,9 and 2.000,1; and 60 more years enabled S. J. Plimpton and W. E. Lawton (*Physical Review,* 1936) to confine it between 1.999,999,998 and 2.000,000,002.

In the section *The Accuracy of the Fundamental Laws,* above, it was noted that the fundamental laws of electricity are known far more accurately than any laws involving properties of specific materials. Here is an example: no property of any specific material has ever been measured with the accuracy of the Priestley-Coulomb exponent, 2.

Moreover (and this also is typical of the contrast between these types of laws), measured properties of specific materials do not turn out to be simple one-digit numbers, like this 2. The laws of electricity are simpler than those of matter.

4. Electrostatic Shielding and the Conservation of Charge.—The law of Franklin and Priestley about charges on surfaces is of inestimable practical importance, because it enables us to shield any electrical apparatus from electrostatic forces caused by other things, simply by enclosing the apparatus within a metal box. Even a cage will often suffice; but to make the shielding strictly perfect the box must form a completely closed surface.

Moreover the shielding works both ways: apparatus within the box cannot exert any electrostatic forces on anything outside it. Shielding in this direction, however, results from one of the most fundamental of the many laws discovered by Michael Faraday along with the law of Franklin and Priestley.

Their law alone gives us Coulomb's inverse-square factor, as just noted, from this factor Carl Friedrich Gauss (1777–1855) made a mathematical deduction called Gauss's theorem. One of the corollaries of this theorem is that, if any charged body is completely enclosed in a metal box, its attraction holds a charge, equal and opposite to its own, on the *inner* surface of the box. This corollary had been discovered experimentally by Faraday, and it was used by him in planning the discovery of the law involved in shielding.

For this he built a box big enough to hold himself and all the apparatus he wanted, including electrical machines of the frictional and Wimshurst types. He coated the outside of the box completely with tin foil, then mounted it on insulators and connected the tin foil to a sensitive electroscope outside it. He got inside, closed the door and tried experiments with the machines, while an assistant outside the box watched the electroscope.

If the machine could produce a charge of either sign without producing any of the other sign, Gauss's theorem would require this charge to attract an equal-opposite charge to the inner surface of the tin foil on the box. With the box up on insulators, this inner-surface charge would have to come through the tin foil from the outer surface; so it would leave on the outer surface a charge like the one produced by the machine. A part of this outer-surface charge would be on the electroscope.

If, on the other hand, the machine charged two bodies with equal-opposite charges, their requirements on the tin foil would also be equal-opposite, so they would ask nothing of its outer surface. Likewise if one terminal of the machine was connected to the inner surface and the other terminal to something within the box, and the machine again produced equal-opposite charges, they too would ask nothing of the outer surface. And Faraday's assistant, watching the electroscope, never saw any effect on it in any of the experiments. Evidently all charges were produced in pairs, those of each pair being equal in strength and opposite in sign.

This law is called the Law of Conservation of Charge because its mathematical form is that the algebraic total of charges in the universe cannot be changed.

With modern knowledge about electrons and protons, Faraday's discovery appears as a simple consequence of the fact that he could not create or destroy either an electron or a proton. He could only transfer electrons from one body to another. So we use the law with far more confidence in its accuracy than if we had to base our belief in it directly on his experiments. This confidence is justified; the law holds, not only in electrostatic shielding, but also in innumerable and infinitely varied applications to other phenomena.

On this basis we extend it still farther. Although Faraday could not create or destroy an electron, we now can do so by using some of the apparatus of modern nuclear-physics laboratories. But when we create an electron, we always create along with it what we call a positron, a particle with the mass of an electron but the charge of a proton; so the rule of equal-opposite pairs still holds. The positron is very short-lived—it has no time for taking part in the conduction of current in a wire; its life ends along with that of some other electron, destroying equal-opposite charges and converting the masses into gamma-ray energy. Positrons and other temporary particles are discussed above, under *Ephemeral Particles.* For now the point is that even in modern high-energy phenomena Faraday's law of conservation of charge seems to be inviolable.

5. Point Charges, Straight Wires and Flat Plates.—Starting from the case of ideal point charges, for which the R^2 of Coulomb's law holds so accurately, we find less perfect accuracy, as already noted, for charged bodies of appreciable dimensions, though the accuracy is still good if these dimensions are all small in comparison with R.

This is true because: (1) the force exerted by one such body on another is really the resultant of all the little forces exerted by electrons or protons (*ie.,* little point charges) in the one body on those in the other; and (2) each such little force has its R practically equal to any R between the whole bodies, so long as they are small.

For larger bodies also we can combine little forces in this way; but naturally we cannot often expect any R for large bodies to serve even approximately for all pairs of electrons or protons in them. Most forces between large bodies therefore obey equations very different from Coulomb's—even though these equations, being found by calculation from his, are corollaries of his.

For example, the pair of very long, straight wires discussed above under *Coulomb's Law in Coulombs* exert forces on each other inversely proportional to the first power, rather than the square, of the distance between them.

In fact, this contrast appears without going so far from the case of two point charges. One point charge and one wire will suffice for it. We may call the point charge Q and its distance from the nearest point on the wire r (to distinguish it from Coulomb's R to any other point on the wire) and assume the charge on the wire to be spread along it uniformly with a line density q (measurable in coulombs per metre, not simply in coulombs like Q). Then if r is short in comparison with the distance from Q to the nearer end of the wire, the force exerted by the wire on Q is approximately

$$F = \frac{qQ}{2\pi\epsilon_0 r}$$

Comparing this equation with Coulomb's, it is evident that when one of Coulomb's point charges is spread out uniformly on a very long line, the relation of force to distance is changed from the inverse square (R^2 in the denominator) to an inverse first power (r in the denominator). Since this change results from spreading in one dimension, one may expect a further change of the same sort on spreading in two dimensions, all over a big flat plate. We find it: the exponent of distance, already reduced from 2 to 1, now drops to 0; *i.e.,* the distance disappears from the equation entirely. At least, it does except for a proviso that the equation holds only if this distance is small in comparison with the distance from Q to the nearest edge of the plate. Then if the charge on the plate has a uniform surface density p (measurable in coulombs per square metre, unlike either q or Q) the force is approximately

$$F = \frac{pQ}{2\epsilon_0}$$

This equation, with its lack of any explicit mention of distance, is very different from Coulomb's. Many other equations of electrostatics seem even more different. Nevertheless, as noted above, the force described by any one of them is a resultant of little forces between pairs of point charges, each pair obeying Coulomb's law.

Moreover the direction of any such force is what one might expect of such a resultant. In these simple cases, the force acting on Q is along the line from the nearest point in the wire or plate.

D. Measuring a Force on One Electron

The law about a point charge near a flat plate was used by Robert Andrews Millikan (1868–1953) to give the most direct measurements of the charge in one electron. Millikan used two plates, both horizontal, with charges of equal strength but opposite sign, 16 mm. apart; and he held his point charge Q on a microscopic droplet of oil, about midway between the plates. By changing the surface densities $+p$ and $-p$ of charge on the plates, he could either drive the droplet upward or let it settle downward or hold it in mid-air like Mohammed's coffin.

With two plates the force on Q was twice as strong as it would have been with only one; and it was even more exactly independent of distance from either plate than with only one, because any motion away from one plate was toward the other. So the force kept a constant value,

$$F=\frac{pQ}{\epsilon_0}$$

even if the droplet moved all the way from the lower plate to the upper. Therefore, calling this distance h, the work done on Q was Fh.

Another view of this work is given by the definition of potential difference, because the work was done by an electrostatic force, acting all the way on the same quantity of electricity, Q. So, calling the potential difference between the plates V, the work was QV.

Now equating these values for the work, we have

$$Fh=QV$$

To measure the charge in one electron, therefore, he had only to find out how many electrons there were in the charge Q and to measure the force F, the height h and the potential difference V.

"Only" to take these few measurements, however, is much easier to imagine than to perform. To be sure, it was easy to measure h, which was 16 mm. in the apparatus described by Millikan in his book *The Electron*. Also V, which was often about 5,000 v. But to measure F was far from easy, because he had to use exceedingly small droplets and correspondingly light forces to enable him to find out how many electrons there were in Q.

To count these electrons, he could start with a droplet having no charge at all. There were many such; they could be identified by watching the droplets in a microscope and finding ones which were completely indifferent to the charges on the plates, so that they would keep on settling just as fast when the plates were highly charged as when they were discharged. Then if the air around such a droplet were ionized by a flash of X-rays, shortly after the flash the droplet might acquire a charge, suddenly, and become responsive to the plates. This charge was the charge of an ion which happened to strike the droplet. An ion in this air was a molecule which had either lost one electron or picked up one electron, astray from some other molecule. When only one ion landed on the droplet, therefore, the charge acquired by the droplet was either an excess or a deficiency of just one electron.

Now, watching the droplet in his microscope, he could adjust the potential difference V so as to stop the slow descent of the droplet and hold it practically at rest. Then he could give the air another flash of X-rays, and the droplet might gain or lose another electron. So it was by counting such gains and losses that he found the number of electrons in the charge Q. This number, therefore, could be either positive or negative. It was often as high as 10, 20 or 30, and sometimes even over 100.

Nevertheless the force F needed for supporting the droplet or for driving it upward, not too fast for timing it, was exceedingly weak. Therefore the droplets had to be much too small for weighing. They were even too small for measuring their diameters in the microscope. So the only way to find their weights was from the speeds with which they settled when F was reduced to zero by discharging the plates.

Even this was not so simple as it might look. Such a very small droplet does not "see" the air as a continuous medium but as a lot of fast-moving molecules, filling only about a thousandth of the space in which they move and bumping irregularly into the droplet. So the calculation of weight, and from it of F, required very careful thought and mathematical analysis, with data on many droplets, each droplet being alternately pushed up by F and allowed to settle back many times.

Altogether, though the basic equation $Fh = QV$ is simple, and it is easy to say $Q = ne$, with n the number of electrons and e the charge in one electron, so that

$$e=\frac{Fh}{nV}$$

nevertheless in the work involved in his use of this equation Millikan earned thoroughly his Nobel prize.

1. Best Value for the Electronic Charge.—While Millikan's method for measuring e is the most direct, it is not the only method. Most others, however, determine the values of various fractions or products, containing e and other fundamental constants, such as the mass of an electron or the speed of light; and usually these constants appear with different exponents. To find the "best value" for any one of these constants, therefore, requires much ingenuity and study. It requires repetition, also, every few years, because of new experiments giving data more accurate than older ones.

Pioneer work on methods of analysis, to give "best values" for all the fundamental constants in one research, was done in the 1920s and 1930s by Raymond T. Birge. Further improvements were made in the late 1940s and in 1950 by Jesse W. M. DuMond and E. Richard Cohen and by J. A. Bearden and H. M. Watts. In 1953, DuMond and Cohen revised the "best values"; and another revision appeared in 1957 in a book, *Fundamental Constants of Physics*, by Cohen, K. M. Crowe and DuMond. It is from this book that the values of e, c and other such constants in this article have been copied or calculated. These authors also estimated the probable errors in the values of the constants. Thus e = (1.60206 $\pm$ 0.00003) $\times 10^{-19}$ coulomb, meaning that the probable error in the 1.60206 is 3 units in the last digit, either up or down.

E. Condensers (Capacitors)

Millikan's pair of plates may serve here not only with regard to e but also as a prototype for a very useful device, known long before his experiment and called a condenser or capacitor.

To see how a condenser condenses electricity, we may recall that the charges on his plates had surface densities $+p$ and $-p$, which came into the equation for the force F acting on the charge Q of his droplet, namely

$$F=\frac{pQ}{\epsilon_0}$$

Solving this equation for p we find

$$p=\frac{\epsilon_0 F}{Q}$$

and the work equation $Fh=QV$ enables us to write this as

$$p=\frac{\epsilon_0 V}{h}$$

From this equation it is evident that if the plates are kept at a constant potential difference V, say by connecting them to the terminals of a battery, we can increase the surface density p—condense more electricity on each plate—simply by decreasing the distance h between the plates.

1. Condensers as Analogues of Springs.—To see why condensers are useful, we may compare them with a clock spring, or

still better with the spring in a wound-up toy. Such toys run vigorously at first but more slowly as they "run down." A close analogy to this change appears in the circuit shown in fig. 9.

When the switch S, shown there, is closed, electrons flow as shown by the arrowheads, escaping from the plate which has excess electrons and entering the plate which is deficient. Since this flow decreases both the excess and the deficiency, it decreases the electrostatic forces which are causing the flow. So the flow slows down—just like the wound-up toy.

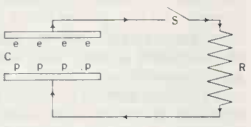


FIG. 9.—CONDENSER C WITH (e) EXCESS ELECTRONS AND (p) EXCESS PROTONS (ELECTRON DEFICIENCY) ON ITS PLATES: CONNECTED TO RESISTOR R THROUGH SWITCH S

One may wonder why anyone should want a circuit like this, which runs down, when a similar circuit with a battery in place of the condenser (as in fig. 5) can keep running steadily. So it should be noted here that condensers and resistors are used together in every radio receiver. To be sure, they do not form the whole circuit, as in this simple case; but the tendency for the current to "run down," shown here, is the essential feature for which condenser-resistor lines are used there.

Another property of mechanical springs is their ability to make things vibrate freely, with well-defined periods. This is illustrated best by the hairspring and balance wheel of a watch or clock; but it appears also in most musical instruments, though less clearly because the spring and the thing controlled by it often are combined in the same body, as for example in a violin string. This property of a spring also is found in condensers; but not with circuits consisting primarily of resistors. To show free vibrations, more technically called oscillations when they are electrical, the rest of the circuit must have low resistance and strong magnetic properties. These magnetic properties will be defined later; for now, we may note that such a circuit can be "tuned" like a musical instrument, and that tuning a radio receiver to a station is primarily changing its condensers in a way analogous to tightening or loosening a violin string.

This is done by narrowing or widening the area of each plate which faces the other plate, as shown in fig. 10.

If the plates are staggered as shown there, almost all the charge on each plate is on the part which faces the other plate, being held there by the electrostatic influence of the charge on the other plate. If the dimension a is widened, therefore, each plate can hold more charge than before, at any given potential difference; or, conversely, there is a lower potential difference with any given charge. So this widening of effective area corresponds to loosening the violin string.

The tuning condensers in a radio receiver are built on this principle. Usually they differ from fig. 10 only in having more than two plates and in accomplishing the change of area by rotation of each moving plate in its own plane, around a shaft near one of its edges.

FIG. 10.—METHOD OF VARYING CAPACITANCE OF CONDENSER PLATES BY ALTERING AREA a IN WHICH THEY OVERLAP (*See* TEXT)

2. Capacitance.—To put these rules into quantitative form we may call the effective area A (corresponding to the width a in fig. 10) and use the equation for charge density,

$$p=\frac{\epsilon_0 V}{h}$$

Then calling the charges on the plates $+Q$ and $-Q$ (and noting that this Q does not refer to any oil droplet) we have

$$Q=pA=\frac{\epsilon_0 A}{h}V$$

When Q and V change (as in the running down of the condenser-resistor circuit) while the plates keep fixed positions, this fraction $\frac{\epsilon_0 A}{h}$ stays constant. Therefore it is given a name, capacitance, and a conventional symbol, C, and we write

$$Q=CV$$

This name "capacitance" is derived from an older name, "capacity." Capacity was partially appropriate, because at any given V the quantities of electricity in different condensers are proportional to their C's; but the name "capacity" was misleading. A jug, of capacity one gallon, can hold one gallon of water, but no more. The quantity of electricity Q held by either plate of a condenser has no such relation to its C. We do better, therefore, to keep only a suggestion of the idea of capacity, in the name derived from it; and there is also an advantage in following the rule, stated above under *Other Terms and Units*, that gives capacitance a termination like resistance.

Since a condenser is characterized by its capacitance, as a resistor is by its resistance, it is often called a capacitor.

To define a unit of capacitance we may recall how the ohm was defined. We wrote Ohm's law in the form

$$R=\frac{V}{I}=\text{constant}$$

then we defined unit of R as

$$1\text{ ohm}=1\frac{\text{volt}}{\text{ampere}}$$

So now we write

$$C=\frac{Q}{V}=\text{constant}$$

then the unit named in honour of Faraday is

$$1\text{ farad}=1\frac{\text{coulomb}}{\text{volt}}$$

Since the coulomb itself was defined as one ampere-second, the farad is evidently related closely to the ohm. More specifically,

$$1\text{f}=\frac{1\text{ coulomb}}{1\text{ v.}}=\frac{1\text{ amp.-sec.}}{1\text{ amp.-ohm}}$$
$$\therefore\ 1\text{ farad}\times 1\text{ ohm}=1\text{ sec.}$$

This relation has a direct application to any condenser-resistor circuit, either the simple one of fig. 9 or any of those derived from it and used with radio tubes. The product CR of the capacitance C and resistance R of that circuit is evidently expressible in seconds; so it is a quantity of time. Its significance is this: the law by which the circuit "runs down" is such that in a time interval CR, starting at any instant we choose, the charge Q and the potential difference V and the current $I\left(\text{which}=\frac{V}{R}\right)$ all decrease to 0.368 of the values they had at our chosen starting instant.

To a mathematician this means that

$$Q=Q_0 e^{-\frac{t}{CR}}$$

where $e=2.71828\ldots$, with similar equations for V and I; and the 0.368 is e^{-1}.

The capacitances of most condensers are very small fractions of a farad. Most of them are small enough to be expressed most conveniently in microfarads or millimicrofarads, with *micro-* and *milli-* having their usual meanings of one millionth and one thousandth, respectively. So the time constant CR of any circuit of this type is usually a very small fraction of a second unless the resistance is very high.

3. Insulating Materials in Condensers.—If the plates of a condenser do not have to be movable, the easiest way to make sure that they will keep their separation h constant is to attach them to opposite sides of a sheet of glass or some other good insulating material.

Such condensers need not be flat. Sheets of tin foil, glued to the inner and outer surfaces of a glass bottle, make a condenser of a type that was historically the first and is still useful. Its

ability to "condense" charges was discovered in 1745 by E. G. von Kleist of Pomerania; but it is called a Leyden jar because it was first investigated thoroughly by Pieter van Musschenbroek (1692–1761) of Leiden, who discovered its properties independently in 1746.

Tin foil or any other thin metal foil is useful also in condensers of a later and very common type, in which the two sheets of metal foil are separated by very thin waxed paper. The foil and paper, together, are then folded into a compact mass and held firmly within a metal box. These condensers usually have capacitances of the order of one microfarad, whereas a Leyden jar, though much bigger in over-all dimensions, rarely exceeds a few millimicrofarads.

On the other hand, if the potential difference in the paper condenser exceeds a few hundred volts the paper is punctured electrically and so badly burned that the condenser is ruined; but a good Leyden jar can stand a potential difference 100 times as great.

In addition to the purely mechanical convenience in this use of insulating materials, there are two electrical advantages. One is that any good insulator, such as waxed paper or glass, increases the potential difference which can be applied without breakdown. The other is that the insulator increases the capacitance obtainable with any given values of the area A and separation h, in a way that is both practically useful and theoretically significant.

4. Polarization Charges.—This increase of capacitance is a result of displacements of charges within the molecules of the insulator, under the electrostatic influence of the charges on the plates. If the material really is a good insulator, no electron or charged atom can leave the molecule in which it is; so their displacements are very short. Nevertheless, being systematic in their directions, they create an excess of electrons on one surface of the insulator and a deficiency on the other surface. And, of course, the excess is on the surface facing the metal plate which is deficient in electrons, and vice versa, as shown in fig. 11.

Charges of this kind on the surfaces of an insulator are called polarization charges. The word "polarization" alone as applied to an insulator denotes a vector whose strength in such a case as this equals the surface density of the positive polarization charge and whose direction is from the negative polarization charge (e in fig. 11) to the positive.

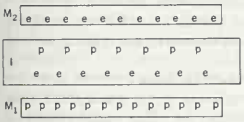


FIG. 11.—POLARIZATION CHARGES

Arrangement of charges in the surfaces of insulator I under the electrostatic influence of conduction charges of metal plates M_1 and M_2; (e) excess of electrons; (p) deficiency of protons

For distinction from these polarization charges the charges on the metal plates are called either conduction charges (since they usually get there by conduction) or free charges (since they are not confined within certain molecules).

Since the polarization charge near each metal plate is of opposite sign from the conduction charge, it tends to offset the conduction charge in all electrostatic effects. These include the potential difference V between the plates: V here is less, with the polarization charges present, than it would be without them but with the same conduction charges.

In any use of a condenser, such as in a CR circuit modelled on fig. 9, it is the conduction charge that is useful. So, keeping Q for the conduction charge, we still define the capacitance C as $\frac{Q}{V}$. With this definition, and V less with the insulator in place than without it, C is greater.

5. The Specific Inductive Capacity.—How much greater C is depends on the material used as an insulator. It does not depend on A or h; and even if the metal plates are neither flat nor parallel, the ratio in which C is increased depends only on what material fills the condenser, provided it fills all the space not filled by metal. So this ratio is given a name and a symbol.

Unfortunately it is given several names and some of them are used in other ways, so they can easily cause confusion. For present purposes, therefore, we shall choose one of them, specific inductive capacity, which is not used for anything else, even though it is not the capacity of anything but is a ratio of capacitances. The word "inductive" got into this name from the term "induction," which was superseded as an electrostatic term by the word "influence."

For a symbol, we choose one of the most widely used ones, the Greek letter κ (kappa).

If now we set up a pair of parallel metal plates so that the space between them may be either left empty or filled with an insulating material (for example, they may be in a glass jar which can be filled with oil), we can use C and C_0 for the capacitances when full and empty. Then, by the definition just given for specific inductive capacity, and the equation written above for the empty condenser,

$$C = \kappa C_0 = \frac{\kappa \epsilon_0 A}{h}$$

If the plates were not parallel, or even if they were but the dimensions of the plates were not large in comparison with h, we should need a more complicated equation; but in any case, κ and ϵ_0 would still appear only in the product $\kappa\epsilon_0$. So this product is often denoted by ϵ without the subscript 0; and it, like κ, has several names.

One of these, which is not used otherwise, is permittivity. Another, which is more widely used but is spoiled by indiscriminate use for either κ or ϵ, is dielectric constant; and insulating materials, when under discussion of this sort, often are called dielectrics.

6. Mechanisms of Dielectric Polarization.—In most solid dielectrics, such as glass or waxed paper, the displacements of charges within the molecules occur in two ways. One is by displacements of electrons, relative to nuclei, within each atom. The other is by displacements of charged atoms, positive in one direction and negative in the other. In glass, for example, positive ions of sodium and calcium are held within meshes of a tangled three-dimensional network of tightly bonded atoms of silicon and oxygen, and this network as a whole has an excess of electrons, or negative charge. If the glass is hot, even if not quite hot enough to be soft, sodium ions can wander through these meshes and enable the glass to conduct electrolytically. In cold glass each positive ion, sodium or calcium, is confined within one mesh; but about three-quarters of the polarization in glass is a result of displacements of these ions, each within its mesh, while the other quarter results from displacements of electrons, relative to nuclei, within the atoms.

In many insulating liquids and gases there is one more kind of displacement of charge, namely orientation of molecules in which charges are asymmetric. The simpler alcohols, for example, show this effect especially well. If the space between the plates of a condenser is filled with ordinary ethyl alcohol, and the alcohol is frozen very hard, the capacitance is only about three times that of the empty condenser, roughly as if it were filled with wax. If now the condenser is allowed to warm up until the alcohol approaches its melting point, and molecules can begin to orient themselves, the capacitance rises to 55 times that of the empty condenser. Further warming reduces the ratio by increasing the energy of the molecular motions and their tendency to disorient the molecules; but at room temperatures it is still about 25.

F. Electrostatic Fields

Under *Transmission of Power*, above, the term "electrostatic field" was introduced to denote a region where electrostatic forces could be found; and it was noted that in the case of a power line this region includes and surrounds the whole system, from the generator through the transmission line to all motors, lamps and other receivers of power. Naturally the electrostatic forces vary greatly from place to place, both in strength and in direction.

The variety has two causes: (1) differences in the charges on which the forces act; and (2) differences in their locations in relation to other charges which exert these forces on them. The first cause is simple: the force acting on any body is proportional

to its charge, so long as all other charges stay as they are. The second is not so simple; therefore it is well to consider it separately. We may ask what variety of forces we would find in different locations if we carried a test charge of fixed strength from place to place.

An ideal test charge would be a single electron, or a Millikan oil droplet with a one-electron charge. Practically, however, the force on it usually would be too small to measure. It would be too small for convenience, even on paper, because it would be a very minute fraction of any practical unit of force. The traditional pith ball, described under *Test Charges*, above, would serve better. But with how strong a charge? Certainly not a whole coulomb; as noted already, that would be as destructive as a bomb.

Even a small test charge makes some trouble in actual tests. It attracts or repels other charges, shifting them around on the bodies which hold them and changing the fields it has to test. For describing the field, therefore, we do not use the force actually observed on any fixed test charge. Instead, we consider what happens if we weaken our test charge more and more, making its strength Q_t approach zero.

Naturally this makes the force acting on it also approach zero. But, calling this force F_t, the ratio F_t/Q_t usually stays away from zero. So we define the electric field strength as the limit approached by this ratio when Q_t approaches zero. We write this definition algebraically as

$$E = \lim_{Q_t \to 0} \frac{F_t}{Q_t}$$

We can think of this approximately, of course, as simply the ratio F_t/Q_t for any Q_t small enough to let us neglect its electrostatic influence on bodies around it.

1. Special Cases.—To illustrate this definition we may apply it to some cases for which we have already specified the forces.

First, for the two point charges Q_1 and Q_2, Coulomb's law was written in this form:

$$F = \frac{Q_1 Q_2}{4\pi\epsilon_0 R^2}$$

So now, making Q_2 the test charge Q_t, this F is F_t, and the ratio F_t/Q_t is

$$E = \frac{Q_1}{4\pi\epsilon_0 R^2}$$

Second, near a long, straight wire, charged with a line density q, the force acting at a point charge Q was

$$F = \frac{qQ}{2\pi\epsilon_0 r}$$

Making this Q serve as Q_t, the field strength here is

$$E = \frac{q}{2\pi\epsilon_0 r}$$

Third, using Millikan's oil droplet as a test charge, with surface densities $+p$ and $-p$ on his plates,

$$F_t = \frac{pQ_t}{\epsilon_0}$$

so here,

$$E = \frac{p}{\epsilon_0}$$

In all cases, as in these, the strength Q_t of the test charge disappears as we get to the equation for E; but this equation always contains the charges which exert the force on the test charge, as in these cases with Q_1, q or p. The unit of E is of course the ratio F_t/Q_t, namely newton/coulomb; and the definition of the volt as a newton-metre/coulomb enables us to equate this unit of E to volt/metre.

2. Lines of Force.—The directions of these forces are, as already noted, just what one would expect. If Q_t is positive, in each case F_t tries to drive the test charge either directly away from the nearest positive charge or directly toward the nearest negative charge. These three cases, of course, are especially simple ones. In more complicated cases, however, the direction of F_t can be estimated roughly by extension of this same idea: it is always such as to drive a positive test charge away (perhaps obliquely) from some other positive charge and/or toward some negative charge.

To specify it somewhat more closely we may use one of Faraday's many good ideas, the concept of what he called "lines of force." In the simplest of these cases, that of the point charge Q_1 and the test charge Q_t, if they are of the same sign F_t tries to drive Q_t directly away from Q_1. If we let Q_t move a little way in this direction, and then stop it and test the field in its new place, we find F_t still in the same direction, as shown in fig. 12.

Repeating this test again and again, we find Q_t moving along a straight line, which leads directly away from Q_1 as far as we care to go. This line is a line of force.

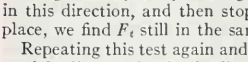


FIG. 12.—TEST CHARGE Q_t IN TWO POSITIONS ON LINE OF FORCE F_t DIRECTED AWAY FROM FIXED CHARGE Q_1

Naturally we can find other such lines by starting Q_t at points not on this one. Wherever we start, outside the body holding the charge Q_1, there is such a line; so these lines form what mathematicians call a family of lines.

In the case of the long, straight wire we can find in this way another family of straight lines of force, all leading perpendicularly away from the wire.

Between the parallel plates the equations for F and E were subject to the proviso that they must not be applied at any point near the edges of the plates. With the same proviso, we find still another family of straight lines of force, all these lines leading perpendicularly from one plate to the other.

But what if this proviso is not obeyed? Here is where lines of force become useful. We can place Q_t at any point we choose and proceed to find a line of force, just as in the case of the point charge Q_1; but if we start near the edge of one of these plates, the line will not be straight. Lines found in this way, however, will still form a mathematical family; that is, at any point we choose in a definable region of space there will be such a line, and any two lines very near together have nearly the same shape. Therefore we may portray them all by drawing typical ones, as in fig. 13.

Perhaps one detail of this procedure should be noted more specifically now. When we let the test charge Q_t move a little way in the direction of the force F_t, and the line of force is curved, we must not let the test charge "fly off on a tangent." Static tests of the forces at various points are what we need for finding the line of force in any experiment that shows curved lines.

3. Some Uses of Lines of Force.—To Faraday, and to many others after him, these lines seemed likely to be much more substantial than a mere mathematical family. The metal plates of fig. 13 attract each other; rubber bands, stretched along these lines, would pull the plates together; so Faraday thought of these lines as having something of the substantial quality of rubber bands.

They could not be exactly like rubber bands, to be sure, which like the curved lines in fig. 13 would try to shrink to straight lines. So Faraday concluded that each line must not only be trying to shorten itself but also trying to "fatten" itself, to push neighbouring lines away, sidewise. If the plates are held apart by nonelectric forces of any sort, the lines must find their proper shapes by a balance between these shortening and "fattening" forces.

Faraday was no mathematician, even though he was a genius in planning and carrying on researches. So the mathematics of these forces was formulated by another genius, James Clerk Maxwell. Maxwell's work on many topics, as on this one, was largely the composition of mathematical sequels to Faraday's notebooks; but Maxwell was no mere Boswell. His mathematics

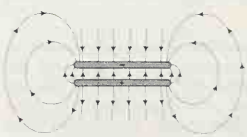


FIG. 13.—LINES AND DIRECTIONS OF FORCE BETWEEN CHARGED CONDENSER PLATES

revealed much that Faraday had never seen—notably the fact that it should be possible to produce electric waves. Radio waves resulted from the work of both Faraday and Maxwell.

For Faraday's shortening and "fattening" forces in his lines, Maxwell found a remarkably simple rule: the force per unit area, technically called stress, in each of these forces was proportional to the square of the field strength E; and at every point the shortening and "fattening" stresses were equal to each other.

The simplicity of these stresses made them very useful in qualitative thinking about what to expect in electrostatic systems. It led, therefore, to great extensions of Faraday's habit of thinking in terms of lines of force. Sir Joseph John Thomson even made some valuable calculations about electric waves (for example, X-rays) as waves running along lines of force, much as waves run along a rope if one end is tied to a post and the other end is shaken sidewise.

Nevertheless the lines of force were not real strings. Neither Maxwell nor any of the other eminent mathematicians who became interested could find any way to specify elastic constants like those of real substances so as to match his stresses. This failure, and others like it, and difficulties arising with relativity led physicists early in the 20th century to abandon the view of lines of force as something substantial, with real stresses; but this change does not impair the utility of the lines. They still give us a mental grasp of the essential features of a field more readily than any less pictorial description. And the real forces the field exerts on real bodies are often calculable most readily through the equations of these ghostly stresses.

4. Lines and Field Strengths.—The mental grasp given us by the lines of force includes not only the directions of the field at all points but also comparisons of its strengths. The latter depend on two rules.

One of these is the familiar rule relating work to potential differences. In the case of fig. 13, since neither plate is carrying any current, any two points in the same plate are at the same potential. But since the plates are oppositely charged, a point in one plate and a point in the other are at different potentials. This difference of potential is the same for all such pairs of points. Therefore if a test charge is driven by the field from one plate to the other, on one such trip along one line of force and on another trip along another line, the field must do the same amount of work on it on both trips. On either line, the motion is in the direction of the force. So this work is the product of distance and force; or, if the force varies in strength from point to point, the work is the product of the distance and the average strength of the force. To make this work the same for all lines of force, between conductors not carrying any current, long lines of force have weaker average field strengths than short lines. Quantitatively, their average field strengths are inversely proportional to their lengths.

By this rule, evidently the very long lines, far out beyond the edges of the plates, have field strengths which are very low in comparison with those on the straight lines well within the space between the plates.

While this rule enables us to compare field strengths on different lines, we need another for comparisons along any one line. This other rule is a corollary of Gauss's theorem, for which another corollary was stated above under *Electrostatic Shielding and the Conservation of Charge*. The corollary needed here is qualitatively as follows: wherever neighbouring lines of force diverge, the field gets weaker as we go along any line in the direction in which they diverge; and where they converge, it gets stronger.

For example, in fig. 13, if we start on the rounded edge of either plate, along one of the roughly semicircular lines shown there, neighbouring lines diverge from this one until we get halfway along this line; then they converge as we approach the edge of the other plate. The strength of the field, therefore, is high near the edge of either plate and low halfway along this line.

Edges of plates generally have especially strong fields near them, just as in this case and for the same reason. A similar convergence of lines of force, even stronger, is found on sharp points such as those of lightning rods.

With two corollaries of Gauss's theorem already stated, we might logically introduce the theorem itself at this point. An even better place for it, however, will be below, under *Maxwell's "Displacement Currents."*

5. Electrostatic Fields With Direct Currents.—Many of the laws of electrostatics illustrated by these cases of charges without currents carry over to the static charges which accompany direct currents; but this is not true of all these laws. One which does carry over, for example, is this corollary of Gauss's theorem, by which the field in empty space near a point or edge of a conductor is stronger than at places on the same lines of force, farther away from the point or edge. On the other hand, a law which does not carry over is the one by which we compared the average values of the field strength on different lines of force.

This comparison, as noted, depended on the equality of the potential differences between pairs of points, each pair being the ends of a line of force. This equality, in turn, depended on the fact that neither plate was carrying any current, so that all points in either plate were at the same potential.

If, instead, we consider a pair of parallel wires, like the wires of a lamp cord or any other two-wire transmission line, carrying currents in opposite directions, Ohm's law requires differences of potential between points in the same wire. These differences upset the equality on which our comparison depends. They do not upset it badly, to be sure, in any short length of lamp cord, but only just enough to illustrate the principle: lines of force between conductors carrying currents do not all have average field strengths inversely proportional to their lengths.

Quantitative comparisons and other details of such cases would require too much space here, but two other points should be noted. One of these is that, while a conductor without any current is equipotential, because there is no electrostatic field within it, this is not true of a conductor with current. On the contrary, as noted very early in this article, electrostatic forces are the driving forces in most conductors carrying currents; *i.e.*, the field strengths within them are not zero.

The other point to be noted is closely related to this one: where there is no current, every line of force meeting the surface of a conductor meets it perpendicularly; with current, it meets the surface obliquely. Usually the obliquity is slight, but it is there. For example, near a U-turn in a conductor (such as an especially simple lamp filament), if the charges happen to be symmetric about the middle of the turn, the lines of force can be roughly as shown in fig. 14.

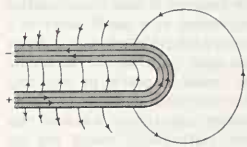


FIG. 14.—LINES OF FORCE NEAR A SYMMETRICALLY CHARGED U-TURN IN A WIRE

Mathematically speaking, the obliquity at the surface is just enough to give the field strength immediately outside the conductor a tangential component equal to the field immediately within the conductor. "Mathematically speaking," however, is avoided in this article as far as is practicable. Therefore a vast array of interesting statements about electrostatics must be left unsaid.

G. Radio Tubes

The little tubes of radio receiving sets illustrate some of these principles of electrostatic fields and of the forces they exert on electrons in free space. Not that radio waves are static; on the contrary, they are propagated with the speed of light, fast enough to go around the earth $7\frac{1}{2}$ times in a second. Paradoxically, it is because their speed is so high that the field in a radio tube can be treated as if it were static. Typical broadcast waves, at 1,000 kc. per second, go 186 mi. during each cycle of changes of the field. Any radio tube is so small in comparison with this distance that there is plenty of time, during every change, for the field to acquire almost exactly the form of a static field.

There is plenty of time also at this frequency for an electron to be driven by this field through the vacuum in the tube, from one electrode to another, before the field changes its strength appreciably. So we may discuss the tube and its electrons in terms

of electrostatics, provided we do not make the frequency too high.

Tubes are of many types and cannot all be considered here—not even, in detail, any of the types that suggest the generic name "tube." Neon signs are of such a type, being really tubular, long and slim. If a neon sign is run on direct current there is a gap in its glow, near its negative end. We call this gap the Faraday dark space.

"Vacuum tubes" have undergone much evolution since Faraday: most of the modern species have lost the ancestral tubular form, but have greatly improved their vacuums. We now use the term "gas-filled tube" for anything like an old vacuum tube; and a tube that Faraday could not come near to making we now call "vacuum-filled"—and we write whole books on how to fill tubes with vacuum, though our best vacuum is only about halfway, in terms of pressure ratios, from Faraday's best to what nature most evidently does not abhor, out between the stars.

For the story of gas-filled tubes *see* ELECTRICITY, CONDUCTION OF: *Conduction in Gases*. Here we shall discuss only a few kinds of tubes, and even these kinds only nonmathematically, to cover the main ideas and their relation to electrostatics.

1. The Diode.—The simplest of these tubes has been given various names. First it was the Fleming valve, because in 1904 John Ambrose Fleming (1849–1945) found it useful in apparatus for receiving messages by wireless telegraphy. Another name for it is "kenotron"; a third is "diode," because a tube of this type contains only two electrodes—but so does a neon sign.

In a diode one of the electrodes is kept hot, so that electrons can evaporate from it into the vacuum, much as water evaporates from leaves into the air around them; and the other electrode is left cool enough to prevent any appreciable evaporation of electrons from it. The hot electrode is called the cathode. The cooler electrode is likewise called the anode or, more descriptively, the plate.

The shapes of these electrodes vary greatly. The plate may be flat, as suggested by this name, or it may be bent so as to surround the cathode. The cathode may be a simple wire like a lamp filament, heated by current passing through it, or it may be a piece of sheet metal, flat or bent, heated by radiation from a filament behind or within it. Whatever shapes the electrodes have, however, the essential principle of any diode is this: electrons can pass through the diode from the cathode to the plate, but they cannot pass in the other direction. Conventional symbols for these electrodes are shown in fig. 15.

Fleming called his diode a valve because it was the electrical equivalent of a check valve in a water pipe, through which water can pass only in one direction. He found it useful in detecting the very weak currents in his receiving apparatus because his best instruments would not respond to a current alternating at any very high frequency. The waves gave forces tending to produce alternating currents; his valves rectified the currents by allowing them to pass only one way.

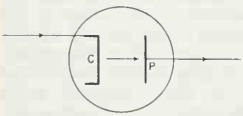


FIG. 15.—CONVENTIONALIZED DIODE WITH CATHODE C AND PLATE P

Scientists later devised better means for detecting such weak currents but continued to find diodes useful for rectifying stronger currents. They are useful, for example, in power packs, systems for supplying direct-current power for tubes of other types and for various other devices. The essential principle of this use of a diode is shown, omitting most other parts of the power pack, in fig. 16.

Without the condenser shown as C in this figure, electrons would flow through the load R only during the half of the alternating-current cycle in which transformer terminal T_1 is charged with excess electrons and T_2 is deficient. With C in place as shown, electrons flow through R—though not through D—all the time. During most of the cycle the electrons flowing through R are draining away some of the excess on the upper plate of the condenser and reducing the deficiency on the lower plate; but during a small part of the cycle, when the potential difference between

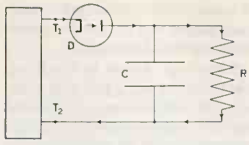


FIG. 16.—SIMPLIFIED POWER PACK: (T_1 AND T_2) TRANSFORMER TERMINALS; (D) DIODE; (C) CONDENSER; (R) RESISTOR

T_1 and T_2 is in the right direction and near its maximum value, the charges in the condenser are replenished by flow through the transformer, the diode and the connecting wires.

Of course the changes of strength of these charges make changes in the potential difference between them, and this is the potential difference that drives the electrons through the load R. These changes are often undesirable, but there are various ways to reduce their magnitude. The most obvious way is to increase the capacitance of the condenser. Other ways, however, involving other condensers, resistors, inductors or more diodes are usually more economical; but details such as these need not concern us here.

2. The Coolidge X-Ray Tube.—A special form of diode that illustrates some important principles very well is an X-ray tube invented in 1913 by William David Coolidge (1873–). In this tube the plate is usually a heavy block of tungsten and is usually called the anode. The cathode is a small spiral of tungsten wire, held in a cup that is open toward the anode.

X-rays are produced wherever electrons moving at very high speed strike matter of any kind. Half the speed of light, for example, does very well; and this speed is found in any electrons that have started from rest and have been driven by an electric field in vacuum through a potential difference of 79,000 v. To produce X-rays in a Coolidge tube, therefore, one needs only to connect its terminals to those of a transformer which gives some such very high voltage and to use some low-voltage source of power (a six-volt automobile battery will do) for heating its cathode, as shown in fig. 17. Here the X-rays are produced at anode A by electrons which have evaporated from cathode C and then been accelerated to a high speed by a very strong electrostatic field between C and A. The heating current for C runs only around the circuit CBJRC; so the wire leading from T_1 to J carries only enough current to replace the electrons lost from the heating circuit through evaporation from C. This current is small in comparison with the heating current, being measured in milliamperes while the heating current is about five amperes; so J may be located at will, anywhere on the circuit CBRC.

In this system practically all the electrons that evaporate from cathode C get out into the strong electrostatic field between it and anode A and are driven to A. To regulate the current they carry, therefore, one needs only to regulate the temperature of the cathode. This is done with rheostat R, if the heating circuit is of the type chosen for illustration here. (Other types, using transformers, differ in this detail but not in any way essential for present purposes.) Even a little experimenting with such a circuit shows that the current which the evaporating electrons can carry increases very much faster than in proportion to the temperature of the cathode.

A little experimenting shows also that this current is rectified as one might expect because of the character of the tube as a diode. All the X-rays, therefore, are produced on the anode, because it is the only thing hit by the high-speed electrons.

They hit hard, however, and most of their energy becomes heat rather than X-rays. Even one milliampere of these electrons, at 79,000 v., carries 79 w.; and the heating current in the cathode

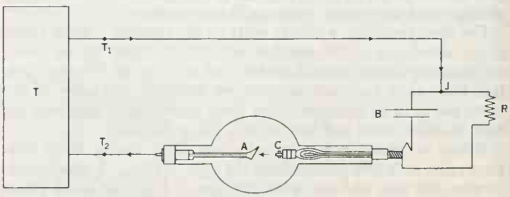


FIG. 17.—CONNECTIONS FOR A COOLIDGE X-RAY TUBE (*See* TEXT)

carries only about 20 w. (roughly four or five amperes at four or five volts). So the anode gets heated very fast if many milliamperes are used at any good X-ray voltage, well up in the tens of thousands of volts.

If the anode gets hot enough for electrons to evaporate freely from it, there is trouble, wrong-way current in the other half of the alternating-current cycle, not only producing X-rays in wrong parts of the tube but also overheating those parts. So it is fortunate that X-ray pictures need only short exposures, and the current can be turned off in time to avoid such trouble. It is fortunate also, in other uses of tubes of this type, that such trouble can be avoided even in long running by rectifying the current with other diodes, as explained under *The Diode* in relation to fig. 16.

3. Space Charges.—Looking again at fig. 16, one might expect offhand that the electrons in the diode would pass through the vacuum there, whenever the current is in the right direction, without requiring any appreciable force to drive them along. But they do not, because the vacuum is not so good when they are there to "spoil" it. They are there because each electron takes a finite time, called its transit time, to go from the cathode to the plate. This is not much time, as we think of time—usually far less than a millionth of a second—nor is it nearly as much time as a charged atom would take, because an electron weighs only $\frac{1}{1,837}$th as much as the lightest of atoms and therefore goes $\sqrt{1,837}$ or nearly 43 times as fast. Nevertheless the transit time is not zero. At any instant, many electrons are on their way from the cathode to the anode. These electrons, out in what should be a vacuum, are called a space charge.

Concrete evidence of their presence there is given by the way they affect the characteristics of the tube. The space charge, being composed of electrons, repels all other electrons. Any electron that is just leaving the cathode feels this repulsion as a backward force. Different electrons leave the cathode at different speeds; so the slower ones are driven back into the cathode by this backward repulsion unless the electrodes are charged strongly enough to overcome it with their forward force.

How strongly the electrodes must be charged to do this, and how many electrons are driven back if they are not so strongly charged, are questions calling for more mathematics than we want to use in this article. For mathematical treatment of these questions *see* ELECTRON TUBE. Exact answers to the questions depend on the sizes and shapes of the electrodes and the space between them, of course; but some general ideas hold in all cases. If any electrons are being driven back, an increase in the potential difference between the electrodes will rescue some of them and therefore will increase the current through the tube. When all of them are rescued in this way, however, there is usually not much more that any reasonable increase of potential difference beyond this point can do to increase the current. (With some cathodes there is a little more, because of effects in the cathode surfaces; but these need not concern us here.) What can be done at a higher potential difference to increase the current is to raise the temperature of the cathode, as we have seen in the case of the Coolidge X-ray tube. These rules are illustrated here in fig. 18.

4. The Triode.—Practical use of the space-charge effect, with easy control of it, is accomplished in a tube with three electrodes, called a triode. Two of these electrodes are a cathode and a plate, like those of a diode. The third is a sort of a wire fence, descriptively called a grid, set up between the cathode and the plate.

Since the space charge also is between the cathode and the plate, any charge of excess electrons on the grid acts very much like more space charge. So anything the space charge can do to electrons near the cathode can be done also by a charge on the grid. Any increase in the number of excess electrons on the grid, therefore, causes a decrease in the current at the plate.

The importance of this rule comes from considerations of power. Power is what we want for running things. Not much power is necessary, to be sure, for such light duty as running the loud-speaker of a radio receiving set; but even this little power is more than the set can receive from the waves that are "on the air." These waves must control the power, but the power they control must come through the wall socket into which the set is plugged.

Power in anything electric is a product of some potential difference or electromotive force with some current—but never with a charge. A static charge on a grid involves no power. A little current is needed, to be sure, for changing this charge in time with the radio waves, but only very little. This grid current is usually small in comparison with the plate current. Moreover, the potential differences driving the grid current are usually small in comparison with the potential difference between the plate and the cathode. Altogether, the power at the grid is very small in comparison with the power controlled by it at the plate.

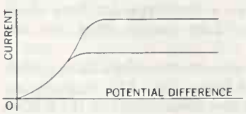


FIG. 18.—DIODE PERFORMANCE CURVES: UPPER CURVE RESULTS FROM INCREASE IN CATHODE TEMPERATURE

This rule is the essential point in the use of radio tubes, but it is not the whole story. It is essential because of the need already noted for controlling the power in the loud-speaker by means of the relatively very weak power that is actually received from the waves. But a look into the back of a radio receiver shows that it is not the whole story: there is more than one tube there, so there must be more to be done than one tube can do.

One thing to be done is to make the loud-speaker speak in tones that we can hear. That is, the changes in its current must occur at what we call audio frequencies, in the hundreds or very few thousands of cycles per second, although the waves that are "on the air" have a radio frequency somewhere near 1,000,000 cycles per second. If the grid of a tube is given an excess of electrons that increases and decreases 1,000,000 times per second, the plate current decreases and increases 1,000,000 times per second. Something more must be done, evidently, to get a current that changes only a few hundred times per second. To make the picture for the triode a little more definite, we may look at a simple circuit involving only one triode, as shown in fig. 19.

A triode used in this way is called an amplifier because the power given to the load L is much greater than the power taken from the signal S. When we say the power is amplified, of course we do not mean to contradict the law of conservation of energy; so we may use this law in locating the source of the amplified power. Fairly obviously this is the B battery $\mathcal{E}_B$. This battery, therefore, must be capable of supplying this power. So it is of rather high voltage, as batteries go, usually of about 90 v.

Such batteries are expensive, though they were used long enough to standardize the letters A, B and C to designate their standard voltages. Nowadays we generally use power packs, as explained in relation to fig. 16.

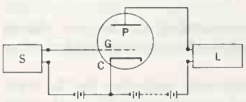


FIG. 19.—SIMPLE TRIODE AMPLIFIER: WEAK A.C. SIGNAL S APPLIED TO GRID G CONTROLS STRONGER LOAD SIGNAL L FLOWING BETWEEN CATHODE C AND PLATE P

5. Other Tubes.—Much more might well be said about triodes, including more about their electrostatic fields and the changes in these fields with changes in the potentials of the three electrodes. But here there are too many other aspects of electricity to be considered. Likewise we must dismiss other tubes with only a few brief notes.

There is the tetrode, for example, a tube containing two grids between the cathode and the plate. The first grid, like that in the triode, controls the current to the plate and is called the control grid; the second, called the screen grid, is essentially an electrostatic shield, to protect the control grid from influence by the plate when its potential changes. Electrostatics covers the ground here, as in the triode, the pentode (a tube with still another grid) and many other tubes, but not all. There is a type of tube called the magnetron, which is very important as a source of power in radar systems. As its name implies, a magnetic field is essential in it—though not without an equally essential electric field, one component of which is static. Another tube in which static and dynamic fields work together is the klystron. Its name means, in Greek, that there are breaking waves in it, as on a

beach; and these are waves of electrons in a stream of fast-moving electrons. There are still other tubes and many varieties of some of the ones already mentioned. To explain many of them would take us far beyond electrostatics; but in most of them, if not in all, electrostatics plays a prominent part.

V. MAGNETIC FIELDS OF DIRECT CURRENTS

Under *Amperes and Volts*, above, the ampere was defined in terms of the force of attraction between long, straight, parallel wires carrying a direct current, and it was said that the law governing this force can serve well as a prototype for laws about electromagnetic forces in general: "Like charges repel, but like currents attract."

Like currents attract, one may say, *because* like charges repel. We saw why in the section *Special Relativity*. This magnetic attraction was required by relativity as a consequence of the motion of the electrons which carry the currents. In general, all magnetic forces can be regarded as requirements imposed by relativity on moving charges. The mathematics of these requirements is usually more complicated than in the case of straight wires, but the main idea is the same. So the contrast between the laws of "like charges" and "like currents" sets the tone for the whole system: the laws of steady magnetic fields are related to those of electrostatic fields by a system of contrasts.

A. Exploring the Magnetic Field

In setting up this system of contrasts, naturally we need something in the magnetic field to be contrasted with the electrostatic field strength E. This vector E, as we may recall, was introduced because electrostatic forces differ among themselves for two reasons: (1) differences in the strengths of the charges on which they act; and (2) differences in the locations of these charges in relation to others. The vector E, varying from place to place, describes the effects of differences in locations.

In the magnetic field, corresponding to the differences in strengths of charges, there are differences in the currents on which magnetic forces act. These differences can be not only in the strengths but also in the directions of these currents. This complication makes it even more necessary in the magnetic field than in the electric to separate the effects of differences in locations and to describe them in terms of a vector analogous to E.

However, we shall not call this vector "magnetic field strength." This name was given long before to another vector, so it is unavailable. Instead we shall use the technical name "flux density," given almost as long before to the vector to be described here, and we shall explain this name later. Corresponding to the symbol E for electrostatic field strength, the conventional symbol for flux density is B.

1. The Exploring Coil.—Corresponding to the test charge used for exploring electrostatic fields, there are several devices available for magnetic fields.

A magnet is the device most obviously suggested by the term "magnetic field." Moreover, a good magnet shows forces stronger than those acting on a wire carrying any practicable steady current in the same field. A good magnet also produces a field stronger, by any test, than any system of wires of the same size. The magnet, therefore, is traditionally the device for exploring the field. Conversely, this is why the field is called "magnetic."

Nevertheless, the magnet does not lead as directly as a current-carrying wire to the principles which are known now to be most fundamental, nor even to the most practical. Antique motors were designed by reference to forces between magnets. Modern motors, which really evolved from the antique through many mutations, look as if they had been designed directly from the law of forces between parallel wires.

To explore a magnetic field with a current-carrying wire, the most obvious device is a piece of flexible wire held loosely by a frame like a violin bow. This device illustrates the law about parallel wires readily, whenever it is held near another wire and parallel to it and both wires carry a strong current.

In other ways also the flexible wire is useful; but a still more useful device is a small coil of wire, called an exploring coil, hung so that it is free to turn and orient itself in obedience to magnetic forces. (For best sensitivity its leads should be very flexible, like the wires of a thin lamp cord, and it should hang by one lead with the other lead made somewhat longer and left loose.)

For ease of identification it is well to have a pair of arrows painted on each face of the coil, showing the directions of flow of its electrons, as in fig. 20.

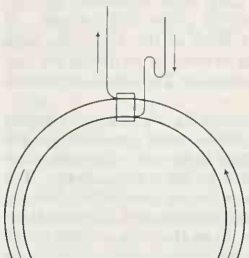


FIG. 20.—EXPLORING COIL FOR STUDY OF MAGNETIC FIELDS (ARROWS INDICATE ELECTRON FLOW)

To illustrate the use of the exploring coil in terms of the parallel-wire law: if this coil is brought near a fixed vertical wire in which electrons are flowing downward, it turns so as to bring its downward-flowing electrons as near as possible to this wire.

Likewise if the exploring coil is hung in the middle of a larger coil, also in a vertical plane, the attraction between like currents and the repulsion between unlike make it turn so as to have its current circulate in the same direction as that in the fixed coil, as shown in fig. 21.

2. The Conventional Direction of the Magnetic Field.—In all cases, as in this one of fig. 21, the exploring coil chooses a plane in which to set itself. The magnetic field, therefore, is characterized by a plane, just as the electric field is characterized by a line, the line of the force F_t which the electric field exerts on the test charge Q_t.

For mathematical purposes a plane is often less convenient than a line. Sometimes, especially in some aspects of relativity, we have to accept this magnetic plane as such. But usually, for convenience, we describe the magnetic field in terms of a line; namely, the line along which the exploring coil sets its axis. (Strictly speaking, the coil must not be restrained by gravity or any other nonmagnetic force if this definition of the line is to apply to all such lines. The simple hanging coil, therefore, tells the whole story only at points where a perfectly free coil would set itself in a vertical plane.)

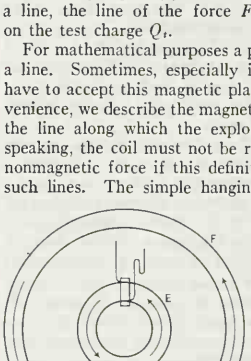


FIG. 21.—EXPLORING COIL E HUNG IN FIXED COIL F AND ORIENTED BY MAGNETIC FORCES

The exploring coil shown in fig. 21 defines not only a magnetic plane but also a direction of circulation in this plane. To specify the field in terms of a line, therefore, we must not only associate a line with this plane but also associate a direction along this line with the direction of circulation in this plane. The rule for choosing the line is natural: the axis of the exploring coil is perpendicular to its plane. But in choosing one direction along this line, rather than the other direction, we have no natural relation as a guide. This choice is made by a purely arbitrary convention: We define the direction of the field as the direction along the axis of the exploring coil which makes its electron flow look counterclockwise.

The conventional direction of the current, being opposite to the real electron flow, is therefore clockwise when we look in the direction of the field. So these two conventional directions are related like the motions of an ordinary right-hand screw, and the rule connecting them often is called the right-hand screw rule.

In the days when motors were designed rather directly from bar magnets, this rule was in constant use, along with others just as arbitrary. Nowadays it does not seem so important. With modern motors the directions of the driving forces are found more readily by the law of parallel wires, without need for any arbitrary conventions.

3. The Direction of the Earth's Field.—Historically this conventional direction of a magnetic field seems to have been based on the mariner's compass, combined perhaps with our ready identification of north by the pole star. "True as the needle to the pole," nevertheless, is a dubious compliment to anyone unless he happens to live on a certain slowly moving line, which crossed London in 1659 and by the 1950s crossed Lake Superior and South Carolina. In London, by the 1950s, the needle pointed about 12° W. of the north pole; halfway to New York city, roughly 25° W.; in Seattle, 23° E. It varies with time and place; nevertheless, throughout most of the habitable lands and navigable seas it points roughly northward. So the direction of an artificial magnetic field was defined first as the direction taken in this field by the end of the compass needle which would point northward in the earth's field. Between this end and the southward end there is no such difference of character as there is between a charge of excess electrons and one of electron-deficient atoms. The subatomic magnets in the needle are essentially like the exploring coil. Looking northward along the needle, they are counterclockwise "electron spins"; southward, clockwise. The conventional direction of the field is as arbitrary when it is defined in terms of the needle as in terms of the exploring coil.

Speaking of the variations in the earth's field, physicists are practically certain that this field is not due to any magnetism like that of hard steel, but almost entirely to electric currents conducted by the molten metal which forms the earth's core. This molten metal flows, slowly but restlessly. So the magnetic map, changing from century to century, somewhat as a weather map changes from day to day, reflects a sort of slow, ponderous, subterranean weather.

4. A Magnetic "Line of Force."—The axis of the exploring coil, as we may recall, defines a line for the magnetic field at its centre. Therefore, recalling the prospect for counterparts to laws of electrostatics, we may expect this magnetic line to develop into something analogous to an electrostatic line of force. The direction chosen for the magnetic field, however, is arbitrary; so we must not expect anything analogous to charges of different kinds at ends of a magnetic line.

In the electrostatic field we defined the direction of the electric vector E as that of the force F, which acts on a positive test charge Q_t. Then we traced the line of force by moving Q_t always in the direction of E. So in the magnetic field, having defined a line for its vector B at any one point, and even an arbitrary direction along this line, we can move this coil always in the direction of B and thereby trace the line.

Starting, for example, from the centre of the large fixed coil of fig. 21, where the axis of the exploring coil coincides with that of the fixed coil, and moving as just specified, we find the exploring coil keeping its axis in this direction, regardless of how far we move along it. The line we trace from the centre of the fixed coil, therefore, is simply the axis of the fixed coil; and it extends as far as we choose to go in either direction.

How can we call this a "line of force"? Of what force? The coil keeps this orientation because of forces acting on its sides, in directions suggested by the law of parallel wires. These are certainly not along this line. Yet we do call it a line of force.

Inquiring further: what if we try the single flexible wire on the frame like a violin bow? Placing this at the centre of the fixed coil, and turning it in various directions, we find various forces; but none of them is along the so-called line of force. Moreover this line of force is the only line through the centre of the fixed coil along which we can lay the flexible wire and find no force at all.

The flexible wire acts like this not only on this particular line of force but also on any other. So we can define the term "magnetic line of force" in either of two ways: (1) it is a line traced by the axis of an exploring coil, free to face just as it pleases; or (2) it is a line along which a wire can lie and feel no force whatever, even when carrying a strong current.

Perhaps the name "line of force" never would have been given to such a line as this if the line had been defined first in terms of the exploring coil or the flexible wire. It was defined first by means of the compass needle. For this, the term "line of force" looked appropriate. Now, therefore, regarding the compass needle as an assemblage of subatomic exploring coils, and dealing in contrasts with electrostatics, we may justify the term "line of force" in the name of contrast.

5. The Field of the Circular Coil.—To trace other lines of force in this field, we may start at other points in the plane of the large coil. At any point in this plane the law of parallel wires gives the orientation of the exploring coil: at any such point it lies in the plane of the large coil; within the large coil its current circulates the same way, just as at its centre; outside the large coil, it circulates the opposite way. So all lines of force cross this plane perpendicularly.

Where they lead from this plane is not so easy to predict exactly; but all the lines through points at the same height as the centre of the coil continue horizontally; so they are easy to trace with the hanging exploring coil. They are found to be as in fig. 22.

Comparing these magnetic lines with the electrostatic lines of a condenser, shown in fig. 13, we continue to find contrasts. Every electrostatic line there runs from the positive plate to the negative, beginning on a proton and ending on an electron; but no magnetic line begins or ends on anything. Moreover, every magnetic line here loops around a side of the coil; but no electrostatic line ever loops all the way around anything.

B. Some Other Magnetic Fields

The contrasting laws stated above hold very generally; and where they do not hold in quite these simple forms they need but little change. For example, if the charges on the plates of fig. 13 are made unequal by a slight increase in the charge of the positive plate, some of its lines of force (from the side not facing the negative plate) have to miss the negative plate. If they can find some other negative charge on which to end, they lead to it. If not, they lead away in all directions to infinity. So the law relating electrostatic lines to charges can be generalized as follows: every electrostatic line of force either begins on a positive charge or ends on a negative charge or does both. Lines of the magnetic vector B never have any such beginnings or ends. They are not always so simple as the ones in fig. 22, but they always contrast with electrostatic lines in this way.

Both this contrast and the other one, between magnetic lines as loops and electrostatic lines as never looping all the way around anything, need a word of caution: static. The electric field in a transformer does loop all the way around something, but this field is not static.

Historically, magnetic lines have behaved very much like electric. They have seemed practically substantial, with real stresses obeying a simple formula like that of the electric stresses. Later, their appearance of substantial reality faded away, leaving some helpful thoughts, but also some perplexity.

Magnetic lines have been even more helpful than electric and

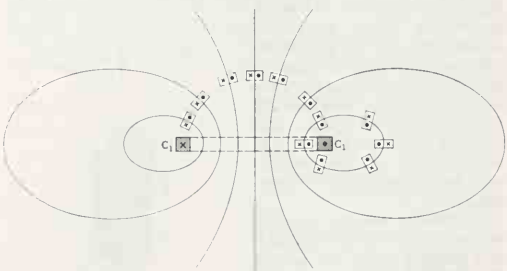


FIG. 22.—CROSS SECTION OF HORIZONTAL COIL C_1, SHOWING MAGNETIC LINES OF FORCE CREATED BY CURRENT IN COIL X IS ELECTRON FLOW TOWARD THE PLANE OF THE PAGE; DOT, FLOW AWAY FROM THE PAGE. RECTANGLES SHOW THAT ATTITUDE IN WHICH HANGING COIL COMES TO REST DEPENDS UPON LOCATION OF ITS CENTRE

have left much more perplexity. How they are still helpful, but no longer perplexing, will appear, therefore, on consideration of some more sets of magnetic lines and of their uses.

1. Lines of Force Around a Straight Wire.—Around a long, straight, current-carrying wire the magnetic lines are simple circles all having the wire as an axis. If the wire is vertical, for example, and the exploring coil is hung near it, the law about like or unlike currents tells the whole story: the side of the coil with current like that in the straight wire gets as near as it can to this wire; so the axis of the exploring coil points along a circle around this wire.

To relate these lines of force to those of the circular coil, we may regard the straight wire as one side of an infinitely big circular coil. If the distance between the cross sections in fig. 22 is increased indefinitely, without increasing their own size, the lines around either of them approach circles centred on it.

For the direction of the field on any of these lines, of course we can use the rule stated above under *The Exploring Coil.* A well-known alternative rule, however, is as follows: If you hold the wire in your left hand, with your thumb pointing the way its electrons are flowing, your fingers point in the direction of its field. In terms of conventional current, this rule takes the same form except that the left hand is replaced with the right. In this form this rule is often called the right-hand rule.

This rule and the one for the exploring coil, setting itself in the field, have been used together ever since the early days of electrical engineering. To find out which way a motor will turn, it has been customary to use this right-hand rule first, finding the direction of the field which acts on the moving parts of the motor. Then the rule for the exploring coil is applied to a coil contained in the moving system, on the basis that this coil must be turning like an exploring coil orienting itself in this field.

This chain of reasoning is confusing, but fortunately it is not needed with modern motors. In one motor, for example, there are moving and stationary wires that are parallel and very close together. Therefore, the simple rule that like currents attract is all that is needed for telling which way the motor turns. The old conventional rules, with their confusing angles, are like sharp right and left turns in an old road, replaced by a modern highway. If they are forgotten, therefore, the loss is not serious; but still we need them occasionally, in odd quests, off the beaten track.

2. The Solenoid.—Many electrical devices contain cylindrical coils called solenoids; the field of a solenoid is shown in fig. 23. Comparing fig. 23 with fig. 22, it is evident that the field of the solenoid is much like that of the circular coil. The chief difference is that the solenoid contains a region in which the lines are almost straight. The longer the solenoid is, the straighter these lines are.

In having such a region the solenoid is the magnetic counterpart of the condenser for which electrostatic lines of force were shown in fig. 13. So we may recall that, throughout its region of practically straight, parallel electrostatic lines, its field strength E is practically uniform. This law can be taken over to the magnetic field of this solenoid, with no change except in terminology.

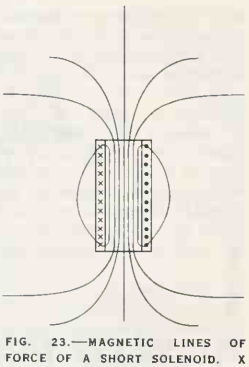


FIG. 23.—MAGNETIC LINES OF FORCE OF A SHORT SOLENOID. X IS ELECTRON FLOW TOWARD THE PLANE OF THE PAGE; DOT, FLOW AWAY FROM PAGE

To give it concrete meaning, we shall have to define a quantitative measure of the flux density B, which is the magnetic counterpart of E; meanwhile, the essential point is that the relation between these fields includes more than just the shapes of their lines.

It includes also some contrasts. One of these, already familiar, is between the plus-to-minus character of the electrostatic lines and the endless-loop character of the magnetic. A second contrast is that the charges which cause the uniform electrostatic field to exist are on its ends, whereas the currents which cause the uniform magnetic field to exist are on its sides. A third contrast, closely related to this second and very important, will appear when we compare the effect of inserting an iron rod into the solenoid with that of inserting a glass plate into the condenser. Conversely, the absence of such things as this plate or rod is a necessary condition for the law about practically straight, parallel lines.

3. Magnetic Lines, Endless but Not Closed.—In most of the simple, idealized magnetic fields traditionally used as bases for designs, the lines are simple closed loops, like the ones shown in figs. 22 and 23 for the circular coil and the solenoid. So it should be noted here that this simplicity is not universal.

The simplest lines of all are the circular loops around an infinitely long, straight wire. But these are the lines of a field, idealized by omission of the earth's magnetic field. Taking the earth's field into account, we cannot expect the exploring coil to align its axis exactly on a circle around the wire. On the contrary, its axis must incline northward from such a circle. If, for example, the wire lies exactly along the lines of the earth's field, and we trace a line of force around it, this steady northward inclination makes the line bring us around to a point north of our starting point.

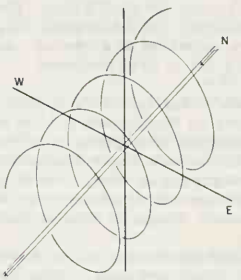


FIG. 24.—HELICAL MAGNETIC LINE OF FORCE AROUND A WIRE IN THE EARTH'S FIELD

Keeping on along the line, we come around again to a point still farther north; and so on, indefinitely. The line is a helix, looping around the wire endlessly but never bringing us to our starting point again, as shown in fig. 24.

In most magnetic fields, as in this one, we may start at any chosen point and follow a line, looping around some current, time after time, without ever returning to the starting point. Closed loops like those of figs. 22 and 23 are the exception rather than the rule.

Nevertheless, they do illustrate the most important rule: that magnetic lines do loop around the currents to which they owe their existence. Moreover, there are so many cases, albeit simplified or idealized, in which each line is a single, closed loop that most practical designs are based correctly on closed loops.

C. Flux Densities

Returning to the closed loops of the solenoid of fig. 23, we may recall the statement that in the region where these lines are practically straight and parallel the flux density is practically uniform. To define a measure of flux density, we use the exploring coil.

In fact we can use it in either of two ways. These look very different at first glance, but they turn out to be exactly equivalent. Therefore we shall describe first the one which is the simpler in principle at this point, but note that the other is easier in practice. This other way is described below in the section *Induced Electromotive Forces.*

Qualitative comparisons of flux densities are the easiest of all. One can hang the exploring coil in various places, in each place at first with no current in it and with its axis not along any magnetic line; and then turn its current on. It starts at once to align its axis with the field; and the stronger the field is, the more vigorously it starts.

By this simple test it is easy to show that the carry-over from the electrostatic field includes more than just the law about practically straight, parallel lines. In the solenoid field, if we follow any line from within the solenoid to points outside it, we find neighbouring lines diverging from this one, as shown in fig. 23. At the same time, the exploring coil becomes less vigorous.

In short, we can take over, word for word, a law stated for electrostatic fields in the section *Lines and Field Strengths*, above: wherever neighbouring lines of force diverge, the field gets weaker as we go along any line in the direction in which they diverge; and where they converge, it gets stronger.

In taking this law over, word for word, however, we are carefully avoiding the words "magnetic field strength," though coming very close to them. They denote a vector for which this law does not always hold. So, even when we use the word "strong" and other adjectives related to it, we must avoid the noun "strength." In place of it, we may use the technical term "flux density."

In this term, the word "flux" is justified more by history than by modern knowledge; but the word "density" is appropriate. Converging lines can be described as becoming more dense. Moreover, when we do define a quantitative measure for flux density, it will be such that the measured quantity does change as we go along a line, exactly in proportion to the density of a properly defined group of neighbouring lines.

1. Measuring the Flux Density.—The last statement above is based on quantitative measurements rather than qualitative judgments of the vigour of an exploring coil. The same principle applies, however; the only change is to make it quantitative.

The most obvious way to do this, perhaps, is to hold the coil at a right angle to the attitude it prefers and to measure the forces it exerts on whatever is holding it, and then to calculate something from these measurements. A little analysis by laws of mechanics shows that this "something" can be defined very simply. It can be reduced, in fact, to the force on a single straight wire.

In the section *A Magnetic "Line of Force,"* above, after description of tests with the single flexible wire on the violin bow, it was stated that a force acts on this wire whenever it carries a current in a magnetic field, except when it lies along a so-called line of force. Conversely, the wire shows the strongest force when it lies across these lines at right angles to them. Now, therefore, the flux density can be defined quantitatively in terms of this strongest force.

As one may expect, wires of different lengths, in a uniform field, show forces differing in proportion to their lengths. So, to obtain a measure of the field, we want to divide any such measured force F by the length l of the wire. Less obviously, perhaps, but still exactly, the force F is proportional to the current in the wire. Calling this i, therefore, the flux density B is defined quantitatively as

$$B = \frac{F}{li}$$

with the specification that the wire must be at right angles to the lines of force.

Measuring F in newtons, as usual in the metre-kilogram-second system, and l in metres and i in amperes, the unit of B is evidently

$$1 \frac{\text{newton}}{\text{metre-ampere}}$$

2. More About the Force on a Single Wire.—Digressing briefly, several steps are needed to complete the laws of this force F.

The laws stated above are for wires directed in only two ways: wires lying along the lines of force have $F = 0$; at right angles, $F = liB$. Oblique angles give F intermediate values. Starting from zero angle (along the lines) and turning the wire so as to increase its angle with the lines, the strength of F changes most rapidly at first, and less and less rapidly as the angle approaches a right angle. Beyond this, F decreases, reaching zero again when the wire again lies along a line of force. These changes are in proportion to the trigonometric sine of the angle between the wire and these lines. Calling this angle λ, the law is

$$F = liB \sin \lambda$$

Considering now the direction of F: it is always perpendicular to the wire, and also to the lines of force, as shown in fig. 25.

These rules give us a definite line of action for F, but they do not say which of the two directions along this line is the direction of F. The choice between these two directions is not a matter for arbitrary definition, like the choice for B on the axis of the exploring coil. When the current is turned on, we can see the wire move. Seeing is believing, therefore, on the direction of F; and its direction must be governed by a real natural law.

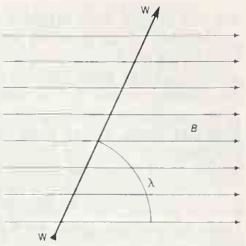


FIG. 25.—WIRE W CROSSING MAGNETIC LINES OF FORCE B AT ANGLE λ. WIRE AND LINES OF FORCE ARE ALL IN THE PLANE OF THIS PAGE; RESULTANT FORCE IS PERPENDICULAR TO THIS PLANE

Not that it is usually found by means of a natural law; it is usually found by a "thumb-and-fingers rule," in which the thumb and two fingers of one hand represent the directions of i, B and F. There are six alternative rules of this sort—and six "misremembered" ways of going astray. Any one of the first six is a convenient short cut, good if used often enough to become firmly fixed in the memory.

However, none of them can tell us anything about the nature of electricity; so for present purposes it is better to look at the real natural law.

To do this, the first step is to repeal the arbitrary rule for B. If we have an exploring coil, we can use it directly, without this rule. Even if we have only someone's specification of a set of lines of force and a direction for B on these lines (as in fig. 25), we can translate this artificial direction of B back into a mental picture of an exploring coil.

In either case, the next step is to recall how the exploring coil in the centre of the big fixed coil of fig. 21 aligned itself with that big coil. Now we may think of the field at our single wire just as if it and this wire were at the centre of these coils. So we may use the law of like currents: if our single wire does not lie along the axis of our exploring coil, its current is somewhat "like" the currents in one side of our exploring coil and somewhat "unlike" the currents in the other side. Therefore, F is directed toward the like-current side, as shown in fig. 26.

3. Lines and Flux Densities.—Returning now to the relations of B to the shapes of its lines, we may recall that the law about diverging and converging lines was taken over, word for word, from a law of the electrostatic field, stated in the section *Lines and Field Strengths*. Along with this law was another law which can be taken over, though not verbatim. The law already taken over enables us to compare fields at different points along any one line of force; the other law permits us to compare different lines.

Before leaving the law about different points on one line, we should note that for the electrostatic field strength E this law holds only for points between which there are no charges on this line. Charges may invalidate this law for E, even in the so-called vacuum in a radio tube, where a space charge of free electrons is an important feature.

On the surfaces of insulators this law for E usually breaks down, because most of these surfaces have polarization charges. But in the magnetic field there is nothing related to the flux density B as charges are related to E; so for B this law holds everywhere.

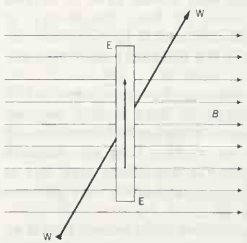


FIG. 26.—INTERACTION OF EARTH'S MAGNETIC FIELD B AND FIELD ABOUT WIRE W. WHEN CURRENT FLOW IN COIL E IS IN DIRECTION SHOWN BY ARROW, RESULTANT FORCE IS DIRECTED TOWARD VIEWER

Turning now to the other law, the law for different lines, its magnetic form is as follows: comparing closed-loop lines of force around the same currents, long loops have weaker average flux densities than short loops. Quantitatively, their average flux densities are inversely proportional to their lengths.

For example, in the field of the circular coil the short loops, which are close to the wires all

the way around, have stronger fields than the long loops.

The field of the solenoid shown in fig. 23 looks, at first sight, hard to reconcile with this law. Every line which leads around all of its wires leads through the region of practically uniform flux density within it. In this region, therefore, all these lines have practically equal values of B, even though some of these lines are much longer than others. Notwithstanding this, the law holds: to give the long lines low averages, B is very weak indeed at all points far outside the solenoid.

A word of caution should be added here about iron. Whenever iron is present, the "currents" mentioned in this law must include its electron spins as well as the conduction currents in wires.

4. Biot and Savart's Law.—The field to which these laws can be applied most easily is that of a long, straight wire. Since its lines are all circles having the wire as an axis, they neither diverge nor converge; so, on any one line, B has the same value all the way around. Its average for this line, therefore, is just this value.

The lengths of these circles of course are proportional to their radii. So their values of B are inversely proportional to their radii. In short, for all points in this field, B varies inversely as the distance from the wire.

This special law was discovered by experimental investigation in 1820 by Jean Baptiste Biot (1774–1862) and Félix Savart (1791–1841).

Long, straight wires, as we may recall, are the basis for the definition of the ampere. As stated under *Coulomb's Law in Coulombs* above, the force exerted by either wire on a piece of the other, of length l, is

$$F_m = -\frac{\mu_0 I_1 I_2 l}{2\pi r}$$

if the ampere is defined through the equation

$$\mu_0 = 4\pi \times 10^{-7} \frac{\text{newton}}{\text{ampere}^2}$$

Each of these wires is at right angles to the lines of force due to the other; so each wire is in a position for measuring the field due to the other through the definition $B = \frac{F}{li}$. Using F_m and I_2 for the F and i of this definition, and dropping the minus sign (since it was only an indicator of the direction of F_m) we find the field due to I_1 to be

$$B_1 = \frac{\mu_0 I_1 I_2 l}{2\pi r l I_2} = \frac{\mu_0 I_1}{2\pi r}$$

I_1 may be the current in any infinitely long, straight wire; so we may now drop the subscript 1 and express Biot and Savart's law as

$$B = \frac{\mu_0 I}{2\pi r}$$

5. Some Other Flux Densities.—This denominator $2\pi r$ is the length of one of these circular lines of force. So this equation says B is inversely proportional to this length, as it must.

For mathematical exactness this wire would have to extend infinitely far in each direction along its line. Practically, it never does; but the equation holds very well if the wire extends each way for a distance many times the distance r for which the equation is used.

This fact permits us to follow the wire beyond either end of its straight part. Here the wire can be bent in any way we choose; yet the law about averages of B on lines of force around it must continue to hold. So we can replace Biot and Savart's $2\pi r$ with the circumference length C of any closed-loop line of force around any single wire, and the equation will then give the average of the values of B on this line.

Bending the wire in any way we choose includes winding it on a spool, to make a coil. If we do this by turning the spool through N complete revolutions, we say the coil has N turns. Such a coil has the same current I all along the wire, so it is like a single wire with current NI. Any closed-loop line of force around all its N turns, therefore, has for its average flux density

$$B_{ave} = \frac{\mu_0 NI}{C}$$

provided no part of its circumference C is in iron or any other magnetic substance.

Even with this proviso, this equation is one of the most important, both in theory and in practice. Deference to it is the explanation for the 4π in the definition of μ_0. Before the adoption of the metre-kilogram-second system this 4π was written explicitly in this equation; now it is hidden in μ_0 and the equation is simplified.

Almost every equation of electromagnetic theory is related to this one. An especially good example is the equation for the flux density well within a long solenoid. If the solenoid's length L is many times its diameter, this flux density is

$$B_{sol} = \text{almost } \frac{\mu_0 NI}{L}$$

6. The Two Laws as Partners.—The word "almost" in this equation illustrates well what happens in practical uses of the two laws of the field. If B could be exactly $\frac{\mu_0 NI}{L}$ at every point within the solenoid, and exactly zero at every point outside it, the law about the average of the values on any line would still be satisfied; but the law about changes where lines diverge or converge would not. This law would not permit any such abrupt change as the one proposed, from full strength to zero at either end of the solenoid. On the contrary, it forbids any reduction to zero, in any distance along any of these lines. When we substitute nonzero values for the proposed zeros, in averaging on any line, we raise the average. To bring the average back so as to satisfy the law about average values, we insert that word "almost."

For an exact formula we should replace "almost" with a mathematical function having fractional values almost, but not quite, up to 1 at points well within the solenoid. For this article, however, we shall leave it as "almost." The essential point here is that neither one of the two laws of the field, alone, could compel the flux density at any one point to have any definite value. To govern the field, these laws must act as partners.

But, one may ask, how can the law about averages on closed-loop lines help to govern fields which don't have any closed-loop lines? Such fields are common, as described in the section *Magnetic Lines, Endless but Not Closed.* So it should be stated here that we can choose our closed loops at will: when we choose a closed-loop line of force, the law tells us the average of all the values of B on that line; when we choose some other loop, it tells us the average of all the values of what we call the component of B along the loop. (This component, at any point, is B times the cosine of the angle made at this point by the loop and a line of force; so if the loop is a line of force, itself, the component is B, itself.)

This generalization of the law makes it include many loops very different from lines of force.

Nevertheless the qualitative idea, to which this law gives quantitative form, is always the same. It is that every current either makes magnetic lines loop around it or at least tries to do so. This idea, when freed from the artificial element in the definition of these lines, is simple: every current is ready to attract any like-current side of an exploring coil.

The quantitative forms of this law and its partner, the law of changes where lines diverge, can be combined by means of integral calculus into a single equation. This equation, however, is much less readily understood than these two laws, and in most cases it is less easy to use in calculations. Consequently we usually do better to reason in terms of the two laws.

This sort of reasoning, with these laws written as a pair of differential equations and the corresponding laws for the electric field as another pair, gave Maxwell his greatest discovery: the discovery that light is a form of electromagnetic field and that it should be possible, in some way, to produce the waves later called "radio."

D. Masses of Electrons and Ions

At the beginning of this article it was said that an electron was only $\frac{1}{1,837}$th as heavy as a hydrogen atom, but no evidence for the value of this fraction was presented there. The evidence involves magnetic fields and the forces they exert on electrons.

The forces that give us this evidence are forces acting on free electrons in vacuum tubes. Similar forces act on charged atoms, or ions, in tubes that we call gas-filled because their vacuums are more obviously imperfect. So the most direct comparison between the electron and the hydrogen atom is given by these forces as they act on electrons and hydrogen ions. Even the most casual comparisons of this sort show that electrons are much lighter than these lightest of all ions.

1. Cathode Rays.—Considering electrons only, for now, we may recall some of the facts stated about them under the heading *Radio Tubes*. One of these tubes, the diode, was described as having a hot electrode called the cathode, from which electrons evaporate into the vacuum, and a cooler electrode, called the plate or anode, which attracts these free electrons whenever it is positively charged. The stronger this charge is, the higher the speed of an electron, by the time it reaches the anode.

Now let us introduce a new feature: a small hole in the middle of the anode. Electrons that happen to be going toward the hole go through it and keep on going. In an ordinary diode they could not go very much farther without hitting the wall of the tube, but of course they can be given more room in tubes of other shapes. One such is the tube that shows the picture in a television receiving set. It is more elaborate, to be sure; but what is used in drawing the picture is a beam of electrons that have come from a hot cathode, gained speed in an electrostatic field, gone through a hole in an anode and kept on going.

Such electrons are like any others in their basic laws of forces. They are repelled by stationary electrons and attracted by stationary protons. They also obey the law of like or unlike currents: if a wire is held parallel to their line of motion and is made to carry current, they are attracted by it if its electrons are flowing in their direction and repelled if its electrons are flowing the other way. Any magnetic field, like that of this wire, exerts a force on these electrons moving in the vacuum just as if they were flowing in a wire. Of course we cannot measure this force in just the same way, because we cannot handle these electrons just as we can handle a wire; but we can, and do, let them hit a screen like a television screen—which is, indeed, a direct descendant of the first screens used in these experiments—and observe how their places of impact on the screen are changed when we subject them to forces, either electrostatic or electromagnetic. Conversely, the television picture is drawn by use of just such forces.

Of course television did not come first historically, nor did electrons from hot cathodes. The electrons used in the discovery experiments came from cathodes that were fairly cool, as a whole. These cathodes were made to release electrons by impacts of ions travelling like atoms in matter at a temperature in the millions of degrees; and the ions got such speeds by being attracted electrostatically to the cathodes.

To have ions in a tube, there must be gas in it. But to let electrons move freely enough for these experiments there must not be much gas—not as much as there was in Faraday's best vacuum. Only after his time was the now-familiar glow of the television screen first seen, by Julius Plücker (1801–68), as a yellowish-green glow on the glass wall of a tube, near its cathode. Not till 1879 was this glow studied well enough to enable one of its leading investigators, Sir William Crookes (1832–1919), to decide that its cause must be the impacts of negatively charged particles of some sort, repelled electrostatically by the cathode.

What sort of particles these were was uncertain until 1897, when Thomson made the first accurate measurements of their deflections in electric and magnetic fields. By calculations from these measurements he discovered how extremely light these particles were. He discovered also that they were all alike, even when taken from cathodes made of different metals. Evidently they were more important in nature than any one chemical element. So Thomson's work on them was the discovery of electrons.

Strangely enough, though we credit him for the discovery, we do not use his name for them, "corpuscles." And when we treat them as he treated his, shooting them at a target as little guided missiles, we give them the old, noncommittal name that they had in the Crookes tube: cathode rays.

2. Positive Rays.—In the Crookes tube the electrons were released from the cathode as described above, because of impacts of positive ions that had been attracted electrostatically by the cathode. If there is a hole in the cathode, therefore, some of these ions go through the hole and on beyond it, much as electrons go through and beyond a hole in an anode.

These positive ions, now called positive rays, are not exactly like cathode rays, because they are much heavier and therefore slower. A positive ray spends much more time than a cathode ray near any gas molecule that it passes; so the positive ray has more chance to use its electrostatic influence on electrons in the molecule and to cause disturbances that result in emission of light. As we look at the tube, therefore, we see quite different lights from the rays of these two kinds: cathode rays are shown by the light they produce at the glass wall of the tube; but positive rays are shown by a streak of light in the gas within the tube.

3. Measuring the Speeds and Masses.—For rays of either kind there are many ways to apply electrostatic and magnetic fields to cause measurable deflections of the rays, from which the speeds and masses of the particles composing them can be calculated. Details of such measurements would be out of place here, but it may be well to say enough to show their main characteristics. For this purpose let us think primarily of cathode rays.

For electrostatic deflection of these rays Thomson's method, which he used in the discovery of the electron, was so good that it is still used, especially in cathode-ray tubes called oscillographs, for studying or testing alternating currents. It consists simply in charging a pair of parallel metal plates, which are placed so that the cathode rays are aimed along a line halfway between them, as shown in fig. 27.

If the plates shown here as D_1 and D_2 are given charges of opposite signs and equal strengths, like those of condenser plates, they produce a downward or upward force on every cathode-ray electron while it is passing between them. So these rays, which would follow the straight dotted line if the plates were not charged, are deflected about as shown by the angled dotted line.

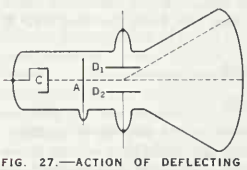


FIG. 27.—ACTION OF DEFLECTING PLATES D_1 AND D_2 ON CATHODE RAYS ORIGINATING AT CATHODE C AND PASSING THROUGH HOLE IN ANODE A

Not bothering with exact calculations here, we may compare the field between the plates D_1 and D_2 to the field between the plates used by Millikan in his measurements of the electronic charge, described under *Measuring a Force on One Electron*, above.

For his plates we had a formula,

$$Fh = QV$$

in which F was the electrostatic force acting on an oil droplet with a charge Q when the potential difference between the plates was V, and h was the height from one of his horizontal plates to the other. So we may apply this formula to these deflecting plates, at least as a first approximation, with Q set equal to the charge of any electron or ion that is passing between these plates.

For more accuracy one would need much more elaborate calculations. One would have to allow for the fact that the electrostatic field strength E is not exactly V/h when h is not small in comparison with the dimensions of the plates. But such complications need not concern us here.

For the magnetic deflection the simplest method with the tube shown in fig. 27 is to hold a long wire, straight and horizontal, in the plane of the figure, above or below the tube, and pass a strong current through it. Still better is a bundle of such wires,

all with current in the same direction, above the tube, and another bundle of wires below the tube with current the other way. This apparatus does very well indeed in a lecture-table demonstration on cathode rays. As noted above, the law of like currents makes any wire attract the cathode rays if its electrons are flowing in the direction of the rays, and repel if flowing the other way. With the wires just described one may calculate the force acting on all the electrons in the beam at any instant, with fairly good accuracy if the deflection is small, by using the equation for parallel wires.

With other magnetic fields the calculation would not be quite so simple; but it is easy to set up a more general formula, based on the equation

$$F = liB \sin \lambda$$

which was written under *Flux Densities*, above, for the force F that acts on a wire of length l if it carries a current i and lies at an angle λ with the lines of force of a field having flux density B. A charge Q, moving in the direction of i with a speed v, is equivalent to this wire if $Qv = li$, for reasons to be discussed below, under the heading *Moving Wires, Moving Magnets and Their Relativity*. So the force on this charge is

$$F = QvB \sin \lambda$$

In any apparatus really well designed for measurements of this sort we can have the rays perpendicular to the lines of the magnetic field, so that $\sin \lambda = 1$.

Having seen ways to deflect the rays, both electrostatically and magnetically, let us now consider briefly how to calculate their mass and speed. The magnetic deflecting force is always perpendicular to the direction of motion; so for it at least we can use the general equation of mechanics for forces perpendicular to motions, namely

$$F = m\frac{v^2}{r}$$

where m is the mass of the particle and r is the radius of the curve on which it is moving. In any experimental setup, this radius r can be calculated from the observed deflection and other dimensions of the system. Equating this F to the magnetic F with $\sin \lambda = 1$, we find

$$m\frac{v^2}{r} = QvB$$

$$mv = QBr$$

In the electrostatic field the force is not necessarily perpendicular to the line of motion; but if the deflection is kept small it need not be far from perpendicular. Treating it as perpendicular, therefore, and using r' for the radius of curvature under the electrostatic force QE, we may write

$$m\frac{v^2}{r'} = QE$$

$$mv^2 = QEr'$$

So here we have two equations, one for mv and the other for mv^2. Dividing one by the other, we find

$$v = \frac{Er'}{Br}$$

Then substituting this value of v into either mv or mv^2, we find

$$m = \frac{QB^2r^2}{Er'}$$

So here is an equation enabling us to calculate m from quantities all of which can be measured and made available for this use of them. It is not the only possible equation for m. Neither is it always the most convenient, because it is restricted to apparatus in which both the fields E and B are perpendicular to the rays.

If they are not perpendicular, the mathematics is not quite so simple. Nevertheless the final formula is essentially of the same type: it contains Q, B and E in the same way; also some pair of geometric factors for the deflected rays, corresponding to the radii of curvature r and r'.

These characteristics of any such formula require positive-ray apparatus to differ greatly from cathode-ray apparatus—in degree, rather than in kind. Fields that work well with cathode rays do not work well with positive rays. For example, the straight wires described above as good for a lecture-table demonstration with cathode rays are useless with positive rays requiring the same E, because their effect on the positive rays is too small to be visible.

For good measurements we want about the same radii of curvature, r and r', with rays of either kind. Yet the fraction written above for m must be thousands of times as great for positive rays as for cathode rays. So we must get greater values for positive rays from the factors B and E: we must give positive rays either much stronger magnetic fields or much weaker electric fields, or both. Good positive-ray apparatus, therefore, must differ greatly from good cathode-ray apparatus, even though the basic principles are the same.

Here we are concerned primarily with basic principles. For details on positive-ray apparatus, *see* Mass Spectroscopy.

4. Masses of Isotopes.—The significance of the small mass found for cathode rays, $\frac{1}{1,837}$th of the mass of a hydrogen atom, has been discussed already; but not much has been said about the masses found for ions. Much could be said: the masses found in this way are not the same for all the ions formed from atoms of any one element by taking electrons out of them; these different masses were the first evidence, and are still in many ways the best evidence, for the existence of isotopes; that is, atoms whose nuclei contain the same number of protons but different numbers of neutrons. These were discussed very briefly near the beginning of this article, under *Neutrons and Atomic Nuclei*. For more information about them, *see* Isotope.

VI. MAGNETIC EFFECTS OF IRON

From the magnetic fields of direct currents in wires without any iron, it is a short but important step to fields with iron. The essential principle is one already stated briefly in the section *The Direction of the Earth's Field;* namely, that the subatomic magnets in iron are essentially like little exploring coils.

The compass needle, aligning itself with the earth's field, illustrates one way in which these subatomic "exploring coils" sometimes orient themselves in magnetic fields. Another way, not involving motion of the iron as a whole, is for each subatomic "exploring coil" to turn around just where it is. One way or the other, unless it is restrained by forces it cannot overcome, it orients itself like an exploring coil of visible size.

A. Amperian Circuits

Historically these subatomic "exploring coils" provided us some surprises. There is nothing in the outward appearance of a compass needle, as a simple steel rod, to suggest a coil. Compasses were used, moreover, for hundreds of years before anyone knew of electric circuits of any kind. But electrostatic forces, as well as forces exerted by magnets on each other, were known to Thales (640–546 B.C.). Rather naturally, therefore, philosophers confused these two types of force.

The confusion was dispelled by William Gilbert (1544–1603). His book *De magnete*, published in 1600, contains a chapter on electrostatics and is the first report of really scientific research on forces of either kind. On a firm basis of observation, and with Elizabethan vigour of expression, he condemns this confusion.

Forces between magnets indeed are quite distinct from electrostatic forces. This is not to deny that a magnet can be charged and then can exert electrostatic forces on other charged bodies. Rather, it is to deny that the magnet is at all different in this respect from a piece of nonmagnetic metal (brass, for example) of the same size and shape. Magnetism has no electrostatic effect.

Certain basic differences, nevertheless, had to be recognized. The forces which one bar magnet exerts on another do not behave like forces between electrostatically charged metal bars, when the magnets are moved from place to place. On the contrary, they do behave like forces between bars made of insulating materials,

each bar having a positive charge near one end and a negative charge, of the same strength, near the other. So the very simplest pair of magnets had to be compared to a system of four charges instead of two.

Reconciling themselves to this complication, physicists gave a name, magnetic pole, to each of these hypothetical analogues of charges.

Before reconciling themselves to it, however, some tried the experiment of breaking a bar magnet between its poles so as to get two bodies, each with a single pole. But they never got even one such body. On each piece, a new pole was created at the break; each piece then had a pair of equal-opposite poles.

Although no magnet ever could have only one pole, it was found easy to make a magnet with three, four or more poles. In every such magnet, however, the poles could always be analyzed into combinations of equal-opposite pairs.

The next great advance came in 1785, when Coulomb discovered his inverse-square law of electrostatics; he also measured forces between magnets and found these forces to be calculable through an inverse-square law for poles.

The simplicity of Coulomb's law led to great mathematical developments from it. Among the greatest was a paper by Siméon Denis Poisson (1781–1840) which set forth many of the most useful laws of electrostatics. Many of these laws could be used as well for magnets, but not all. The limitation on poles to equal-opposite pairs ruled some laws out. Other laws, those about conductors, were ruled out because poles could never be conducted like charges. Magnetically, all bodies seemed to behave like insulators; poles were like polarization charges—or vice versa. Notwithstanding these restrictions, however, the theory of magnetism seemed to be growing well as a younger brother of electrostatics. At least, so it seemed when Poisson published his discoveries in 1812.

1. Oersted and Ampère.—But soon, in 1820, came Hans Christian Oersted (1777–1851) with the discovery that an electric current in a wire could deflect a compass needle. This inspired Ampère to his researches, in which he soon found that electric circuits could exert forces on each other. Thinking this over mathematically, Ampère discovered also that the forces between magnets, in the absence of any wires, could be explained in terms of subatomic electric circuits, without any poles. Here these Amperian circuits provided their first surprise.

Scientists should have reacted to this surprise better than they did—but scientists are human. The philosophical principle of parsimony in hypotheses should have been their guide. Instead, their guide seems to have been habit. Parsimony would have dictated as follows:

1. Whatever we believe about magnets, we must recognize currents in wires as currents.

2. The pole theory of magnets requires us to believe in two types of field producers, poles and currents, whereas Ampère's theory requires only currents.

3. The pole theory requires two very different sets of laws for magnetic fields, one for fields due to poles and the other for fields due to currents, whereas Ampère's theory requires only one set of laws.

4. Therefore, we shall follow Ampère.

But poles were treated as real for nearly another century.

2. Barnett's Discovery.—Ampère's theory, however, was too good to be completely forgotten. Maxwell thought of the possibility that if Ampère were right there might be matter of some kind circulating along with the electricity in the Amperian circuits. If so, these circuits would be miniature gyroscopes and would act like gyroscopes of visible size.

A gyroscope (*q.v.*) does not merely resist all efforts to change the directions of its axis. It resists only on its own terms—that it be allowed to move its axis, slowly but surely, at right angles to the forces it resists. ("Gyro" instruments for blind flying are based largely on ingenious uses of this law.) Applying this law to Amperian circuits, Maxwell thought it might be possible to magnetize a piece of iron merely by spinning it fast. So he tried this experiment; but he did not find any magnetization. However, this did not disprove Ampère's theory; it only proved that magnetism did not depend on circulation of anything heavy.

Truly, an electron is not heavy. It is only $\frac{1}{1,837}$th as heavy as the lightest thing known to Maxwell, a hydrogen atom. So the discovery of electrons threw new light on Maxwell's question; in 1909 it occurred to Samuel Jackson Barnett (1873–) that Maxwell's test should be tried again, this time with very much greater refinement, so much greater, according to Barnett's calculations, that the test would be a very difficult experiment. It was, indeed. Preparation of the apparatus took five years.

Meanwhile the objective of the test came into clearer focus. Atomic nuclei were discovered in 1911 by Ernest Rutherford (1871–1937). In 1913 Henry Gwyn-Jeffreys Moseley (1887–1915) found out how many electrons there were in an atom of any element; for hydrogen, only one electron.

In the same year Niels Bohr (1885–1962) published his famous theory of the hydrogen atom. In this theory its one electron was supposed to circulate around its one-proton nucleus, much as the earth goes around the sun. Bohr could not say just how the 26 electrons in an iron atom should move; but at least his theory suggested strongly that Amperian circuits must be orbits of electrons. This gave Barnett a better basis than ever for his calculation of the amount of magnetization he could expect to find in his spinning iron.

When at last the apparatus was ready, he found only a little over half what he expected. Amperian circuits had given their second surprise. They were not electron orbits.

On what they were, a good suggestion was made in 1921 by Owen Willans Richardson (1879–1959): a spinning motion of an electron, on an axis through its own centre, like the daily rotation of the earth, would make the gyroscopic effect exactly half that of an orbital motion. Barnett's ratio, a little above one-half, implied that the Amperian circuits in iron are mostly "electron spins," with a little help from "electron orbits."

Nevertheless we must not take this picture quite literally. Bohr's orbits were replaced in the late 1920s by structures loosely called "electron clouds." In these an electron is not pictured as having any definite position or orbit. Instead it is given only definite probabilities of being found in certain regions or with certain velocities. The "picture" is blurred; but the mathematics is definite and its confirmations by experiment are far better than any ever given to the orbits. Electron spins are blurred even more, but they too are mathematically improved.

Moreover the new theory, now called quantum mechanics, told which of the 26 electrons in an iron atom can orient their spins like exploring coils. It also told why; why iron, nickel, cobalt and gadolinium have electrons free to do this; and why other elements usually have all their electrons too firmly restrained.

In short, even though we must not take the description of an Amperian circuit as an electron spin quite literally, we can take it as the best picture readily conceivable.

3. Amperian Currents.—Fortunately we do not have to deal with all the individual electron spins in a magnet. On the contrary, taking them all as a group they are usually equivalent to a simple current, entirely in the surface of the iron. To see how, we may look at fig. 28. Conditions at the surface of the bar are illustrated by what a superhuman observer would encounter first as he approaches the bar from either side. As shown, he finds downward flow in the surface at the left side of the bar and upward at the right. Likewise, at the top he would find flow to the left and so on; in short, all over the surface he would find a counterclockwise flow around the bar.

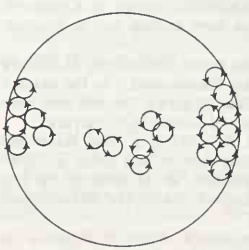


FIG. 28.—EXAMPLES OF ELECTRON SPINS IN A CROSS SECTION OF A MAGNETIZED IRON BAR

Within the bar, on the contrary, a point chosen without reference to the electrons is no more likely to be in a downward flow

than upward; within the bar, there is no predominant flow in any direction.

Of course no one electron here goes all the way around the bar; this equivalent flow is like a relay race; but its magnetic field is the same as if electrons ran the whole distance. This equivalent flow of electrons, or the corresponding conventional current, called the Amperian current, is easy to calculate when we know how strongly the bar is magnetized.

Thanks to Ampère, therefore, we can put all magnetic fields on the same basis: we can ascribe them all to currents.

B. Co-operation of Currents

Amperian currents usually co-operate very simply with conduction currents in wires. An iron bar will have electron spins like those of fig. 28, for example, if it lies within a solenoid whose electron flow is counterclockwise, as shown in fig. 29.

The controlling law here is simply the law of like currents, by which each Amperian circuit aligns itself with the solenoid. Its result, as shown here, is an Amperian current like the conduction current in the solenoid. So the field caused by the Amperian current is like the field caused by the solenoid. Therefore these fields combine, almost everywhere, into a resultant field stronger than either one alone.

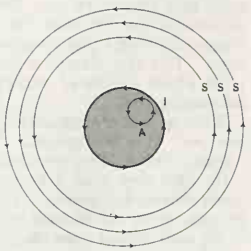


FIG. 29.—ELECTRON SPIN IN IRON BAR I WITHIN COILS OF SOLENOID S. AMPERIAN CIRCUIT A WITHIN BAR ALIGNS WITH CURRENT FLOW IN S

This case is typical. Whenever a mass of iron is magnetized by a field caused by wires lying along its surface, its Amperian current co-operates magnetically with the conduction current in the wires.

1. Contrasting Iron With Glass.—This case helps also to extend the system of contrasts between electrostatic and electromagnetic laws. The electrostatic counterpart of the solenoid is the condenser, and the counterpart of this iron bar is the insulating plate, say of glass, shown in fig. 11. That plate has polarization charges on its surfaces. Each of these is opposite in sign to the conduction charge on the metal plate adjacent to it, but not so strong. So the resultant electrostatic field is weaker almost everywhere than it would be if the glass were removed.

This is typical of the electrostatic effects of insulators. Therefore, in general, polarization charges conflict electrostatically with conduction charges; but Amperian currents co-operate magnetically with conduction currents.

2. Solenoid and Plunger.—A good illustration of this co-operation occurs in a common use of solenoids—opening or closing switches which carry strong currents but must be controlled by weak currents. The starter of an automobile, for example, may be controlled through a push button on the dashboard, yet the push button would be burned out by a current only one-tenth as strong as that which the starter motor must have. The current through the starter motor requires a fairly heavy switch; but this switch can be closed by the use of a push-button current in a solenoid. For this purpose the solenoid must have an iron rod, or plunger, as shown in fig. 30.

In the plunger the electron spins orient themselves, as already described, like exploring coils within the solenoid. So the plunger becomes equivalent to another solenoid, within the real one, with electron flow as shown by the dots and crosses in its surface.

Considering now any point in its surface, a little below the bottom of the solenoid, the electron flow at this point is "like" that of the nearest wires in the solenoid, in the sense of the law of "like currents." This law, therefore, makes the solenoid pull the plunger upward.

This motion closes the starter switch, which is attached to the plunger. At the same time, the motion brings the Amperian current of the plunger, as a whole, more nearly into coincidence with the conduction current in the solenoid. In so doing, it allows these currents to co-operate, magnetically, even better than before the motion.

Co-operation of currents, therefore, is achieved in this case partly by the orientation of the electron spins and partly by the subsequent motion of the plunger. We may regard both the orientation and the subsequent motion, therefore, as manifestations of what may be called, anthropomorphically, a desire on the part of currents to co-operate magnetically.

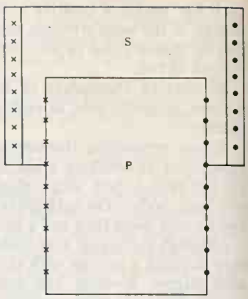


FIG. 30.—ELECTRON FLOW IN SOLENOID S AND PLUNGER P

X's and dots represent electron flow into and out of the plane of this page, respectively; flow shown in the surface of the plunger is the Amperian equivalent of all its electron spins

At the same time, both the orientation and the subsequent motion are consequences of the law of attraction between "like currents." So it may be well to recall in this connection the simplest case under this law: the case of two parallel wires, with lengths many times the distance between them. If these wires are allowed to come together, under their mutual attraction, they produce a magnetic field stronger than either one of them produced when alone. In short, the attraction between "like currents" and the "desire to co-operate magnetically" are just two views of the same law.

Both views are useful, even mathematically. "Like currents" in an exploring coil and other wires gave us the laws of the field. "Desire to co-operate magnetically" takes a mathematical form also. The energy contained in a magnetic field due to two circuits, or to more than two, is mathematically a good over-all estimate of the extent of their co-operation. Also, this energy is useful in calculations of what to expect them to do. They move as if trying to increase it. Calculations in terms of this energy, therefore, express mathematically their "desire to co-operate magnetically."

3. Motors.—Co-operation of currents takes an especially practical form in motors. As noted already, the most important conductors in most modern motors are wires or copper bars lying parallel to the axis of rotation. Some of these are parts of the rotor or rotating system of the motor; others, of the stator or stationary frame around the rotor. So the law of "like currents" (and of "unlike") gives the directions of the forces which drive the rotor around.

Now, therefore, in regard to co-operation: both the rotor wires and the stator wires have Amperian currents alongside them; so these Amperian currents co-operate with the wires and very greatly increase the strengths of the driving forces.

The conductors of a typical motor are embedded in iron. Each conductor, therefore, orients the electron spins in the iron near it, so as to form a co-operating Amperian current in the surface of the slot in which the wire is inlaid. The question arises about how the currents in these wires can be changed, from time to time, so as to make their forces drive the rotor around, consistently in one direction. Actually there is a great variety of ways of accomplishing this. (*See* Motor, Electric.) Here it is sufficient to state only the main point. A large part of the variety depends on the fact that some motors are designed for use only with alternating current, some only with direct current and some with current of either type. One might well expect great differences in appearance in these different motors. So the main point here is that wires parallel to the axis of rotation, either embedded in iron or at least laid along surfaces of iron, are characteristic of good motors of all types. The laws common to them all are the "like-current" law and the law of co-operation of Amperian currents with conduction currents.

4. Relative Strengths.—This co-operation is reminiscent of Aladdin's lamp: conduction currents alongside a piece of iron align its electron spins so as to conjure up Amperian currents; the latter, like the genie of the lamp, are promptly subservient and relatively gigantic.

Without their aid, conduction currents can do but little work. Motors without iron do exist; they turn the pointers in the instruments on our basement walls to tell the power company, each month, how many kilowatt-hours we have used. Telling about kilowatts is easy; producing them as mechanical power is a task for Amperian currents.

Therefore engineers are much concerned with factors affecting the strengths of Amperian currents. So also are physicists, though for more abstract reasons. Such factors are discussed at length in the article MAGNETISM; *see* also *Magnetization of Iron*, below.

VII. INDUCED ELECTROMOTIVE FORCES

The problem of the strength of Amperian currents is essentially one of measuring the flux density B within a piece of iron. For free space, B was defined in terms of forces acting on an exploring coil, to align it with the field. But how can anyone use this definition within a piece of iron? The exploring coil cannot be where iron is. If a hole is cut in the iron, to admit the coil, the hole is free space, not iron.

Fortunately there is another way to use the coil, practicable for fields in iron; and in free space, it is exactly equivalent to the first way of using the coil. Like many other electrical phenomena, it was discovered by Faraday.

Faraday used iron in making this discovery, but for present purposes it seems best to consider free space first. For this we may refer back to the exploring coil in the centre of a large, fixed coil, shown in fig. 21. In this figure both the fixed coil and the exploring coil were assumed to be carrying steady currents, and the exploring coil had aligned itself with the field of the fixed coil.

Now let us hold the exploring coil in this attitude but disconnect it from its sources of current, though keeping the current in the fixed coil constant. Then let us connect the exploring coil to a very sensitive ammeter, so that the exploring coil and ammeter, together, form a closed circuit. Thus far, the ammeter reads zero; nothing has been done yet to make any current run through it.

But now, if the current in the large, fixed coil is suddenly stopped (say, by opening a switch in series with it), the pointer of the ammeter swings suddenly away from the zero mark. It shows current, but only temporarily. It keeps moving only as long as the inertia of the pointer and other moving parts can keep it moving against the spring within the ammeter which restrains them; then the spring brings the pointer back to zero. Meanwhile, however, if we have read the scale where the pointer stopped its upward swing, we have got what we want.

1. Induction and Flux Density.—If we repeat this experiment with different values of the flux density B within the exploring coil (due to different currents in the large coil, or even to different large coils), the scale readings at the ends of the swings are proportional to the flux densities. We have, therefore, the main features of a measure of B.

Other features include the size of the exploring coil, its number of turns, the Ohmic resistance of the exploring coil and ammeter and wires connecting them, and various constants of the ammeter. All these features of the experiment, however, are independent of the value of B at the coil; so for present purposes we need not go into detail.

A matter of terminology, perhaps worth noting here, is that a very sensitive ammeter, when used in this way, is called a ballistic galvanometer. It is ballistic in the sense that the observed quantity is a swing, often called a throw. The name "galvanometer" honours Luigi Galvani (1737–98) for pioneer work on weak currents.

In such an experiment as this, the current in the large, fixed coil cannot actually be stopped instantaneously. When a switch in series with such a coil is opened there is always a flash of blue light in the switch, caused by current through the air. It lasts for only a small fraction of a second, to be sure; but its duration is not zero. Even with electricity, "nothing happens in no time."

The time required for stopping this current can be changed in many ways, sometimes merely by changing the rate of opening of the switch. When this is done, the throw of the galvanometer is practically unaffected, provided only that this time is a very small fraction of the time taken by the galvanometer in its swing. Evidently, the same throw may be produced either by a strong force lasting only a very short time or by a force not so strong but lasting longer.

Putting all the data together, and calculating as needed, one can make reliable deductions as follows: (1) While the flux density in the exploring coil is being reduced from its initial value B to zero, there is an electromotive force in the coil, and we call this an induced electromotive force; (2) if the time for reduction to zero, which we may call t_0, is lengthened, the induced electromotive force is weakened; (3) during this time t_0, the induced electromotive force may vary in an unknown way, but its average value $\mathcal{E}_{ave}$ is inversely proportional to t_0; (4) also, $\mathcal{E}_{ave}$ is directly proportional to the initial flux density B; (5) and to the area A enclosed by the exploring coil; (6) and to the number N of its turns (just as in a battery of N cells).

Combining these proportionalities, they make $\mathcal{E}_{ave}$ proportional to $\frac{NAB}{t_0}$. In this fraction the only factor subject to any arbitrary definition is B. This was defined previously in terms of forces orienting the coil. Now, using this previous definition, it turns out that $\mathcal{E}_{ave}$ is not merely proportional but actually equal to this fraction. Conversely, therefore, an alternative definition for B, exactly equivalent to the definition by forces on wires, is

$$B=\frac{E_{ave}t_0}{NA}$$

2. A Question of Units.—The definition by forces, when reduced to the force F on a single wire of length l, at right angles to the lines of force, with current i, was

$$B=\frac{F}{li}$$

This gave us for a unit of B the combination

$$1\frac{\text{newton}}{\text{metre-ampere}}.$$

Does this new definition give us the same unit of B? The answer is not obvious. Any electromotive force, including $\mathcal{E}_{ave}$, can be measured in volts; t_0, in seconds; A, in square metres; and N is just a pure number. So the unit of B given most directly by this definition is

$$1\frac{\text{volt-second}}{\text{metre}^2}$$

To compare this with the first unit of B, we may recall the definition of a volt as:

$$1\text{ v.}=1\frac{\text{watt}}{\text{ampere}}$$

and the definition,

$$1\text{ w.}=1\frac{\text{newton-metre}}{\text{second}}$$

Putting these together,

$$1\text{ v.}=1\frac{\text{newton-metre}}{\text{ampere-second}}$$

Now inserting this value of the volt in the new definition of the unit of B, we have

$$1\frac{\text{newton-metre-second}}{\text{ampere-second-metre}^2}$$

and simple cancellation reduces this to the first unit of B. These compound units of B, therefore, are strictly equivalent.

Often the product volt-second is called the weber, in honour of Wilhelm Weber (1804–91) for his contributions to the mathematical theory of induction. So the unit of B is often called

$$1\frac{\text{weber}}{\text{metre}^2}$$

The denominator metre^2, incidentally, recalls the law that as

we go along a line of force, B changes in proportion to the density of neighbouring lines; that is, to the number of such lines per square metre. Engineers often have measured B in "lines per square inch"; and for historical reasons their "lines" are defined so that 100,000,000 of them constitute one weber.

3. Lenz's Law.—If the current in the large coil is started again (quickly, by just closing the switch), the galvanometer swings again, the same distance on its scale, but in the opposite direction.

If now one traces the wires to find the directions of electron flow in the exploring coil during these tests, it appears that when the current in the large coil was stopping, the current in the exploring coil ran in the same direction as in the large coil—just as if the exploring coil were consciously trying to keep the field within itself from being lost. The use of the term "induced," like that of "influence" in electrostatics and other such terms, has a basis in psychology.

Likewise when the electrons in the large coil start to flow again, the electrons in the small coil seem to be trying to prevent any change in B from its previous value, zero.

Since these induced currents are short-lived or, as we say, transient, they do not succeed in preventing the changes in B but only in delaying them. So the rule for the directions of all induced currents, discovered by H. F. E. Lenz (1804–65), is that every induced current delays the change of B which is inducing it.

4. Induction With Iron.—With B defined and measured by a fixed coil, we can make tests on iron. Keeping the same apparatus, we have only to insert a long iron rod through the exploring coil, as shown in fig. 31.

Repeating the experiment now, the throw of the galvanometer is much greater than it was without the iron. Often it is a hundredfold greater, or more; sometimes more than a thousandfold. The value of B within the iron rod, due to its Amperian current, is evidently hundreds of times the B given by the large, fixed coil alone.

More exactly, what the galvanometer tells us in such a case is not the value of B at the centre of the exploring coil but the average of all the values of B in the area A enclosed by the coil. To make this average represent the iron well, the exploring coil should be wound tightly around the iron rod.

Still more exactly, we have to recognize another fact, that when we stop the current in the large coil, which has aligned the electron spins in the iron, these spins do not become completely disorganized. On the contrary, a part of their alignment remains in what we call residual magnetism. What part remains depends on several variables. One is the shape of the rod. The longer the rod is, in proportion to its diameter, the better its electron spins become aligned and the better they stay aligned when the magnetizing current in the large coil is stopped. This is reasonable: as they become aligned, the field aligning latecomers is due much more to the Amperian current than to the conduction current; and the longer the rod is, the more Amperian current it has.

Among other significant variables are the composition of the iron or steel and any heat treatment it may have had for hardening or softening. The magnetism of a compass needle is residual magnetism, and the compass needle is hard steel.

Although residual magnetism invalidates a measurement of B in iron made by simply stopping the magnetizing current in the large coil, it does not invalidate a measurement made by reversing this current. Without any iron, a quick reversal of the current gives the galvanometer a throw just twice as great as a mere stoppage. With the iron, a few preliminary reversals are needed to get rid of residual effects of its previous magnetic history. Thereafter, reversals are all alike, so they tell us the value of B within the iron.

A. Magnetization of Iron

As stated above, a long rod gets magnetized and stays magnetized better than a short one of the same diameter, simply because the long rod has more Amperian current than the short one to help align still more electron spins. For the same reason, in any rod it is easier to align the spins with their axes along the rod than across it. This difference affects strongly the behaviour of pieces of soft iron in magnetic fields.

An ordinary soft iron nail illustrates the difference well, when hung by a thread near a coil of wire. If the nail happens to be oblique to its field when the current in the coil is turned on, its electron spins do not immediately align themselves directly with this field. Instead, they point their axes in directions between that of this field and that of the nail. Then, to improve their alignment, they turn the nail, like a compass needle, to point along the lines of force.

Watching this, one might suppose the nail was already magnetized like a compass needle. But, trying it again, starting with the nail turned end for end from its first starting position, one soon finds that the nail does not seem to care which of its ends points in the conventional direction of the field. The nail's behaviour, therefore, does not depend on residual magnetism but rather on the principles just stated.

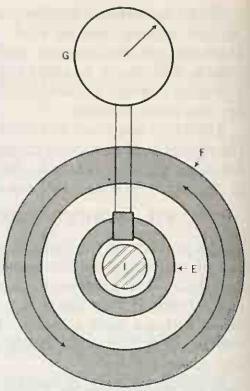


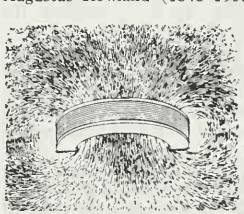
FIG. 31.—EFFECT OF IRON CORE ON LARGE FIXED COIL F; (E) EXPLORING COIL CONNECTED BY WIRES TO GALVANOMETER G; (I) CROSS SECTION OF LONG IRON ROD PERPENDICULAR TO COILS

Even the compass needle is not immune to these principles. If the same pair of experiments is tried with it, and the field of the coil is very strong, the needle's magnetization may be reversed before it has time to turn around. Thereafter, in the earth's field, its end marked "N" will point southward.

1. Mapping a Field With Iron Filings.—Another good illustration of these principles is the behaviour of small, oblong iron filings. If these are sprinkled on a table in a fairly strong, horizontal magnetic field, each filing sets itself like the nail, along a line of force. Then, if the table is tapped so as to make them jump, they collect into strings along these lines. In this way they map the field as shown, for example, in fig. 32.

Magnetization Curves.—This principle about long rods sets the ideal form for a specimen of iron to be tested for magnetization, namely an infinitely long, slim rod. Since no real rod can be infinitely long, the most obvious practical method of applying this ideal is to use a very long, slim iron wire, within a still longer solenoid, and then apply mathematical corrections.

A still better practical form of specimen, devised by Henry Augustus Rowland (1848–1901), is a circular ring wound uniformly with wire so that the wire forms a "ring solenoid," shown diagrammatically in fig. 33.



FROM A PHOTOGRAPH BY BURTON G. STUART

FIG. 32.—IRON FILINGS ON A TABLE PERFORATED AND PATCHED SO AS TO HOLD A VERTICAL CIRCULAR COIL WITH ITS CENTRE ON THE LEVEL OF THE TABLE

In addition to this ring solenoid, or primary winding, which magnetizes the iron ring, there is a secondary winding, corresponding to the exploring coil around the straight rod in fig. 31, for use with a galvanometer in measuring changes of B.

If the ring solenoid consists of many turns, close together, there is practically no field anywhere except in the ring itself. Here all the lines of force are circles, centred on the axis of the ring. These are like the circles around a very long, straight wire. Therefore, by the reasoning given in the section *Biot and Savart's Law*, the part of the flux density which is due directly to the conduction current in the ring solenoid is

$$B_0 = \frac{\mu_0 NI}{2\pi r}$$

where N is its number of turns, I is the current, r is the distance from the axis to the point where B_0 is to be calculated and μ_0 is the so-called permeability of free space.

For the whole flux density we conventionally write

$$B=\frac{\mu NI}{2\pi r}$$

and we call μ the permeability of the iron.

Since, as usual, the Amperian current is like the current in the wires but vastly stronger, B and μ are correspondingly greater than B_0 and μ_0, respectively.

Unlike most quantities specifying electrical properties of materials (such as resistivity), the permeability of any one piece of iron varies greatly when other quantities in its equation vary. Therefore it is customary to plot graphs rather than to rely too much on any given value of μ. A typical graph of this sort is shown in fig. 34.

Starting at o, with the specimen well demagnetized, and increasing B_0 continuously to $B_{0\,max}$, the changes of B are as shown in graph (1), called a magnetization curve. Changing B_0 to $-B_{0\,max}$ and back again, the changes of B are as in graphs (2) and (3), which together form a hysteresis loop. Any given interval on the B axis represents 10,000 times as much change as an equal interval on the B_0 axis.

With such graphs we can compare an Amperian current quantitatively with a conduction current. The conduction current gives us B_0; both together give us B; so the Amperian current gives us $B-B_0$. Noting the 10,000-to-1 ratio of scales, it is evident that $B-B_0$ is scarcely distinguishable from B. The Amperian current is hundreds of times the sum, NI, of the conduction currents I in all the N turns.

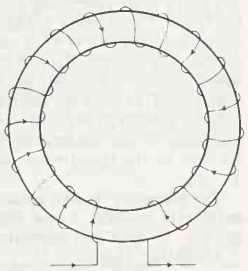


FIG. 33.—SCHEMATIC REPRESENTATION OF RING SOLENOID

There is a limit, however, to what we can expect, even from the genie of the lamp. Iron of any given quality, after all, has only a certain number of orientable electron spins per unit volume. Good iron gives 1,000,000 amp. around a rod one metre long; but to ask for 2,000,000 is too much.

All magnetization curves, therefore, level off more or less like the one in fig. 34, when the orientable spins are all fairly well aligned or, in technical terms, the iron is approaching magnetic saturation. The last possible bit of magnetization usually costs more in magnetizing current than it is worth. Any design, therefore, has to be checked, to make sure it is not asking too high a value of B. A million amperes per metre goes with B = 1.257 v.-sec. per square metre (this 1.257 being 1,000,000 times the $4\pi\times 10^{-7}$ occurring in μ_0).

The graphs for different kinds of iron and other magnetic metals have the same general characteristics, but they differ greatly in proportions and details. For them, therefore, and for their explanation in terms of subatomic physics, *see* the article MAGNETISM.

2. The "Magnetic Field Strength."—The term "field strength" was explicitly tabooed for application to B, at the beginning of the description of magnetic fields, because it is conventionally applied to another vector. In the Rowland ring, this other vector is $\frac{B_0}{\mu_0}$, or

$$H=\frac{NI}{2\pi r}$$

The symbol H seems to have come from "horizontal." In pioneer work on the earth's field, Gauss called its horizontal and vertical components H and V, respectively. Later, physicists seem to have been so much concerned with exact measurements of the horizontal component, for use in standardizing galvanometers of certain types, that its name, H, came to be used for all magnetic fields.

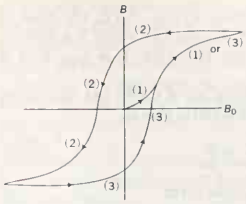


FIG. 34.—HYSTERESIS LOOP IN FLUX DENSITIES IN AN IRON BAR (*See* TEXT)

Naturally this raises the question of the origin of B as the symbol for flux density. On this the present article has only a very tentative suggestion: In the mathematical theory of fields, B often is formulated in terms of a more easily calculable vector called the vector potential and conventionally denoted by A; so perhaps both A and B were adopted merely for lack of any guide but the alphabet. In Maxwell's great book *A Treatise on Electricity and Magnetism*, published in 1873, A and B are introduced without explanation of the choice but in a way strongly suggesting this reason.

In other cases less simple than the Rowland ring, H is not so easy to calculate. In general it can be defined, for any point, as a measure of the field that would exist at that point if the pole theory were correct. This is why, in the days when the pole theory seemed to be correct, the name "magnetic field strength" was given to H.

The unit of H in any case is that of $\frac{NI}{2\pi r}$; so it is conventionally called the ampere-turn per metre.

This compound unit is inherently different from the unit of B, and to most modern physicists the difference is repugnant. They see no corresponding difference in character between the whole flux density, B, in a Rowland ring and the part of it, B_0, due to the wires alone. The use of these different units was determined by engineers, who greatly outnumber physicists.

One thing, nevertheless, can be said for the difference: when a field is calculated from data taken with an exploring coil, the data give directly a number of volt-seconds per square metre (or newtons per metre-ampere); but when the field is calculated from data on the current which maintains it, these data give directly a number of ampere-turns per metre.

Perhaps, altogether, there is some right on each side.

B. Transformers and Alternating Currents

The great development of long-distance power transmission in the 20th century was made possible by the use of very high voltages and correspondingly low currents and low losses of power in resistance. These high voltages and low currents were made practicable by transformers, which require alternating currents.

Most transformers are essentially Rowland rings. They are not circular, to be sure; instead, their iron cores are usually rectangular frames or square-cornered figure-8s, so they look quite different. Also they are modified in a few more important ways, which will be discussed briefly later. But they keep the essential principles of the Rowland ring, especially the principle of induced electromotive force.

In fact they make even more use of this principle: it applies to the primary winding as well as to the secondary, and with transformers both applications are useful.

For a Rowland ring measuring flux densities the most convenient algebraic form for this principle was

$$B=\frac{\mathcal{E}_{ave}\,t_0}{NA}$$

Neither $\mathcal{E}_{ave}$ nor t_0 was known or readily knowable; but neither of them needed to be known because their product $\mathcal{E}_{ave}\,t_0$ was found with the ballistic galvanometer. With transformers, on the contrary, electromotive forces are just what we need to know; but here t_0 is rigidly fixed by the use of alternating currents.

1. Alternating Currents.—The term "60-cycle alternating current" means current which changes through 60 cycles in each second; and the cycle includes all the changes from maximum strength one way, through zero, to maximum strength the other way and back, through zero, to maximum strength the first way.

The flux density changes in similar cycles. So t_0, the time in which it changes from maximum to zero or vice versa, is a quarter of a cycle. With 60-cycle current, therefore, t_0 is $\frac{1}{240}$th of a second.

The term "frequency" means number of cycles per unit time, in this case 60 per second. Algebraically, this is

$$f = \frac{60}{\text{second}} = \frac{1}{4t_0}$$

This gives the equation for $\mathcal{E}_{ave}$ the form

$$\mathcal{E}_{ave} = \frac{NAB_{max}}{t_0} = 4fNAB_{max}$$

with the subscript "max" inserted for special attention to the fact that the B required here is the maximum B.

With this equation we could return now to transformers; but it may be better, while on the subject of alternating currents, to see where this $\mathcal{E}_{ave}$ stands in relation to such values as the customary 110-v. rating of household circuits. This 110 v. is not $\mathcal{E}_{ave}$. Neither is it $\mathcal{E}_{max}$. $\mathcal{E}_{max}$ is useful, both as a basis for mathematical functions specifying $\mathcal{E}$'s at all parts of the cycle and for seeing how strong one has to make insulators. B_{max} likewise is useful in mathematics and for comparison with the practical limits on B described in the section *Magnetization Curves*. But of course no maximum can be truly representative of the cycle as a whole.

$\mathcal{E}_{ave}$ is obviously more representative than $\mathcal{E}_{max}$; yet is not the best representative value. Forces between wires carrying the same current vary during the cycle in proportion to the square of the current. So does the rate of heat production in a lamp filament. The square is more useful than the current itself. Usually, therefore, we average the square. Then we calculate the square root of this average and use it as the most representative value of the current. It is called the effective, or root-mean-square, or for short, rms, current.

Voltages are treated in the same way. One hundred ten volt household wiring has 110 v. as the rms value of the potential difference between the wires.

The rms value of any quantity is always greater than its average. It is greater, even, than such an average as this $\mathcal{E}_{ave}$, taken with all values of $\mathcal{E}$ treated as positive. (Algebraically, with $\mathcal{E}$ negative half the time, its average is zero.) How much any rms value exceeds an "arithmetic average," such as this $\mathcal{E}_{ave}$, depends on how the quantity varies during the cycle.

The smoothest sort of alternation, the ideal toward which alternating-current machines are designed, is sinusoidal variation with time, as shown in fig. 35.

If this ideal is attained the rms value is $\frac{\pi}{\sqrt{8}}$, or practically $\frac{10}{9}$ times the average, and the maximum is $\frac{\pi}{2}$ times the average, or $\sqrt{2}$ times the rms.

The equation

$$\mathcal{E}_{ave} = 4fNAB_{max}$$

does not demand this ideal variation. It demands only that all four quarters of the cycle be alike. For $\mathcal{E}_{rms}$, if $\mathcal{E}$ does vary sinusoidally, the 4 in this equation may be replaced with 4.44; for E_{max}, likewise, by 2π, or 6.28.

2. Step-Up and Step-Down Transformers.—Armed with these principles of alternating currents, we may return to transformers and their places in the transmission of power. If 110,000 v. are needed, for example, on a long transmission line, a transformer at the generating station has to take the power from a generator, for which design principles may have dictated 5,000 v., and step the voltage up to the required 110,000 v. At the other end of the line this highly dangerous potential difference must be removed. Usually it is removed in steps. A transformer in a distributing station steps it down to something of the order of 10,000 v. for local distribution; then others step it down further for industrial and household uses. All this must be done without serious losses of power.

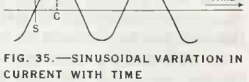


FIG. 35.—SINUSOIDAL VARIATION IN CURRENT WITH TIME

Starting from zero at time S, ordinate is proportional to sine of an angle which increases uniformly with time; starting from zero of time C, ordinate is proportional to cosine of same angle

All this is done according to the law of induced electromotive forces. Most of the factors in the equation

$$\mathcal{E}_{ave} = 4fNAB_{max}$$

are practically the same for both windings. Of course f is exactly the same. Usually the cross section A is the same. Then B_{max} is almost the same, because almost all the lines of force stay within the iron core, all the way around. Even if the cross sections are different, the law about changes of B where lines diverge makes the product AB_{max} practically the same for both windings. But there is nothing to prevent a designer from giving N, the number of turns, whatever different values he may need in order to get any required electromotive forces. In short, the $\mathcal{E}$'s are practically proportional to the N's. Using subscripts 1 and 2 for primary and secondary, respectively, to a good approximation,

$$\frac{\mathcal{E}_2}{\mathcal{E}_1} = \frac{N_2}{N_1}$$

Both these $\mathcal{E}$'s may be average values, or both may be root-mean-squares. Or, even, both may be instantaneous values; that is, values occurring at some arbitrarily chosen instant of time.

These $\mathcal{E}$'s are practically equal to the potential differences on the lines. The primary is like a battery "on charge," receiving energy from the pair of wires connected to it. Electrostatic forces deliver this energy. In the battery "on charge," electrostatic forces overcome adverse chemical forces; that is, the potential difference V exceeds the electromotive force $\mathcal{E}$. But it exceeds $\mathcal{E}$ only by enough to drive the current through the small internal resistance of the battery; so, to a first approximation, $V = \mathcal{E}$. So also, in the transformer primary, to a first approximation, $V_1 = \mathcal{E}_1$.

The secondary is like a battery in use as a source of power, again with opposing forces, but now with $\mathcal{E}$ winning by a small margin. So, to a first approximation, $V_2 = \mathcal{E}_2$.

Altogether, therefore,

$$\frac{V_2}{V_1} = \frac{N_2}{N_1}, \text{ approximately}$$

To step up from 5,000 v. to 110,000 v., for example, one has only to make the ratio N_2/N_1 equal 110/5—plus a little more to allow for losses—and be sure not to ask for too high a value of B_{max} nor use wires too small for the required low resistances, and so on.

Likewise, to step the voltage down again, one has only to make N_2 less than N_1, with similar precautions.

3. Primary and Secondary Currents.—In transforming power to a higher voltage one must expect a lower current. The law of conservation of energy demands this; but this law does not tell how its demands must be met. This is a problem for other laws.

These others include the ones just discussed and also the law by which conduction currents magnetize an iron core. For a Rowland ring with no current in its secondary this law took the form

$$B = \frac{\mu NI}{2\pi r}$$

For a transformer with current in both windings this NI must be changed to include both the primary N_1I_1 and the secondary N_2I_2. With a plus between them or a minus? This question raises one about their directions. The primary is taking power from its line, so I_1 is being driven *against* the electromotive force. The secondary is delivering power to its line, so I_2 is being driven *by* the electromotive force. I_1 and I_2, therefore, circulate around the core in opposite directions. We must use a minus. NI must be replaced with $(N_1I_1 - N_2I_2)$.

In a good transformer under full load this difference $(N_1I_1 - N_2I_2)$ is small in comparison with either N_1I_1 or N_2I_2 alone. This makes N_1I_1 and N_2I_2 approximately equal. So, with the N's practically proportional to the V's, we may write

$$V_2I_2 = V_1I_1, \text{ approximately}$$

One might expect exact equality here if there were no losses of energy. Even without losses, however, there are changes. There is energy in the magnetic field in the core except when this field is zero. It is zero twice in every cycle. Between zeros it rises and sinks back.

To raise and lower it, power must be fed to the core and taken away again. This requires differences, first one way and then the other, between instantaneous values of $\mathcal{E}_1 I_1$ and $\mathcal{E}_2 I_2$. These differences are passed on along with others due to power losses, to the instantaneous values of $V_1 I_1$ and $V_2 I_2$. The law of conservation of energy is satisfied—it always is—but not quite so simply as one might expect.

4. Lagging Currents and Power Factors.—Instantaneous values for power, such as the $V_1 I_1$ and $V_2 I_2$ here, have many uses in the theory of circuits; but in the practical operation of a motor or a lamp we are concerned more often with average power. This average is not related most simply to $V_{ave} I_{ave}$. On the contrary, it is often practically equal to $V_{rms} I_{rms}$. In many cases, nevertheless, it is far below this value. In most of these cases it is low because the current lags, as shown in fig. 36.

Under the heading *Lenz's Law* the currents induced in an exploring coil by a changing magnetic field were described as seeming like conscious efforts on the part of the coil to delay the changes in the field. Lags in simple circuits are essentially delays of this sort. They occur in circuits whose currents create strong fields (large coils or solenoids, for example) and especially in circuits wound around iron. In such circuits the fields created by the currents have no choice but to change whenever the currents change. The only way to delay changes in the fields, therefore, is to delay the currents; that is, to make the currents lag.

The forces doing this to the currents are induced electromotive forces. Conversely, weakness of induced electromotive forces suggests smallness of lag. More specifically, if the electromotive force induced in any circuit is very weak in comparison with the potential difference applied to its terminals to drive its current, the lag is very small. A lamp filament, for example, shows practically no lag.

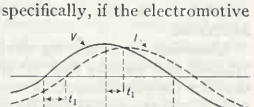


FIG. 36.—POTENTIAL DIFFERENCE V AND LAGGING CURRENT I AS FUNCTIONS OF TIME

T_1, time interval by which current lags, may have any value less than quarter-cycle time t_0

The lamp filament, moreover, recalls one of the main reasons for using root-mean-square values of currents and calling them the "effective" values: its rate of heat production varies during the cycle, in proportion to the square of its current. This rate is the power being taken by the filament. Algebraically, with R for the resistance of the filament, it is I^2R. Its average, therefore, is $I^2_{rms}R$.

Having no appreciable induced electromotive force, the lamp filament responds simply to the potential difference applied to it: at every instant its $I=\frac{V}{R}$. This makes its $I_{rms}=\frac{V_{rms}}{R}$. So its average power, $I^2_{rms}R$, can be called $V_{rms}I_{rms}$.

Some currents have zero lag for reasons less simple than Ohm's law. Motors of certain types can take currents with zero lag. If such a motor and a lamp are connected to the same source of power, the electrostatic field specified by V does just as much work on an electron going through the motor as on one going at the same time through the lamp. For this motor, therefore, or for any circuit with zero lag, the average power is $V_{rms}I_{rms}$.

Many other motors, especially those of the commonest type, called induction motors, take currents with rather large lags. When such a motor and a lamp are connected to the same source of power, the motor's maximum current does not come, like the lamp's, at the time of maximum potential difference. Instead, it comes late, after the potential difference has started to decrease. So an electron going through this motor at the time of its I_{max} receives less energy than one in the lamp's I_{max}. This difference alone makes this motor's power average less than $V_{rms}I_{rms}$.

Moreover there are times when its power is negative. The first of the intervals marked t_1 in fig. 36 is such a time. Here V has just become positive, while I is still negative. During this time the motor is returning energy to the power company. It should be credited with this return. Therefore we want to calculate the average power taken by it in such a way as to allow this credit: we want an algebraic average, one recognizing negative values as negative. This reduces the average still further below $V_{rms}I_{rms}$.

Naturally the ratio of this algebraic average power to $V_{rms}I_{rms}$ is of considerable practical importance. So it is given a name, the power factor.

Whenever a lag is small, the average power is almost up to $V_{rms}I_{rms}$, so the power factor is almost unity. As the lag increases, the power factor decreases. It decreases slowly at first, then more rapidly. If the lag time t_1 could equal the quarter-cycle time t_0, the intervals with negative power would equal those with positive. Then the average power and the power factor would be zero.

Mathematically the power factor takes a simple form if both V and I vary sinusoidally with time, as defined with fig. 35. In this case the power factor is the cosine of an angle, called the angle of lag, which is the fraction $\frac{t_1}{t_0}$ of a right angle.

5. Reactance and Impedance.—For a case of extreme lag, very near a quarter-cycle, one of the best is the primary of a transformer with its secondary not connected to any load circuit.

In this case the potential difference V_1, applied to the primary terminals, has almost nothing to do but to balance $\mathcal{E}_1$.

Taking the term "induced" once more as a licence for animism, one may think of the electrostatic forces in V_1 as putting almost all their efforts into starting and stopping the flow of electrons in the wires; also as never needing to put much effort into maintaining the flow, against forces of resistance, because the electrons never get flowing fast enough to arouse strong forces of resistance.

"Their flow," V_1 might say, "is impeded, not so much because of forces resisting it as because they react against my efforts to start and stop them." Now, therefore, we need two more technical terms, reactance and impedance.

To define these terms exactly, we shall first idealize the case to one of pure reactance. That is, we shall assume the resistance to be strictly zero, and also assume B to be strictly proportional to I_1, instead of being subject to curves like those in fig. 34. In this ideal case (dropping the subscript 1 now) if V_{max} is doubled, so is I_{max}. Conventionally we write for this case,

$$I_{max}=\frac{V_{max}}{X}$$

and we define the term "reactance" to mean this X.

The form for this equation is obviously patterned after Ohm's law; but his law holds for instantaneous values of V and I, at every instant, whereas if I lags, V_{max} and I_{max} do not occur at the same instant.

Nevertheless their ratio X is also the ratio between the rms values or between the arithmetic averages (the algebraic averages being zero). Therefore, to complete the definition:

$$\text{reactance}=X=\frac{V_{max}}{I_{max}}=\frac{V_{rms}}{I_{rms}}=\frac{V_{ave}}{I_{ave}}$$

provided there is no resistance and B is proportional to I.

With resistance not zero, but B still proportional to I (in a coil with no iron, for example) we define

$$\text{impedance}=Z=\frac{V_{max}}{I_{max}}=\frac{V_{rms}}{I_{rms}}=\frac{V_{ave}}{I_{ave}}$$

and we find, either by mathematics or by experiment, that

$$Z=\sqrt{R^2+X^2}$$

This last equation includes the case of pure reactance by making $Z=X$ if $R=0$. Also it includes Ohm's law by making $Z=R$ if $X=0$.

With iron, behaving in the unruly manner of fig. 34, these sim-

ple equations do not hold exactly. The open hysteresis loop of that figure involves a loss of energy. Currents induced within the iron involve further loss. Together these core losses have somewhat the effect of an increase in the resistance. Even allowing for iron in this way, however, the equations are only approximate.

6. Eddy Currents.—These currents induced within the iron would cause serious losses of power if they were allowed to run wild. Being driven by electromotive forces due to changes in B, they circulate around the lines of B. This makes them resemble eddies in water, enough to give them the name "eddy currents."

To reduce the power loss to a tolerable value, most transformer cores are built up out of thin sheets of iron, like a pack of cards, with low-voltage insulation (often only surface films of oxide) between the sheets. From the Latin word *lamina* meaning sheet, such cores are described as laminated. Less classically, the sheets themselves often are called "laminations."

Not only in transformers, but also in many parts of motors and generators, eddy currents would cause serious losses of power if they could. Therefore such parts also are laminated.

7. Self-Inductance.—The reactance of any simple circuit increases with any increase in the frequency of alternation. Thinking of the efforts of the applied potential difference to start and stop the current, one may see that if less time is allowed for each start the current will never get started so well. That is, I_{max} will be less. With reactance X defined as it is, X must therefore be greater.

It is, in fact, proportional to the frequency. Conventionally we write this proportionality as

$$X = 2\pi fL$$

with f for the frequency, and we call the constant L the self-inductance of the circuit.

This self-inductance is closely analogous to the inertia (one aspect of mass) of a moving body. Water in a circuit of pipes shows effects like those of reactance if it is made to flow alternately one way and the other.

Nevertheless there is a significant difference between these cases. The reactive effects of alternation in the flow of water can be calculated correctly from the mass of a molecule of water, the numbers of molecules and the geometry of the pipes. But if a similar calculation is made for the electrons in a coil of wire (even without any iron core) the result is of a much lower order of magnitude than the reactive effects actually observed. Self-inductance is not a result of the masses of the electrons as individuals.

Instead, it is a result of mutual action between them. Whenever their speed of flow is changing, each electron is affected by induction, tending as usual to delay the change of speed. The force acting on it, therefore, is like the effect of its own individual inertia; but being due to such a vast number of other electrons, it is much stronger. Self-inductance, therefore, is mutual inertia.

8. Resonance and Tuning.—The comparison made between self-inductance and ordinary mechanical inertia recalls another comparison between electrical and mechanical things, made much earlier in this article under the heading *Condensers* (*Capacitors*). There a condenser was compared to a spring, and fig. 9 showed a circuit in which a condenser would act like the spring of a wound-up toy; it would drive electrons through a resistor, vigorously at first, but then with less and less vigour. The electrostatic forces, getting weaker as the charges were drained off the condenser plates, were therefore like the mechanical forces of the spring.

It was noted also that springs not only could drive things that simply slow down without changes of direction of motion, but also could make things vibrate, changing their directions of motion with definite frequencies of vibration. So it was said that a condenser could do the same for electrons—though for electrons we use the word "oscillate," instead of "vibrate"—provided we have a circuit of the right sort. That is, instead of a resistor, which corresponds to something with friction, we must have a circuit with something like inertia. Well, here it is: self-inductance is like inertia.

If any mechanical thing vibrates with a frequency of a few hundred cycles per second we can hear it and estimate its frequency by the tone we hear. To show that an electric circuit of the right kind can oscillate, therefore, we may include a radio loud-speaker in the circuit and listen for its tone. A good set of apparatus for this demonstration is shown in fig. 37.

To charge the condenser of the system shown here one may throw the knife of the switch D to the left, closing the circuit through the battery. Then to cause oscillations one may throw the knife to the right, first opening the battery circuit and later closing the circuit through the loud-speaker and the coil. The loud-speaker speaks then with a tone that is recognizable, though the sound is of very short duration, much more like a note of a xylophone than that of a piano.

The tone is changed obviously if the condenser is replaced with another condenser having any very different capacitance, or if the coil is replaced with another coil having any obviously different radius or different number of turns, and therefore a different self-inductance. Moreover the changes are in the right directions: any increase in the number of turns, for example, increases the self-inductance; so this change is like an increase in the mass of a mechanical vibrator; that is, it lowers the frequency.

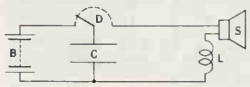


FIG. 37.—CIRCUIT FOR DEMONSTRATING FREE OSCILLATIONS BY USE OF LOUD-SPEAKER S, CONDENSER C, INDUCTANCE L, BATTERY B AND DOUBLE-POLE SWITCH D

The short duration of the sound shows another analogy with mechanics: the short duration of the xylophone note is due to internal friction, which is much stronger in wood than in piano wire; the electrical analogue of friction is resistance; so the resistance of the coil is what makes this system sound like the xylophone.

Altogether an oscillating circuit is so much like a mechanical vibrator that we may look for still further analogies. So we may consider the phenomenon of resonance, as shown, for example, by a piano. If the dampers in the piano are all held up off the strings and a single note is sounded loudly for a second or two on some other instrument, the strings that are in tune with this note pick up enough energy to keep on sounding audibly after that other instrument has stopped; but the other strings, though just as free to vibrate, do not do this. This experiment works well with a piano but not with a xylophone. For good resonance we need not only good tuning but also absence of serious friction.

Absence of serious resistance is not one of the characteristics of the circuit shown in fig. 37, so this circuit is not a very good resonator; but it does resonate well enough to show that there is such a phenomenon as electrical resonance.

To make it show this, one may leave its switch closed to the right and give it power at an adjustable frequency by inducing a suitable electromotive force in its coil. This electromotive force should be weak but persistent. It may well come from a small fraction of the magnetic flux of another coil, which should not contain any iron, so that it will not make any noise; and the current in this other coil should be driven persistently, regardless of any response in the circuit being tested, by a vacuum-tube system called a variable-frequency oscillator, which can be tuned by simply turning a knob with a pointer on a dial. As the pointer passes through a certain short range, which includes the frequency that was obtained with the battery and switch, the loud-speaker gives a little howl. It is not silent at other frequencies; but it is enough louder in this short region to tell us that the circuit really does pick up energy especially well there; in other words, that it really does resonate.

If we could hear sounds of frequencies ten times as high as any that we can hear, they would be much better for this experiment. That is, it would be easy to build a circuit for them with more contrast in volume between the sound at the resonance frequency, where it is loudest, and the sound at a frequency 10% higher or lower. With a resonance frequency in the radio region, it would be very easy to make the contrast still better.

This rule makes resonance very important in the selection of a radio station to be received and the exclusion of other stations. The process of selection is not quite so simple as dealing with a single circuit, to be sure; but the phenomena shown by a single circuit are at the root of it.

What is done to any single radio-frequency circuit is usually to change its capacitance, in a way described in the section *Condensers (Capacitors)*. Such a change is so much like tuning a musical instrument that it is quite reasonable to speak of tuning the circuit, and therefore not wholly unreasonable to speak of tuning the radio as if it were a single circuit.

C. Moving Wires, Moving Magnets and Their Relativity

In old books—old in the history of electrical engineering—transformers were sometimes called "static transformers." This name distinguished them from what we now call motor-generator sets, motors driving generators, which could be used within limits for transforming direct-current power. Transformers are static in comparison with these machines. Also they seemed static in another sense (more significant then than now)—that as the flux density in the core gets stronger or weaker the geometric pattern of its lines of force does not move.

Lines of force have been very useful—almost too useful, in some ways. They appeared likely to be substantial things in electric fields as well as in magnetic. They explained forces certainly acting on visible and substantial things very well, in terms of the stresses imagined for the lines by Faraday and calculated for them by Maxwell. Any experimenter, thinking of the lines in this way, naturally would think of magnetic lines as moving when he moved a magnet. When this motion was found to be related simply to some really observed electromotive forces, the lines seemed more substantial than ever.

The idea of moving lines was extended, therefore, to cover many cases where no magnet moved. "Rotating-field" motors got their name from parallel lines of force which seemed to rotate as the lines of print on this page do if the book is turned while lying on the table. No magnet moves, to move these lines. Instead, several coils, placed symmetrically around the region where they seem to rotate, take turns in creating the field. To try to trace their motion out among these coils is to risk madness. But that is not where electromotive forces have to be calculated with them; nature laughs at mathematical difficulties; and if the joke is on man, it is a small price for such good motors.

Such cases strongly encouraged the idea of trying to explain all induced electromotive forces in terms of motions. Sometimes magnetic lines of force would seem to sweep across wires. Sometimes moving wires were said to "cut the lines of force"—never wounding them incurably. Either way, what seemed to count was relative motion of wires and lines.

The "static transformer" might seem to stand steadfastly against the idea that all induction was of this sort. In some books, however, even it was ingeniously forced into submission: whenever its flux density was increasing, new lines of force were being created in the wires and being crowded out from the wires by still newer lines; whenever the flux density was decreasing, lines were shrinking back and vanishing. The literature of electricity was filled with weirdly moving lines.

Sometimes it was easy to tell how fast the lines should move and from this speed to calculate electromotive forces correctly. But in many other cases, especially where some pieces of iron moved and others did not, one might easily misjudge the speed. Even then, however, one could calculate the electromotive forces correctly by using other equations, chiefly the work of Maxwell, Lorentz and Einstein, which did not mention motion of lines; and from these electromotive forces one could work backward in the moving-line equations and find the speeds at which the lines must move, to explain the electromotive forces.

In what sense is such an afterthought an explanation? This question is especially pointed if the afterthought speed looks like a result of biased judgment. A case in point is a very useful type of generator called the inductor alternator. In this a stationary coil carrying direct current maintains the lines of force. A little more than half the length of each line is in a rotating piece of iron; the rest is in stationary iron. The afterthought speed is the speed of the fastest-moving part of the rotating iron, regardless of all that moves slower or does not move at all.

Is this speed an explanation? Fortunately—thanks chiefly to Maxwell, Lorentz and Einstein, but also to such able experimenters as Harold Albert Wilson (1874–), Earle Hesse Kennard (1885–), John Zeleny (1872–1951), Leigh Page (1884–1952) and others who confirmed their equations—we need not use any speed at all for lines of force. On the contrary: if we have calculated any electromotive force, using the equations which always hold and which contain no speed of motion for the lines, then any subsequent "correction" for motion of the lines is pure error.

1. The "Flip Coil."—Let us return now to the exploring coil in the middle of a larger coil, as in fig. 21, and connected only to a galvanometer. In this position it drives a transient current through the galvanometer when the current in the large coil is stopped, as if trying to prevent or at least delay the loss of the field within itself.

Stopping that current is not the only way to deprive the exploring coil of this field. Another way is to take the exploring coil to some place where there is no field. If this is done while it is kept connected to the galvanometer through flexible leads, such as a piece of lamp cord, again it drives a transient current through the galvanometer. Moreover, if the motion is very quick, so that it is completed before the pointer of the galvanometer has time for any considerable fraction of its swing, the swing is just the same as when the current in the large coil is stopped.

Practical measurements of fields, therefore, may be made in either way. Stopping the field-producing current is often the more convenient way. But when there is iron near where the field is to be measured, merely stopping the current leaves residual magnetism; and reversing the current may be less convenient than quickly removing the exploring coil. Quickness is essential, as already noted. So an exploring coil for this use, with a good handle, is often colloquially called a flip coil.

2. The Lorentz Force.—The electromotive force in a flip coil is closely related to the mechanical forces exerted by the field when the coil carries current. When the coil moves, it takes its electrons along with it. This is one way to move electrons; making them flow as a current is another.

The latter way is known to cause a force, so why shouldn't the former? And shouldn't the same law cover both? Yes. But moving the wire moves its protons as well as its electrons. If there is no current, these motions are alike. The forces, then, are equal in strength and opposite in direction. They balance, mechanically. So long as there is no current, there is no resultant mechanical force on the wire. There is only the electromotive force, the effort to drive its electrons one way and its protons the other way.

Here we should pause to recall that the word "force" is badly misused in the term "electromotive force." Qualitatively, an electromotive force is an effort of the sort just described. Quantitatively, its unit, the volt, is not a force unit like the pound or the newton. So we must be alert for the distinction when using any mathematics on relations between electromotive forces and true mechanical forces.

The basic element common to these phenomena is the force which acts on any one electron because of its motion in the magnetic field. This force, called the Lorentz force, is just what one might expect in view of the force on a current-carrying wire. Its rules about directions are all just those of that force. The only new feature about it is that the equation $F=lIB$, for the force on a wire at right angles with the lines of force, is replaced with

$$f=evB$$

Here e is the electron's charge, 1.60206×10^{-19} coulomb, and v is its speed, and it is crossing the lines of force at right angles.

At any other angle we should have to multiply evB by the sine of the angle. But the simple right-angle equation will cover present needs.

To see why ev is the correct replacement for lI, we may test

it as follows. If *ev* is right, the Lorentz forces on all the electrons flowing in the wire should combine to make the force *F*. The number of these electrons is of course proportional to the volume of the wire, which is *al* if *a* is its cross-section area. Calling this number *nal*, therefore, *F* should equal *nalf*, provided the speed used in calculating *f* is the speed of flow in the current *I*. Calling this speed v_1, the force *nalf* is $nalev_1B$. Regrouping the factors here, nav_1 is the number of electrons per unit time flowing through the cross section *a*. So nav_1e is the quantity of electricity carried through, per unit time; that is, the current *I*. The whole product, therefore, is indeed *lIB*, as it should be.

The Lorentz force passes not only this test but also tests on electrons and protons in vacuum tubes. These tests were described above under the heading *Masses of Electrons and Ions* as tests of deflections of rays in magnetic fields: of electrons as cathode rays and of protons as positive rays in hydrogen. The equation used for the magnetic deflecting force in those tests was just the Lorentz force equation. So this equation holds, whether the electrons and protons are in a wire, where they are subject to subatomic forces as well as to Lorentz forces, or "free" out in the vacuum, where the Lorentz forces can act on them without help or hindrance from anything else.

With confidence in the reality of the Lorentz forces, therefore, we may now examine more closely their manner of causing an electromotive force in a moving wire. To keep the simple equation $f=evB$, the direction of motion must be at right angles with the lines of force. Then to make the electromotive force as strong as possible, the wire should be aligned with the Lorentz forces, as shown in fig. 38.

Since electron e of this figure, with its speed *v*, is equivalent to a very short bit of current-carrying wire, the law of like currents directs its Lorentz force f_e toward the side of coil C which has a current like it. The electromotive force induced in wire W, therefore, tries to drive its electrons downward.

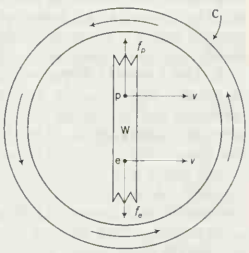


FIG. 38.—LORENTZ FORCES IN WIRE W, MOVING AT VELOCITY *v* ACROSS FIELD SET UP BY COIL C (*See* TEXT)

The general rule for directions of all such electromotive forces is simply what is in this diagram: represent the field by a coil, as here, and use the law of like currents.

To find the strength of any such electromotive force, one has to go beyond this qualitative view of it. One has to use its quantitative definition as a fraction, power divided by current. This implies letting the wire carry a current. The electrons in the wire must not only move to the right, at the speed *v* of the wire, but at the same time flow downward along the wire. The speed v_1 of this downward flow is related to the current *I* by the equation already found, namely $I = nav_1e$. The power given to this flow is the product of its speed v_1 and its driving force. This force is of course the force on all the electrons put together. For a wire of length *l*, with *nal* electrons, this makes the power

$$P = v_1nalevB$$

This product contains all the factors in the value just written for *I*, and some others. That is, this power is

$$P = IlvB$$

Dividing *P* by the current, as required by the definition of electromotive force, the result is

$$\mathcal{E} = lvB$$

The power *P* contains, likewise, all the factors in the mechanical force,

$$F = lIB$$

Altogether, therefore, these equations make

$$\mathcal{E}I = Fv$$

that is, this power *P* can be called either electric power $\mathcal{E}I$ or mechanical power *Fv*.

To have this power given to the current, the electrons had to flow downward. A current like this in coil C is found on its left side, so *F* acts to the left. This makes it try to slow down the motion of the wire. *Fv*, therefore, is mechanical power taken *from* the wire when the electric power $\mathcal{E}I$ is given *to* it. Motion of a wire across a magnetic field changes mechanical power into electrical power (or vice versa if *I* is reversed) with neither loss nor gain. It never transfers any energy to or from the field.

Not to transfer energy means not to do any work. The only way a force can act on a moving body without doing work is to act in a direction perpendicular to the motion. This is precisely how the Lorentz force does act on each individual proton or electron. Therefore the law stated above results directly from the nature of the Lorentz force.

From the electromotive force in a straight wire to that in the flip coil is but a short step. To drive electrons around the coil, the wires on one side of the coil must have stronger electromotive forces than those on the other side. They do, if one side is in a stronger field than the other side. It is, while the coil is being flipped out of the field.

To discuss the flip coil quantitatively is another story, which can be told only with much mathematics. But the point is: if the coil is taken to a place of zero field in a time t_0, first by one path and again by another, the Lorentz forces give the same average electromotive force $\mathcal{E}_{ave}$, regardless of the paths; moreover this $\mathcal{E}_{ave}$ is the same as that given by stopping the field-producing current in the time t_0; namely,

$$\mathcal{E}_{ave} = \frac{NAB}{t_0}$$

3. Pros and Cons of Moving Lines.—It is not often that two causes as different as flipping the small coil and stopping the current in the large coil produce identical effects. One cause, flipping the coil, contains the obvious element of motion through the field. This element is emphasized by the connection between this electromotive force and the mechanical force due to flow of electrons. If motion, relative to the field, is the essential element in flipping the small coil, can it be found also in stopping the current in the large coil? Does stopping this current move its lines of force?

To ask these questions is to reopen a *res judicata;* the case was settled by the decision against moving lines recorded above. This case is reopened here, not with any intent to appeal it, but to show how inevitably physicists and engineers were lured by the will-o'-the-wisp of moving lines and to show where it led them.

With this intent we describe another test, as follows: leaving the exploring coil at rest, and maintaining a steady current in the large coil, the experimenter deprives the exploring coil of its field by quickly moving the large coil away. Again the exploring coil delays the loss of its field by carrying a transient current. And again its electromotive force obeys the same equation. Could anyone doubt that in this test the large coil carries its lines of force away with it? Not without good reason, at least, to overcome such an apparently conclusive test. If all induced electromotive forces could be calculated from moving lines as well as this one we should probably have continued to accept them.

One case where they do not work so well was described briefly above. This case, on the inductor alternator, however, was not chosen as the most conclusive against moving lines. On the contrary, it was chosen because it has been "explained" especially often in terms of moving lines. It has been "explained," that is, in terms of lines moving at a speed so readily identified as to be accepted by anyone wanting simplicity above all else in explanations; nevertheless, a speed that on second thought is unreasonable.

A more conclusive case is that of the transformer. Algebraically this is like the exploring coil with stopping of the current in the large coil. Naturally, therefore, this simpler case is treated first, when using moving lines. It goes well. As the current in the large coil decreases, its lines are assumed to shrink into its wires and vanish there. Each line, as its moves, takes the places of the lines

which have gone before. Thus all the lines which were initially threaded through the small coil sweep across its wires and contribute their shares to its electromotive force. And of course, what works for these two coils should cover the primary and secondary of the transformer.

But just how? Let the transformer core be a circular ring. Let the primary be wound on this ring first, with its turns evenly distributed around the ring. All its lines of force are circles, centred on the axis of the ring. Most of them lie in the iron; all the rest, when there is no current in the secondary, lie in the primary winding itself. No lines due to primary current lie anywhere else. And the secondary wires must be somewhere else if they are wound after the primary. When the primary current changes, therefore, its lines cannot sweep through the secondary wires. Just how, now, can moving lines explain the electromotive force in the secondary?

They can clutch desperately at the case where they had success: two circular coils. Each turn of the primary winding can be given the right to act like the large coil, creating its own mathematical family of lines and moving them. With hundreds of turns, then, there must be hundreds of lines passing through any one point in the surrounding space, crisscrossing in all directions and moving in still other directions. Then too, the Amperian currents must have their own families of lines and must move them, each in its proper direction. This does the trick. But what does it do to our motive for moving the lines: simplicity? *Requiescat in pace.*

4. Induced Electric Fields.—Looking elsewhere, therefore, for the cause of electromotive forces induced in stationary coils, we may recall one type of force already seen definitely to act on stationary electrons; namely, electrostatic force due to other static charges. Can forces of this type account for induced electromotive forces in stationary coils? Not exactly; yet the idea is not without merit. We can profitably extend the definition of electric field strength, given in the section *Electrostatic Fields* above, to make it include what we need now.

The definition given there was stated in terms of a little test charge on which the field could exert a measurable force. Calling this test charge Q_t and the force F_t, the field strength E was defined first roughly as $\frac{F_t}{Q_t}$ and then exactly as the limiting value which $\frac{F_t}{Q_t}$ approaches when Q_t is made to approach zero. All that is needed now is to let the measured F_t include not only the electrostatic force on Q_t but also any force on Q_t due to time changes in magnetic fields.

One may well question the advisability of blending the effects of such apparently different causes as these. Actually, however, this blend becomes inevitable when the charges become as far from static as they are in radio and other high-frequency circuits. If any charge gets notably stronger or weaker during the time required for a radio wave to travel from this charge to the test charge Q_t, there is no reasonably realistic definition for the force it exerts on Q_t that does not refer also, explicitly or implicitly, to changes in the magnetic field. So the causes of the force F_t are themselves blended whenever they change very fast; the defined field strength E, then, cannot help being a blend.

However, it is not so much of a blend as to include any Lorentz force. One of the specifications on Q_t was that it must be stationary during the measurement of F_t. No electron in Q_t can have any force calculable as evB when its speed v has to be zero. So E is no more of a blend than it has to be.

At low frequencies it scarcely has to be a blend at all. A lamp marked 120 v. shines normally in any lamp socket in the house because the laws about potential differences govern the lamp and the wires leading to the sockets almost as exactly at 60 cycles per second as with direct current. The electric field near the lamp and these wires is almost entirely due to charges. At such a low frequency as this, we can call the part of E that is due to charges the electrostatic component of the field.

Only in or near something strongly magnetic, such as a motor or a transformer, do we have to consider seriously the part of a 60-cycle electric field that is due to time changes in magnetic fields. In calculations here we still speak of potential differences but also of induced electromotive forces. In so doing we are actually separating the two ingredients of the blend E. Potential differences are related only to the electrostatic component. Induced electromotive forces are related only to the other component, so we may call it the induced electric field.

Induced electric fields have characteristics quite different from those of electrostatic fields. Lines of force of induced fields do not run from positive charges to negative. Instead, they are loops, much like lines of force of magnetic fields. If the core of a transformer is a thin circular ring of iron, with its windings distributed uniformly around it, the lines of force of its induced electric field are exactly like those of the magnetic field of a circular coil of wire, such as the one of fig. 22. Even the differences, from place to place, in the strength of this induced electric field are exactly like those of that magnetic field.

Since the wires of this transformer are looped around its core, they practically fit some of the lines of this induced electric field. That is why these wires are put there: to let this field drive their electrons along them. Their electromotive forces are simply results of this driving. These electromotive forces are proportional to the numbers of turns of wire because the electrons are driven the same way along the wire throughout its length, whether this is only one turn or many.

Induced electric fields generally are much like this one; their lines are always looped. Thus they are always much like magnetic fields and are quite different from electrostatic. And they always cause electromotive forces in coils placed like their loops.

5. The Relativity of Moving Coils.—Among the coils with electromotive forces due to induced electric fields we may include the exploring coil in the presence of the larger, current-carrying coil, discussed above, whenever we keep this exploring coil stationary. And keeping it stationary we could induce these electromotive forces either by stopping the current in the larger coil or by moving that coil away. But when we moved the exploring coil, keeping the large coil stationary and its current steady, we ascribed the electromotive force in the exploring coil to Lorentz forces.

Who can say which coil is stationary? We could refer to relativity, but must take great care. In physics at least—not to speak of other disciplines such as ethics—relativity is a system of law, not licence. Relativity does not tell us we may call a coil stationary if we are picking it up and starting to carry it away or turning it around. Acceleration or rotation of any object disqualifies it as a thing to be called stationary.

Rotation, especially, will need attention in this connection because it is often treated as purely relative in explanations of motors and generators. Of course one would expect a false basic assumption to lead to wrong conclusions. This one does, but not on the questions which happen to have the greatest practical importance. On these questions, resemblance to the truth is not purely coincidental, but the reasons for this resemblance do not endorse the false assumption. More is said about this in the section *Rotation, Relative to What?* below.

The rotation of the earth, therefore, raises a question on our right to call the earth stationary. Its rotation is a major factor in some calculations, such as those of meteorologists on the relations of winds to air pressures. When we call the earth stationary, as we shall here, we should recognize that we are not calling rotation purely relative but merely calling this particular rotation too slow to affect present problems appreciably.

Now, therefore, we may call the large coil stationary if it rests on a firm support. Alternatively—but not in the same breath—we may call the exploring coil stationary if it moves straight along the axis of this firmly supported large coil, always keeping its speed constant and never rotating.

If now one observer stands on the ground, calling the large coil stationary, while another observer rides along with the exploring coil, calling it stationary, we may compare their reports on the electromotive force induced in the exploring coil. The ground observer ascribes this electromotive force to Lorentz forces; the rider, to an induced electric field. Under these carefully prepared conditions both observers have unquestionable rights to their view-

points. How, now, can relativity give them a right to differ in their reports? In a preceding section, *Amperes* and *Volts*, we had occasion to examine carefully the force of attraction between two parallel wires with "like currents." There, under *Special Relativity*, we found just such a difference. One observer called the attraction magnetic; another observer called it electrostatic.

In short, with the parallel wires as well as with the coils, two observers, both qualified to call themselves stationary, are found to differ as to the absence or presence of an electric field.

Their right to differ extends to magnetic fields as well as electric. One of Rowland's epoch-making researches was an experiment in which he observed and measured a magnetic field due to motion of a set of electrostatically charged disks. Therefore if we have any charged body moving straight and steadily we can observe a magnetic field around it, but we have to admit that an observer riding along with it and calling it stationary will deny the existence of any such magnetic field.

He will even differ with us, though only very slightly, about its electric field. As we see the body come nearer to us and then go away, we say its magnetic field, where we stand, is at first getting stronger and then getting weaker. So we can observe electromotive forces in circuits we call stationary. Therefore we see an induced electric field, along with the electrostatic component, where he sees nothing but electrostatic.

If now we have any temptation to ascribe our electromotive forces to motion of the magnetic lines, saying the body carries them along, he can protest, "No such thing! I find no magnetic lines for it to carry!"

This is a far cry from the old view of the fields, both magnetic and electric, as practically substantial, with lines of force almost like stretched rubber bands. The change came in steps. First the lines of both fields lost this bodily strength, although the ghostly forms of the magnetic lines still stalked through the laboratories, inducing electrons to move aside if they could. Now even these ghosts, if not otherwise laid to rest, can often be exorcised by appeals to relativity.

Thus either field, magnetic or electric, can depend for its very existence on the state of motion of him who observes it. Nevertheless there is no way to exorcise both fields at any one point in space and time. That is, if any observer finds a field of either kind at this point, there is no way for any other observer to move so as to observe zero strength for both fields there.

What remains, however, or more specifically what stays strictly objective, in the sense of having the same value for all observers qualified to call themselves stationary, is a rather abstruse mathematical combination of the fields (*see* RELATIVITY).

Fortunately, any one qualified observer's pair of fields can be translated into any other qualified observer's pair by means of a rather simple set of equations. These equations were discovered by Lorentz, and are called the Lorentz transformations for fields; but they are among the equations taken over by Einstein into relativity.

Relativity, therefore, not only allows but even compels observers to differ on the natures of forces, as in the cases discussed above. But on the strength and direction of any force (as on the lengths and directions of their yardsticks) no observers in any such case ever differ by any amount large enough for detection by any instruments at our command.

After all, now, if one observer calls the force acting on an electron a Lorentz force, and he calculates it as evB, while another observer ascribes it to an induced electric field and calculates it as eE, anyone may well ask, "What's in a name?" There is this: a warning, in making any calculation, to consider which observer's viewpoint we are taking. To slip over into the other's viewpoint unwittingly is an easy way to omit a force from the calculation, or to insert the same force twice under different names or even to bog down in hopeless confusion.

Many problems are simple if any one observer's viewpoint is used consistently, regardless of whose viewpoint it is, but hopeless if another's viewpoint is brought in by the use of his name for some force.

D. Magnetic Flux

In motors and generators, Lorentz forces and induced electric fields often act on the same wires at the same time. This happens whenever a wire is moving in a magnetic field while this field is changing with time. Fortunately this situation does not involve any new force; there are only these two, and they add themselves together quite simply.

Sometimes, indeed, they add themselves so simply that they seem to be merged into a single force, because we need only one equation for calculating their combined electromotive force. This equation is a good short cut, but, like most short cuts, it can be tricky. Once in a while it is easy to go astray here. So it is well to note that the basic laws, by which the short cut may be tested in any dubious case, are those of the electric field and the Lorentz force; and, especially, to note that these laws must be used with the viewpoint of only one observer and that he must be qualified by relativity to call himself stationary.

1. Defining Magnetic Flux.—The short-cut rule does not apply to single, straight wires, but only to loops of wire. The reason for this restriction is simple: the rule uses a quantity called magnetic flux, which can be defined only for a complete loop.

We have seen this quantity already in some special cases; namely, the exploring coil and either winding of a transformer. In these the flux is the product AB of the area A enclosed by the loop and the flux density B within the loop. We saw this product in the equation

$$\mathcal{E}_{ave} = \frac{NAB_{max}}{t_0}$$

where $\mathcal{E}_{ave}$ was the average of the values of the electromotive force $\mathcal{E}$, during the time interval t_0, in which the flux density B changed from zero to B_{max} or vice versa. Here N was the number of loops, or "turns," in the coil and they were all alike, so that each loop enclosed an area A in its plane. The product AB, then, was the flux through any one loop at any instant when the flux density had the value B.

These cases, however, were especially simple in several ways: (1) each loop was flat; (2) the flux density was uniform, all over the area A; and (3) the magnetic lines of force were all at right angles to the plane of the loop. But fortunately the simple product AB needs only rather obvious modifications when we depart from these conditions.

Departing first from condition (3), let a flat loop of wire be made oblique to the magnetic lines, as in fig. 39.

Obviously there are some of these magnetic lines which could lead through the loop L if its plane were at right angles to them but cannot when it is oblique to them. If the lines of B were lines of flow of water (and this analogy is behind the term "flux"), less water would flow through the loop in its oblique attitude than if it were at right angles. The flux is defined for the oblique attitude on just this basis: it is proportional to this imaginary flow.

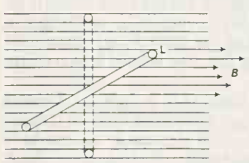


FIG. 39.—CHANGE IN MAGNETIC FLUX IN FLAT LOOP L AS ITS ANGLE TO LINES OF FORCE B IS ALTERED (*See* TEXT)

Of course there are many ways to specify this proportionality mathematically. For example, one can set up a plane at right angles to the magnetic lines and calculate the area on this plane which includes all the lines that do lead through the loop in its oblique attitude; this area is obviously less than the full area A enclosed by the loop. Mathematically this area is A times the sine of the angle between any magnetic line and the plane of the loop. So the flux is AB times this sine.

Now let us remove another of the conditions listed with the initial case, where the flux was AB; namely, its condition (2) that B was uniform over the area A. Keeping the principle that flux is analogous to flow of water, we may now treat the area A to the process called by realtors "subdivision." The dividing lines in it may be laid out along the threads of a fine net stretched over the

loop. If the meshes of the net are so fine that there is not much difference between any values of B within any one mesh, the flux through any one mesh can be calculated roughly with any value of B within it. Roughly, therefore, the flux through the whole loop is the sum of the values thus calculated. Exactly, of course, the flux through the loop is the limit which this sum approaches when the net is made finer and finer. To calculate this limit by arithmetic would be laborious; by integral calculus it may be easy.

Finally let us remove condition (1) of the initial case, that the loop was flat. If it is not flat, it is unable to lie in any one plane. On what sort of surface, now, do we measure the area A? Again the idea of a net, with the wire as its edge, is useful. And fortunately the net does not have to have any special shape. If, for example, it has the shape of a soap-bubble film held by the loop of wire, it will have the least possible area for that loop; but it need not have that shape. It may be wrinkled in any way we choose, like the loose nets used by anglers in landing their fish. The only requirement is that the frame holding the net shall be our chosen loop of wire. All nets meeting this requirement give the same value for the flux.

This statement may seem hard to prove or even incredible. Nevertheless, with integral calculus one may prove rather simply that it is only an alternative statement for the law relating point-to-point differences in values of B to divergence or convergence of its lines (see *Flux Densities,* above).

2. Flux Changes and Electromotive Forces.—Having defined the magnetic flux for any arbitrarily chosen loop of wire in any arbitrarily chosen magnetic field, we may now use it with the same freedom. The short-cut rule, for which the flux is defined, is this: if any closed loop of wire experiences a change in its flux during any time interval t_1, the average of the values of the electromotive force induced in it during this time interval is

$$\mathcal{E}_{ave} = \frac{\text{change of flux}}{t_1}$$

Nothing is said in this rule about whether the change of flux is due to motion of the loop or to time changes in the field. It may be due to either one or to both together. The former involves Lorentz forces; the latter, an induced electric field. Their contributions to the electromotive force of the loop are added unobtrusively in this equation.

When motion is one of the causes of this change of flux we are not limited by this rule in any way suggested at all obviously by relativity. Rotation is admissible in this rule. So is a change of shape of the loop; if the loop is made of flexible wire, any change of flux due to bending any part of the loop is admissible here. This rule, evidently, covers a multitude of what relativity might be expected to call sins.

Nevertheless, it covers only what relativity actually lets it cover. As stated before, the test of this rule in any dubious case is to go back to the laws of the electric field and the Lorentz forces, using them with the viewpoint of only one observer, this one being qualified to call himself stationary. In short, this short-cut rule is a mathematical corollary of these laws, though to prove it to be such would take more mathematics than is wanted here.

3. Electromotive Forces in Motors.—Turning now to the practical application of this rule to motors, we may recall first a few of the facts stated above under *Motors.* The most important consideration is of course mechanical power—the motor's reason for existence. Most of this power is produced by forces between "like" or "unlike" currents. Some of these currents are conduction currents in wires or bars, but by far the strongest are the Amperian currents alongside these conductors.

To calculate the power directly from these forces is often very difficult. Not only are the forces numerous and varied; they are also mathematically intractable because the Amperian currents do not vary in proportion to the conduction currents. Instead, the relations between these currents are more like those shown in fig. 34. Such relations are formidable obstacles to calculations. Fortunately, however, we can by-pass such calculations by use of the short-cut rule for electromotive forces now under discussion. With it we can calculate the power in its electric form, before it gets transformed into mechanical power.

In this connection we may recall how the electromotive force in a wire moving through an unchanging magnetic field was treated in the section *The Lorentz Force,* above. There it appeared that the product $\mathcal{E}I$ of this electromotive force $\mathcal{E}$ and the current I in the wire was not only the logical expression for its electric power but also equal to its mechanical power; that is, the electric power was transformed directly into mechanical power, with neither gain nor loss. This rule applied to all electromotive forces due entirely to Lorentz forces.

No corresponding rule has been stated here yet for electromotive forces due to induced electric fields, but there is such a rule. For them $\mathcal{E}I$ represents power taken from or given to the energy of the magnetic field.

Any magnetic field has energy. The stronger the field is, the more energy it has. In a transformer, for example, where the field is continuously changing, the amount of energy in it also is changing. Usually energy is being given to this field by the primary current I_1 at a rate $\mathcal{E}_1 I_1$, and at the same time energy is being taken away by the secondary at a rate $\mathcal{E}_2 I_2$. Only rarely are these rates exactly equal. The difference between them, at any instant, is the net rate at which energy is being given to or taken from this field.

Applying these principles to a motor, we have electric power going into it and mechanical power coming out, with more than one way for the power to be transformed from electric to mechanical within the motor. Some of the transformation is direct, through Lorentz forces. Some of it is via magnetic energy, the amount of which may rise and fall, much as in a transformer. The amounts of power transformed in each way vary greatly from one type of motor to another. But it is not necessary to be concerned over this: the short-cut rule,

$$\mathcal{E}_{ave} = \frac{\text{change of flux}}{t_1}$$

applies to electromotive forces of either type or both together; so, when we multiply such an electromotive force by current, and average this product over a cycle if the motor uses alternating current, we find how much power the motor transforms from electric to mechanical power, regardless of how it does this.

Of course no motor can transform all the power given to it, without losing any by heating its wires and its iron. So its electromotive force is never exactly equal to the potential difference between the wires which deliver power to it through their electrostatic forces. Details on these losses would be out of place here, however. More important is the fact that in a good motor most of the power is transformed. Therefore most of this potential difference is used in overcoming what is called "back-electromotive force."

4. Generators.—Most generators are essentially like motors. Conversely, many motors can be used as generators. In some motors a simple increase of speed, secured by removing the mechanical load and substituting a driving engine, will increase the back-electromotive force so that it is no longer overcome by the potential difference between the wires of the power line but wins the contest and drives the current "backward," thus delivering power to the line. Thus the motor becomes a generator—not generating electricity but electric power. (*See* also GENERATOR, ELECTRIC.)

Whether the electromotive force of a generator is due to Lorentz forces or to induced electric fields, or to both acting together, it comes under the short-cut rule.

5. Rotating Loops.—Since this rule has such a commanding position in practical engineering, one may well ask why we should not give it a fundamental place in the theory of electricity, instead of regarding it as subordinate to the laws of Lorentz forces and induced electric fields. Actually there are several reasons. Especially, there is the fact that it is more limited than those laws; it applies only to closed loops. It gives the value of the electromotive force of a whole loop; but it does not tell how the forces on electrons, producing this electromotive force, are distributed around the loop nor what charges are built up by these forces.

A good case in point is the so-called rotating-field motor. In the simplest kind of rotating-field motor, the magnetic lines of force

are all practically parallel throughout the volume of the drum-shaped rotating body called the rotor; and the direction of these lines changes with time like a very fast clock hand, making 60 revolutions per second. The rotor makes almost as many revolutions per second, but not quite; say 57, for example. The difference, in this case three revolutions per second, is called "slip," and it is often audible as a slight pulsation in the hum of the motor. This slip is closely related to the forces driving the motor. A heavier load, requiring stronger forces, makes a faster slip. Conversely, if the load could really be removed entirely, not even leaving any friction, the slip would be reduced to zero.

If, now, the slip is three revolutions per second, the flux through any loop of wire (or, more usually, of copper bars) in this rotor is changed cyclically at three cycles per second. Therefore this loop has a three-cycle electromotive force and a three-cycle current. This current, and the Amperian current in the iron alongside it, cause the field to exert the mechanical forces (of the type $F = IlB$) which drive the rotor and produce its mechanical power.

Naturally this situation tempts anyone to treat rotation as relative. The electromotive force of any loop as a whole, given by the short-cut formula, really has just the same value as if the rotor were stationary and the direction of the magnetic field made only three revolutions per second. Or, equally well, as if the lines had a fixed direction and the rotor were turning backward at three revolutions per second. So the current, and also the mechanical forces given by $F = IlB$, are the same as if rotation were relative.

From a practical standpoint these quantities, electromotive force, current and mechanical force, are the important quantities. If one's interest in the motor is limited to what is obviously practical, the quickest way to satisfy this interest is to treat rotation as relative.

The quickest way, however, is not necessarily the best way, especially if it is on a false basis that may lead to error in some other case. Nor does a calculation limited to the "practical" quantities satisfy a scientist or anyone else with a scientist's type of interest in "what really makes it tick." Now, therefore, how about the questions raised above? What parts of the loop really feel the forces which drive electrons along its wires, to produce its electromotive force? And what charges, if any, are built up on the loop by the electrons thus driven?

A scientist usually tests first the simplest case he can find. Here this is the case of no load and no slip. His loop may as well lie in a plane containing the axis of rotation and have this axis as one of its diameters. Preferably it should be circular, and the source of the "rotating field" should be as simple and symmetric as possible—say, a hollow magnet, spinning on an axis through its centre, with the loop inside it. Or, best of all, he may combine the loop and magnet; for example, a solid spherical magnet, uniformly magnetized, with an axis of rotation perpendicular to its direction of magnetization, like an idealized earth with its north magnetic pole moved from the shore of Hudson bay to the point on the equator due south of Greenwich: latitude 0°, longitude 0°. The loop now can be like a cable, buried along the meridian of Greenwich and the opposite meridian of longitude 180°, all the way around the earth.

This magnetized globe is mathematically simple. Treating rotation as relative, in the conventional way, one would see no reason for any forces to act on electrons in any part of its loop; therefore no reason for its electrons to build up any charges, anywhere. But is rotation relative in this way? Every meteorologist knows it is not, at least in his line of work. So do many other specialists, such as airplane pilots steering by directional gyros and resetting them now and then to compensate for the earth's rotation.

The question is, do electrons "know" this? Without waiting for the answer to this question, we may grant to anyone who holds his orientation by directional gyros an unquestionable right to his viewpoint. And his mathematical treatment of the rotating globe (at least, if it has electrical conductivity throughout its volume, not merely along one loop) can be made complete and exact in only a few pages of widely spaced handwriting. This treatment shows definitely that the charges are not zero. That is, they are not zero except on the equator of this globe and its 90° meridians, east and west (across its Siberia and North America). The maximum density of excess-electron charge occurs at two points: lat. 45° N., long. 0° (in its France), and lat. 45° S., long. 180° (near its New Zealand). The greatest deficiencies of electrons also are on these 0° and 180° meridians, but at 45° S. and 45° N. (in its South Atlantic and North Pacific oceans), respectively.

The drum-shaped rotor in the rotating-field motor is not so simple, but it must have charges essentially like these. That is, it must have two regions with excess electrons and two deficient regions, located about as on the globe, even when there is no slip and the conventional relative-rotation treatment would predict no charges at all. With slip, that treatment does predict charges, but these are very much weaker than the real charges. On these charges the two relative-rotation treatments (rotor called stationary and field direction called stationary) contradict each other, and both are very far from right.

6. Rotation, Relative to What?—But how, one may ask, can the electrons know the drum is rotating? And especially, how can they know the magnetized globe is rotating, when there is nothing else near it? How, also, can the meteorologist's wind know the earth is rotating? Perhaps they could not, if the earth were alone in space.

But it is not. Strangely enough, there is reason to believe that the gyroscope and the wind get their information from the gravitational field of the stars. If they regard the stars as fixed and the earth as rotating, this field lets them go their own way. But if they regard the earth as stationary and the stars as moving around it, these motions of the stars have to be so fast as to bring on strong gravitational effects. Einstein's theory of gravitation, or general relativity, was shown by Hans Thirring (*Physikalische Zeitschrift*, 1918 and 1921) to suggest that if these effects could be calculated exactly they would "explain" the behaviour of the gyroscope and the wind in this way with a stationary earth. For the wind it is customary in meteorology to treat the earth's rotation as equivalent to some fictitious forces called Coriolis forces, which seem from a stationary-earth viewpoint to be pushing all winds sidewise. Equally fictitious centrifugal forces seem from this viewpoint to be acting on air, water and ground, all alike, so as to keep the earth bulged at the equator and flattened at the poles. Thirring's suggestion, as applied to the earth, is that all these fictitious forces can be regarded as gravitational effects of the stars going around a stationary earth. That is, they can if we choose to take for ourselves this very long way around, mathematically, instead of the very short and easy way to calculate the fictitious forces given us by Newton's laws of motion as applied to a rotating earth.

The point of Thirring's suggestion is not that the longer way is the better. Rather, it is that what the gyroscope and the wind recognize as rotation is rotation relative to the stellar universe as a whole.

This seems to be what the electrons recognize also. If we insist on calling a rotor stationary, we can calculate for the electrons on this basis, provided we introduce certain fictitious charges and currents along with the real ones. Equations for these fictitious charges and currents, covering all steadily rotating systems, were derived from general relativity by L. I. Schiff (*Proceedings of the National Academy of Sciences*, 1939). Use of these charges and currents is therefore a logical extension of the procedure begun in mechanics with the fictitious forces.

In practice, convenience calls for one procedure in meteorology and another in electricity. Coriolis's formula is simple; Schiff's equations are comparatively complicated. They prove that we *can* use the viewpoint of a steadily rotating system if we choose to use it; they tell us exactly how; and they show us its complications.

In short, they tell us exactly why the easiest way to get reliable information about all the charges, currents and forces on a rotating system is to recognize its rotation as its electrons do: to hitch our desks, if not our wagons, to the stars.

7. Why Specify Wire?—Wire was mentioned in the short-cut rule as stated above; yet most mathematical rules dealing with loops let their loops be immaterial, geometric figures. Why, then, should we specify wire here?

Actually it need not be literally wire. The point here is that any loop used in this rule must be composed of real matter, and like a loop of wire it must be composed of the same real matter throughout the time interval mentioned in the rule.

Strangely enough, though Faraday knew this, it is often overlooked; yet it is essential. To see why, let us consider briefly a case described by C. Hering (*Proceedings of the American Institute of Electrical Engineers,* 1908) sometimes called the "trouser-clip experiment."

Hering asked this question: suppose a clip, such as is used by cyclists to keep their trouser legs from touching their bicycles, is put around a long steel bar magnet, with its ends touching each other; then it is slipped off the bar, as off a cyclist's leg, keeping its ends in metallic contact, either with the bar or with each other, at all times; during this operation the flux through the clip is greatly reduced; so, now, does it, or does it not, have the electromotive force given by the rule as applied to this change of flux?

Classes of graduate students in physics and electrical engineering often vote that it does, but the fact is that it does not. This fact can be established by a simple experiment, also described by Hering. The clip can be cut into halves, and its halves can be mounted on a block of insulating material, and then their bases can be electrically connected to each other through a flexible lamp cord leading to a galvanometer. The galvanometer shows no current when the clip, thus prepared, is slipped off the magnet in the way described above, even though it may be sensitive enough to show current readily when the clip is moved along the magnet and off past its end.

In both these motions there is the same change of flux through the clip. In both, metallic contact is maintained at all times, to let electrons flow if they feel induced to flow. But in the former case the loop is not composed of the same matter at all times; in the latter case, it is.

The significance of this test, for electrical theory, is that it emphasizes the mathematical relation of the short-cut rule to the laws of induced electric fields and Lorentz forces. In deriving the short cut from those laws, one has to postulate that whenever the loop moves, real matter, containing electrons, subject to Lorentz forces, must move with it.

Conversely, if the loop runs through real matter which moves, but the loop itself does not move with the matter, again the short-cut rule fails. Faraday discovered this, so the apparatus used for showing it is called a Faraday disk. Many Faraday disks are made of copper and spun on shafts, but mercury swirling around in a glass jar is easier to set up. A pair of wires from a galvanometer dip into the mercury, one near the middle of the jar and the other near the side.

Now the loop consists of the galvanometer, the wires and a line through the mercury. On this line matter moves while the loop does not.

If now there is a vertical magnetic field there, to exert Lorentz forces on electrons in the mercury, there is current in the galvanometer; and this current shows that there is an induced electromotive force in this circuit even though there is no change of flux.

8. What Is Basic About Flux?—The short-cut rule of electromotive force and flux is clearly not so simple as it looks. So we may do well to summarize briefly here some of the most important facts about it.

As applied to moving loops of wire this rule is deducible mathematically from the less restricted and therefore more basic laws of electric fields and Lorentz forces. For moving loops, it is thus not fundamental, though it is extremely useful—provided its restrictions are kept in mind.

Turning to stationary loops, and thus eliminating Lorentz forces, we find this law dealing with electric fields exclusively. It still deals with whole loops rather than single points in space. But in this respect it is like one of the two basic laws of magnetic fields discussed in the section *Flux Densities;* namely, the law about magnetic fields on different closed loops. That law for magnetic fields could not govern those fields alone; it had to work in partnership with another law. So, also, does this law for electric fields. This much they have in common, at least, and much more, actually, as we shall presently see.

VIII. ELECTROMAGNETIC WAVES

Throw a stone into a pond, and once the waves from its splash are gone there is little evidence that the stone is there. The surface of the water must be a little higher, of course, but only a little. If the whole surface had to rise at the same time, so that it could not move any more than this, we could not perceive the motion. Only by waves can energy from the splash be made perceptible at much distance. So also it is with electricity.

This analogy is imperfect, like most analogies; but its main idea is correct. Energy comes to our radio and television receivers in electromagnetic waves. There is not much energy—often so little that only very ingeniously designed receivers can make use of it. But the best of receivers would be silent if a radio transmitter had to spread its effects over all space instantaneously, rather than by waves.

The receiver depends on waves for its very existence. These waves were not discovered by accident, nor by "try everything," nor even by the man whose experiments furnished the first evidence for them, though he was one of the best of all experimenters, Michael Faraday. He did guess that his lines of force might in some way carry the waves we call light; but the reason to believe positively that there could be electromagnetic waves was first found by James Clerk Maxwell.

A. Maxwell's Theory

Maxwell was systematically studying Faraday's research records and conclusions and translating Faraday's verbal statements of his laws into the more exact language of mathematics. Among these laws was the one relating electromotive forces to flux changes. In Maxwell's mind this became an especially good example of mathematical elegance.

In 1855, nine years before his discovery of electromagnetic waves, he compared magnetic lines of force to stream lines in the flow of an incompressible fluid. He was careful to say that he did not mean that magnetic flux really was such a flow. What he did mean, he said, was that a great deal of the mathematics of such a flow could be applied to magnetic fields. By this mathematics he found for us our freedom, when defining the magnetic flux through a loop of wire, to spread the net on our loop in any shape we like.

Maxwell used this mathematics in other ways also; for stream lines of steady electric currents in conductors of all shapes, and even for electric fields. He called his fluid "imaginary" and gave it properties very different from those of any real fluid. It was most obviously imaginary as applied to electric fields: every positive charge had to be a "source" of the fluid, steadily creating the fluid and sending it away along the lines of force; every negative charge had to be a "sink," into which the fluid had to flow, to be sunk without a trace.

Fantastic as this fluid was, it was useful: it gave the mathematics of incompressible flow to electric fields as well as to steady currents and to magnetic fields.

The most valuable feature in this branch of mathematics is its freedom from restrictions on the shape of a net to be used for catching, in one's mind, something that goes through a loop. Magnetic flux is one of the things for which it gives us this freedom. Another is a steady, direct current.

Direct current was what we saw first in this article as going through a loop. We saw it as the current in one side of a circular coil, the loop being any one of its magnetic lines of force, as shown in fig. 22. We had no need for a net there. Now, however, we may check the validity of the idea of a net for direct currents by reference to any one of those loops. The loop may hold two nets, a flat net or a net like an angler's landing net. These nets will "catch" the same current, because the same wires lead through both nets and no electrons are stopping or starting between the nets. So goes the law for all loops around direct currents: on any such loop, all nets are good.

1. Open-Circuit Currents.—How about loops around other

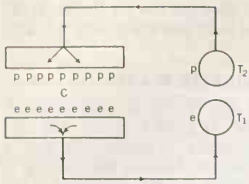


FIG. 40.—OPEN CIRCUIT CONSISTING OF HIGH-VOLTAGE CONDENSER C AND SPARK TERMINALS T: (E) EXCESS ELECTRONS; (P) UNCANCELLED PROTONS (*See* TEXT)

currents; for example, around the current, necessarily not direct, in the open circuit shown here in fig. 40? To prepare this circuit for a transient current, the spark terminals T_1 and T_2 may be moved apart and the condenser plates may be charged through an electrostatic machine. Next, the wires from this machine may be removed. Finally, the spark terminals may be pushed toward each other till a spark flashes between them.

This flash is very short-lived. The charges on the plates may be gone in a millionth of a second. The current lasts somewhat longer, because its self-induced electromotive forces keep it from stopping as soon as the charges are gone. They keep the current going long enough to give the plates charges of reversed signs, almost as strong as the initial charges. Then the current surges back and forth, perhaps 10 or 20 times each way. Even so, it is usually all over in less than a thousandth of a second. Nevertheless it has had time to cause detectable magnetic effects; and some of these effects were known to Maxwell. So the question was before him: what about magnetic loops and currents through nets in such a case as this?

Let the plates be circular, for example, and let us consider a circular loop around them, as shown in fig. 41.

This loop LL′ has no wire at all through its most obvious net, the flat net N_0; yet there must be a magnetic field on the loop. How about net N_1? Or N_2? Wires do lead through them. Does the current through either of these nets give the right value for the field on this loop? If so, what can be done about net N_0, to make it equivalent to either of the others?

To questions such as these most physicists of Maxwell's time gave very elaborate mathematical answers—yet their answers were wrong. Maxwell gave simple answers.

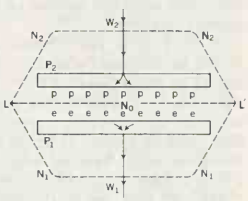


FIG. 41.—OPEN CIRCUIT

Plates P_1 and P_2 are circular so as to give axial symmetry around wires W_1 and W_2 to the plates; (LL′) loop; (N_0, N_1 and N_2) nets; (p) protons and (e) electrons show charges on plates

2. Maxwell's "Displacement Currents."—His answers were based chiefly on what he had done mathematically with his imaginary incompressible fluid, but partly also on ideas for which he gave credit to Faraday and to O. F. Mossotti. These ideas were on what an electrostatic field did to the molecules of any good insulating material, or dielectric, as Faraday had called it. In the condenser of fig. 40 or 41, for example, the space between the metal plates might be occupied by glass or wax, which then would be subject to the electrostatic field due to the plates, as described earlier in relation to fig. 11.

The modern view of such dielectrics was an outgrowth of the ideas of Faraday and Mossotti. Of course they had no data on electrons or on many other factors in the modern view. For lack of such data Mossotti made a guess: each molecule might act like a metal sphere, insulated from all the others. In such a sphere, electricity would be displaced under the electrostatic influence of the charges on the metal plates. The bigger the spheres were, the greater these displacements would be, and the more the displaced electricity would react electrostatically on the plates. Conversely, from measurements of the charges and potentials occurring on these plates, Mossotti could tell what fraction of the space within the condenser would have to be allotted to these spheres. This fraction turned out to be quite reasonable.

Maxwell recognized these ideas as good, though he was skeptical about the details of Mossotti's theory. The essence of it was that the molecules of any dielectric contained electricity which could be displaced slightly, and that this displacement was proportional to the electric field which caused it. This proportionality was like the behaviour of springs. So Maxwell wrote of dielectrics as having "electric elasticity."

On this basis he went farther than Mossotti. He postulated that electricity must be displaceable in this elastic way not only within the molecules of dielectrics such as glass, sulphur or wax but even in "what is called a vacuum." The words just quoted were not directed at the imperfections of vacuum pumps but at the mysterious, half-substantial medium called the aether that was supposed to pervade all space, with or without matter, and to defy an ideally perfect pump.

In Maxwell's theory the electric displacement in vacuum, with an electric field of any given strength, would be about half as great as in wax. This was not an unusual difference, because the electric displacement in wax would be even less than half as great as in glass. The difference between vacuum and wax or glass, therefore, did not seem to be in kind, but only in degree; and at that, not great.

Maxwell's simple answer about those nets was therefore this: electricity could be displaced through net N_0, even if this net were in vacuum; electricity would be moving through this net whenever the electric field due to the plates was changing with time; and finally, this motion of electricity, which Maxwell called a displacement current, would always be exactly equal to the current through net N_1 or net N_2. Thus all three nets became valid for use in the law of the magnetic field, just as if the wire led through them all without any break.

This exact equality of their currents came from another of Maxwell's postulates: that electricity was always strictly incompressible. Whenever any electricity flowed into one of the metal plates through its wire, this incompressibility would force an equal amount of electricity to flow out of the plate by displacement in the dielectric.

Nothing could cause any exception to this rule. In one test the space between the plates might be filled with glass; in a second test, with wax; in a third, it might be a perfect vacuum. A given electric field would cause different displacements in these tests, but the rule of incompressibility would carry these differences to all parts of the circuit. The currents through nets N_0, N_1 and N_2 would be strictly equal in all cases.

Thus, in 1864, Maxwell's mathematics of incompressible fluids took on a greatly increased importance. His incompressible fluid of 1864 was very different from the vintage of 1855. One might call 1864 a rarer vintage, not to be found flowing steadily from every positive charge, nor to be sacrificed to the insatiable thirst of every negative charge, nor to be called "imaginary." It was to be real electricity.

Where, now, can we find this vintage of 1864? We have electrons, also protons, alone and as parts of complex atomic nuclei. All matter is made of electrons and nuclei. In terms of them we see clearly the displacements of electricity that Mossotti groped for within molecules; but in vacuum we see nothing to displace. Nevertheless we still use Maxwell's term "displacement current" and with very good reason. We might not choose this term now, if we were coining a word on the basis of modern knowledge; but as word usages go, the continued use of this term is not exceptionally illogical. At least, in most cases our "displacement current" is mathematically equal, if not physically identical, to Maxwell's.

3. The Fields Are Set Free.—Maxwell's postulate about the magnetic field around a displacement current freed the fields from their last direct attachment to matter. No longer did a magnetic line of force have to encircle a wire or an Amperian current, as one link of a chain encircles a side of another link. A magnetic line of force, like the line around net N_0 in vacuum, could exist now if it encircled nothing but a changing electric field.

Similar freedom already had been given to electric lines of force: they could exist, as lines of an induced electric field, by encircling a changing magnetic field.

Now, therefore, each field could derive its sustenance, as one might say, from a time change in the other field, without any need for anything more substantial. Thus together the fields were free

to slip off the shackles which had bound them to matter. They had become like the motion of water in the surface of a pond, which can start as a splash of a stone and then detach itself from the stone and go on as waves. Now the electric and magnetic fields, if properly linked together, could detach themselves from wires and go on as "wireless."

4. Maxwell's Plane Waves.—How to make the fields detach themselves was not evident. A better prospect for discovery lay in the question as to how they would act after getting thoroughly detached. That is, after getting like waves on the pond, very far from the splash. By "very far" we mean at distances many times the distance from one wave crest to the next.

If now we focus attention on a relatively small area, measuring only a few times the crest-to-crest distance, or wave length, we see only a small part of each crest. This part, though really an arc of a circle, is only a small fraction of the circle, so we may approximate it with a straight line. In short, by "thoroughly detached" waves from the splash we mean waves whose crest lines are straight.

From these to Maxwell's thoroughly detached waves is a transition best made in two steps. The first step is from the two-dimensional surface of the pond to three-dimensional space. Corresponding to the circular waves from the splash, we have spherical sound waves from an explosion. Corresponding to straight waves on the water, we have plane waves of sound. If sound with plane waves is travelling horizontally, due east, for example, its crests of high pressure are vertical, north-south planes. Such planes, therefore, can serve also as crests for Maxwell's plane waves.

But crests of what? Not of pressure but of electric field strength. One of the conditions for the existence of these waves, which Maxwell found mathematically, was that all the electric lines of force would have to be parallel to the crest planes. With vertical, north-south crest planes the electrical field could be vertical or it could have any slant in a vertical, north-south plane, but it could not have any inclination eastward or westward. Choosing vertical lines, for example, to go with the analogy with water waves, we may apply the word "crest" to an upward electric field, and "trough" to downward, and picture the electric lines of force as in fig. 42.

As these waves travel eastward any stationary particle (be it an atom if these waves are light, or a radio receiver if they are broadcast waves) feels the effects of different parts of the waves at different times.

In this respect the particle is like a chip of wood on the ocean, rising and falling on its waves. When a crest is at the particle, it feels an upward electric field; when a trough, downward. The electric field at the particle changes cyclically as the waves go by.

The front of every wave in fig. 42 is therefore a region of fastest change from a downward field to upward; the back is a region of fastest change the other way. And these changes are what we still call "displacement currents." We call them this because they really do have the same magnetic effects as any other currents, just as Maxwell expected and regardless of any slurs we may cast on them as currents for their lack of electrons.

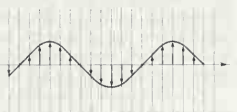


FIG. 42.—DISPLACEMENT IN LINES OF FORCE CAUSED BY TRAVELING WAVE

With these oppositely directed currents, the front and back of a wave are like the two sides of a vertical coil of wire; between them, therefore, there is a horizontal magnetic field. Between them also is the crest of the wave. So this crest consists not only of a vertical electric field but also of a horizontal magnetic field. So also does a trough. The magnetic lines of force are at right angles to the electric; and both sets of lines lie parallel to the crest planes and perpendicular to the direction in which the waves are traveling.

As noted previously, each field in Maxwell's waves depends on time changes of the other field. Now we have seen how the magnetic field depends on time changes of the electric field. Turning, therefore, to time changes of the magnetic field, we find them also to be most rapid in the front and back of each wave. Such changes in a magnetic field with horizontal lines, north or south, can be expected to induce electromotive forces in circuits whose planes are perpendicular to these lines. Such a plane is precisely that of fig. 42. So a circuit should have the strongest electromotive forces if it lies in the plane of this figure; indeed, it does. Radio direction finders are just such circuits. Looking again at the figure we can see that a little circuit in its plane, around any one of the points where the wave crosses the horizontal line, has an upward electric field along one of its sides and downward along the other. In short, the electric field in the waves is an induced electric field, related like any other induced electric field to time changes of the magnetic field.

All's well with each field, therefore, in its relation to time changes of the other field, provided they have the directions shown here: at right angles to each other and also the direction of propagation of the waves. Moreover, this proviso is necessary. The relations could not hold with any other directions. Maxwell proved this mathematically, and the equations relating each field to time changes of the other are known as Maxwell's equations.

5. Comparing the Waves With Light.—Maxwell found some significant resemblances between the properties his theory would give these waves and properties already known for light.

Many experiments had shown that light had a wavelike periodicity, so it had been reasonable to model a theory of light on that of sound. Many of these experiments had shown also that any motion occurring in light waves must be transverse; that is, at right angles to the rays along which the light travelled. This was puzzling; transverse motions could never occur in sound travelling through air or water, though they did occur underground, in sound travelling through solid rock. But light came to us from the sun and the stars. Could it be coming through anything like solid rock? Or if it could, how could the earth move along its orbit through such a medium?

Answers to these questions could not deny light its property of being transverse. The evidence was too clear. Instead, the answers invoked the properties of pine tar, which could seem hard when struck with a hammer, but soft if given time to yield. Such answers were far from satisfactory.

Even if they were accepted, the theory was not good. Sound in rock is not all transverse. Much of it is longitudinal, like sound in air or water. If any sound does happen to have purely transverse waves when in rock of one kind, passage into rock of another kind will usually start longitudinal waves along with the transverse. But light, passing from air to glass or water, for example, persisted in being purely transverse.

So, also, would Maxwell's waves. His equations, like light itself, had no way to make any waves but transverse. Nor did they fill interstellar space with rock, not even with tar. All the worst difficulties of the theory of light were swept away by Maxwell's waves.

Even so, these waves had to pass other tests. The most important was on their speed in vacuum. His equations related each field to time changes in the other field. For given fields he could calculate required rates of time change and thus calculate the speed of the waves. He found that all waves in vacuum would have to travel at the same speed, even if they had different wave lengths. This speed would depend only on the constants of Coulomb's and Ampère's laws, which could be measured by methods independent of light. These constants are the ones described above, in relation to those laws: the permittivity of free space, ϵ_0; and the permeability of free space, μ_0. Maxwell found that the speed of his waves in vacuum would have to be simply

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}$$

On this there was no place in his theory of anything adjustable, no hiding place for any uncertainty. With all the firmness of his strictly incompressible fluid his theory demanded this speed for waves in vacuum. Light in vacuum does travel at this speed.

Light not in vacuum is not so simple. Whether it is passing through a transparent material or being absorbed in something

opaque, any calculation about it involves properties of matter of some kind. When these properties are measured in electrical experiments but used in relation to light, the calculation involves also some theory as to how these properties change when the frequency of subatomic vibrations is increased by a factor of 1,000,000 or more. This element of uncertainty caused differences between calculated and measured quantities. Maxwell wrote, therefore, in regard to the first test of light in a transparent material described in his book *A Treatise on Electricity and Magnetism* (1873):

"The difference between these numbers is greater than can be accounted for by errors of observation, and shews that our theories of the structure of bodies must be much improved before we can deduce their optical from their electrical properties."

Indeed there was much to be done in such improvement. Some of it could be done fairly soon, but most of it had to wait a quarter of a century for the discovery of the electron. Meanwhile the guiding principle was at hand in Maxwell's theory of electromagnetic waves.

B. Hertz's Experiments

For a guiding principle to be at hand is one thing, to be allowed to guide is another. Maxwell's theory was not accepted at once by all physicists. Not all could see, as he did, that the next great need for improvement after his work was in "our theories of the structure of bodies," rather than in the theory of light itself. Many of the "differences between these numbers," theoretical and experimental, were large enough to be disconcerting; the mathematics also was disconcerting to many physicists of that time; so the echoes of sound waves in tar were heard by college students born after Maxwell had died.

Not that the mathematics of sound waves in tar was any simpler than Maxwell's; on the contrary, it was much more complicated. But the ideas were familiar. To many physicists these elastic-solid waves seemed more "real," even though the supertarry "solid" that was supposed to carry them was utterly fantastic. And these waves, by not suggesting any deductions of optical from electrical properties, kept themselves above criticism for any errors in such deductions. Altogether it was natural, even though not completely reasonable, for many scientists to be skeptical about Maxwell's theory.

There was a serious need, therefore, for an experimental test of this theory on its own ground, electricity. If such a test denied the theory, Maxwell's waves would be gone. If not, they would seem "real" enough to come into their own in both optics and electricity.

The need for such a test was seen clearly by one of the greatest of the German physicists, H. L. F. von Helmholtz. So he made electromagnetics the subject of a prize competition for students. The prize was won by his brilliant pupil Heinrich Rudolf Hertz (1857–94), who devised and performed the test in 1888.

1. Hertz's Oscillator.—The first step, as Hertz explained in a lecture, was to devise something that would act quickly enough for speed measurements on waves at the speed of light. Pre-Maxwellian theories said one circuit would act on another at any distance instantaneously; Maxwell's theory would delay the effect, if the circuits were 100 ft. apart, by about a ten-millionth of a second.

The quickest thing at hand was a condenser circuit, such as that of fig. 40. But the quickest oscillation observed in any such circuit took a whole millionth of a second. Trying to time electric waves with such circuits was, as Hertz said, like trying to mark off the length of an almost microscopic object with broad chalk lines. So he had to devise something much quicker.

This "something," the Hertz oscillator, was a pair of metal rods, placed end to end with a small gap for a spark between them. When these rods were given charges of opposite signs, strong enough to spark, the current would oscillate back and forth across the gap and along the rods. The period of this oscillation would be only a hundred-millionth of a second if each rod was about 30 in. long; only a thousand-millionth for 3 in. Here were clock ticks quick enough to time Maxwell's waves.

Not many ticks came with each spark; the oscillations died out too fast. In Hertz's words (as translated by D. E. Jones and G. H. Schott):

"It would suit our experiments much better if this were not the case. But there is still the possibility of success if we can only get two or three such sharply-defined indications. So in the realm of acoustics, if we were denied the continuous tones of pipes and strings, we could get a poor kind of music by striking strips of wood."

Strips of wood, indeed! Little could Hertz foresee the forests of branching television antennas over our residential districts, each pair of branches a lineal descendant of his electrical xylophone.

2. Hertzian Waves.—With this oscillator Hertz solved two problems: the problem of timing and the problem of how to make the electric and magnetic fields detach themselves from wires and go free as Maxwell's waves. He solved the former of these problems in the way just described; the latter, both experimentally and mathematically. Physicists use his name not only for his oscillator but also in the term "Hertz vector." This term denotes a sort of mathematical steppingstone which he devised for use between the current in his oscillator and the fields in its waves. We use this vector still for transmitting antennas and arrays of antennas, in forms compounded according to his rule.

Calculating with it, one can show easily that the waves from an open circuit, such as the condenser circuit of fig. 40, are stronger than those from a closed circuit of the same size and shape with the same current. For still stronger waves, the circuit may be made more open by moving its condenser plates farther apart. Moving them as far apart as their wires and spark gap will let us means laying these wires out along one straight line. When this is done, the charges are not so concentrated on the plates as they were but are spread more over the wires. So now the circuit has become practically a Hertz oscillator. In short, the Hertz oscillator is the best form of simple circuit for producing strong waves.

For it the Hertz vector shows at a glance (with mathematically trained eyes) that the waves are strongest on lines at right angles to the line of the oscillator, and that on these lines the electric field is parallel to the oscillator itself. Accordingly, Hertz set his oscillator with its line vertical, near one side of a large lecture room, and made tests elsewhere on its level for waves with a vertical electric field. He found them with any one of several receivers. One of these was a pair of wires, end to end like the oscillator but with the gap between their ends so small that very much weaker charges would suffice to cause a spark. Detection of electric waves with this receiver was essentially like detection of the sound of a tuning fork with another fork of the same frequency set into vibration by the sound waves. The chief difference is one coming directly from the fact that electric waves are transverse: placing this "interrupted wire" receiver at a point on the same level with his vertical oscillator, Hertz found it would spark if it also were vertical but not if it were horizontal. The electric field of the waves, which drove electricity along the interrupted wire, was indeed vertical, just as his calculations said it would be.

With this receiver he could find out also whether the field really came in waves. For this purpose he hung a large, flat sheet of zinc on the wall of the big room, opposite to the side where his oscillator was. This sheet should reflect electric waves. The electric field in the reflected waves could either reinforce or oppose the field in waves still on their way to the reflector; it should reinforce at some places and oppose at others, in ways depending on distances from the reflector. At least, so Maxwell's theory said. The older theories showed no possibility of such differences from place to place. Hertz found them: wherever Maxwell's theory predicted a deadlock between opposing fields, Hertz's receiver showed no sparks.

He could even verify Maxwell's prediction of the speed of the waves. By calculating the frequency of his oscillations from the size of the oscillator, and measuring the distances between places where the receiver would not spark, he proved that their speed was that of light.

He went on with other tests: reflections of various sorts from sheets of metal, flat or curved; refraction in a big prism, made of asphalt; and other tests suggested by known phenomena of light.

In all tests, Maxwell's theory was confirmed.

C. Wireless

From Hertzian waves to wireless telegraphy was but a step—a long step, but not as long as the one Hertz himself had taken. Wireless waves were not so far from Hertz's waves as his were from light—or, perhaps more exactly, from previously known forms of light. For, as he said of his experiments,

"In carrying them out we are decidedly working in the region of optics. . . . Starting with purely electrical phenomena we have gone on step by step until we find ourselves in the region of purely optical phenomena. . . . Optics is no longer restricted to minute ether-waves a small fraction of a millimetre in length; its dominion is extended to waves which are measured in decimetres, metres, and kilometres. And in spite of this extension it merely appears, when examined from this point of view, as a small appendage to the great domain of electricity. We see that this latter has become a mighty kingdom. We perceive electricity in a thousand places where we had no proof of its existence before. In every flame, in every luminous particle we see an electrical process. Even if a body is not luminous, provided it radiates heat, it is a centre of electrical disturbances."

Comparing electric waves as musicians compare sound waves, we find all the visible colours in just about one octave. At the time of Hertz's experiments physicists knew one octave of higher frequencies, called ultra-violet, and four or five octaves of lower, infra-red. The infra-red faded out gradually, in the detecting instruments of that time, because atoms or molecules in hot bodies radiated less energy at lower frequencies.

So one might call Hertz's oscillators gigantic artificial atoms, artificially forced. As such, they brought the spectrum to life again, about 18 octaves below the red, and extended it 4 or 5 octaves farther. This brought the wave length up to about ten metres. So when Hertz spoke of kilometres he was foreseeing another seven or eight octaves.

Foreseeing was enough. Hertz, as a boy, had studied to be an engineer. At the age of 20 he turned to physics because he had found himself more interested in learning the laws of nature than in applying them. Now at 31, though he foresaw this further extension of the spectrum, he turned to a more interesting problem on light: its relation to cathode rays, then an outstanding mystery. This was a good choice; in these rays, three years after Hertz's untimely death in 1894, Thomson discovered the electron.

1. Wireless Telegraphy.—To others, therefore, went the task of making the extension of the spectrum to wave lengths of a kilometre or more. There was a practical purpose in this. The term "broadcast" is very apt: a Hertz oscillator spreads its energy very broadly; in fact, in every direction except exactly along the line of the oscillator. Very little of this energy can be received by any one receiver far away. To give a receiver far away any detectable amount of energy, the transmitter had to send out a prodigious amount. This requirement involved big oscillators, which made long waves.

In principle, however, they were essentially only Hertz oscillators. The dots and dashes of the Morse code could be sent by simply turning the power on and off at the proper times. Each dot or dash then involved many sparks; each spark, as many oscillations as could be got into it.

This is not to imply that all was simple, nor that it did not require the attention of many very able men; it did. Guglielmo Marconi (1874–1937) was one of them, receiving in 1896 the first of all patents on telegraphy with electric waves. Sir Oliver Joseph Lodge (1851–1940) was another, receiving in 1897 a patent for a method of increasing the number of oscillations in each spark. And there were many more, with a great variety of improvements, including the replacement of sparks with continuous driving of the oscillator throughout each dot or dash.

Regarding wireless telegraphy as an illustration of basic principles, we may recognize Hertz's vertical oscillator in each of the tall towers of a radio broadcast station. We may recognize its vertical electric field and horizontal magnetic field if we take a portable, battery-operated receiver, whose antenna is a flat coil in the back of its box, into an open field, far from wires or pipes, and turn and tip it in various ways, finding what attitudes give weak signals or none. Fliers may recognize also the absence of waves along the line of the oscillator, in what they call the "cone of silence" directly above a station, that tells them they are there. In short, the basic principles are as stated. For other applications, *see* Radio.

2. Wireless Telephony and Broadcasting.—Progress from telegraphy to telephony had been made with currents in wires long before wireless telegraphy was well established. To ask for corresponding progress with wireless waves therefore was simple, but to make this progress was quite another problem. The devices that had made telephony possible with wires would not suffice with wireless waves.

The trouble was with power at the telephone transmitter, not at the receiver; only very little power was needed at the receiver for the sound it had to produce, and only the same low power was needed for the same sound, whether the message it conveyed had come by wires or by waves. With wires the power that had to be controlled by the transmitter was only moderately higher; only enough higher to allow for losses by resistance in the wires. But with waves, as already noted, most of the power sent out by the transmitter did not even start toward any one receiver. So the power that had to be controlled by the transmitter was of a much higher order of magnitude with wireless waves than with currents on wires.

Of course this contrast was present in telegraphy as well as in any prospective telephony, but in telegraphy it presented no serious trouble. The telegrapher's key, which he worked with one finger in tapping out his dots and dashes, could not handle the heavy current needed in a wireless transmitter, but it could handle the current in a relay; that is, a solenoid which pulls an iron rod so as to work a bigger switch. (This, in fact, is how a little push button on the dashboard on a motorcar can replace a foot-operated switch for the starter.) So the high power at the transmitter could be controlled easily in dot-dash signalling.

Telephony, on the other hand, is not just a matter of on and off. Instead, it requires modulation (that is, changes without interruptions) in some characteristic of whatever is transmitted, with every change representing reasonably faithfully some change of pressure in the sound of the speaker's voice. The characteristic modulated in telephony by wires was the strength of the current in the wires. On the floor under a telephone in a house, at the turn of the 20th century, was a battery to drive this current. This would be steady direct current if the resistance of the circuit stayed constant. A large part of this resistance, however, was in contacts between a lot of little grains of carbon in the transmitter, just behind its diaphragm. This resistance changed in response to very gentle changes of pressure on these grains, because of sound waves in front of the diaphragm. A long time had elapsed since Alexander Graham Bell (1847–1922) had invented the telephone in 1876, and this transmitter was much better than his first one; yet the distance over which the telephone could be used was limited because the transmitter could not modulate the current satisfactorily if there was too much other resistance in the line. The essence of the limitation was a pair of requirements: (1) that any transmitter must be light enough to be worked by the weak pressures in the voice; and (2) that it would produce changes of current in proportion to these pressures. The first of these requirements might have been met with a relay, much lighter than those used in telegraphy; but merely switching the current on or off would not be modulating it in proportion to pressures.

What was needed, therefore, was something powerful that would be controlled by the telephone transmitter as the steering engine of an ocean liner is controlled by the light wheel that the helmsman handles—high power, that is, controlled by very low power, not in an on-and-off manner but under a rule of proportionality—and promptly, in response to the quick changes of pressure in sound waves. This became an actuality thanks to Lee De Forest (1873–1961). In 1906 he added a third electrode to the diode, or Fleming valve, making the triode, as described above under *Radio Tubes*. With this tube, which he called the audion, he invented several

circuits, one of which is shown in fig. 19, that are called amplifying circuits because they enable low-power apparatus to control high power in the way just specified. Amplifying circuits, therefore, made it possible for a telephone transmitter—or a microphone—to control the power sent out by a radio station.

Systems for doing this really well contain more than a single amplifying circuit and use tubes with more than three electrodes. Details on them would be out of place here, but a few facts may be pertinent. One of these is that the introduction of the third electrode enabled tubes not only to amplify power but also, in circuits derived from some of the amplifying circuits, to act as primary oscillators and supplant the older sources of high-frequency power.

Some other pertinent facts relate to the term modulation. This word is used most often in "frequency modulation," or "FM." But what was modulated first, and continued to be modulated in the output of most radio stations, is not the frequency of the waves. It is their strength. There is plenty of time during any one vibration of sound, even at a high pitch, for the station to send out many waves. Sending out 1,000,000 waves per second, for example, it sends out about 100 waves during each sound vibration at the highest pitch that most of us can hear. So the strength of these waves can be modulated during each sound vibration, with only very little change between any one wave and the next one; with much less change, in fact, than can be shown satisfactorily in a diagram. So, with the understanding that the diagram is not true to life in this respect, it is represented in fig. 43. In each graph time is shown by horizontal distance. The top graph represents a little more than one vibration of sound, with pressure as the ordinate. In the middle and bottom graphs the ordinates may represent either the current in an antenna (either transmitter or receiver) or the electric field strength at a receiver. The middle graph shows unmodulated waves, and the bottom graph shows waves modulated in obedience to the sound.

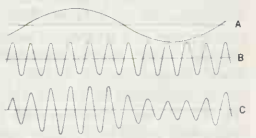


FIG. 43.—AMPLITUDE MODULATION: (A) MODULATING SIGNAL; (B) CARRIER; (C) RESULT OF ADDING (A) AND (B)

Modulation of this sort is called amplitude modulation because the term "amplitude" means the height of the peak of any wave and this height is modulated (along with any other measure of the strength of the waves) in obedience to the sound pressures.

Frequency modulation is not so easily shown on a diagram. As the name implies, it is modulation of the frequency of the waves in obedience to pressures caused by sound. For successful use of it we need many more waves in each period of vibration of the sound; so the frequencies used are roughly 100 times as great as with amplitude modulation. This causes a great difference in the interactions between the waves and gas ions that exist high up in the atmosphere, and thereby it causes the well-known difference in the ability of FM and AM to carry beyond the horizon.

For further details on this and other matters related to wireless telephony and broadcasting, *see* BROADCASTING; ELECTRICITY, ATMOSPHERIC; RADIO.

3. Television.—The principles of amplitude modulation discussed above are used in television also, and many other principles come in there along with them. So very many, in fact, and with so many technical complications, that only a few of the more obvious questions about television can be answered in this article.

The most obvious question, perhaps, is what produces the picture? Under *Masses of Electrons and Ions*, above, we saw that the picture is drawn by a beam of electrons, cathode rays, focused on the screen and deflected to various places on it, one after another, by electrostatic and magnetic forces of the types discussed under that heading. This statement covers the basic principles; the details are complicated and are different in receivers of different types. One more point, however, may be added here: the screen is covered not all at once but bit by bit; the focus of the beam of cathode rays moves from place to place on the screen in a manner called scanning, which is almost what the name implies. The focus does not jump, as if from word to word, along a line, but moves continuously along each line and jumps only in returning to start another line.

To keep the flickering of the picture down to a tolerable amount, it scans the whole screen in a thirtieth of a second. At this speed, with light and dark parts of the picture as close together as they are, the beam has to change its strength in extremely short times—to have its current modulated, that is, in times very much shorter than the periods of vibration of audible sounds. So the electromagnetic waves that transmit the signals for this modulation have to have frequencies far above those used in ordinary broadcasting. This is why television, like FM, does not carry well beyond the horizon.

Another obvious question about television is how the picture gets "on the air." Light carries it at first, from the actors or scene to be viewed; so the light is focused on a plate as in an ordinary photographic camera. This plate, however, is quite different. It is in a vacuum; and its surface is a rather complicated mosaic of little particles of a material that releases electrons wherever light strikes it. This surface is scanned, like the screen in the receiver, by a beam of cathode rays; and each particle in the mosaic accepts enough of the cathode-ray electrons, each time the beam gets to it, to make up for its losses under the influence of the light since the last time the beam was there.

It's a long way from the camera to the screen not only in distance but also in the tasks that have to be accomplished, first to synchronize the scanning in the receiver with the scanning in the camera and then to modulate the cathode-ray current in the receiver with the acceptance of electrons by particles in the mosaic in the camera. *See* also TELEVISION.

4. Radar.—Radar (*ra*dio *d*irection *a*nd *r*ange) is a radio apparatus that gives information on both the direction and the range of objects at a distance.

Direction alone would be too much to ask if the radio waves to be used were in the broadcast band. Waves tend to spread in a way much more obviously suited to broadcasting than to any directional task. If we were not accustomed to the wave theory of light we might well deny the possibility of getting directional information from waves. So it is this theory that gives the clue to such information. Light shows practically straight rays and practically perfect shadows only because its waves are of extremely short length from crest to crest. Wave lengths of visible light are only a few ten-thousandths of an inch. Rays and shadows are good because most things we see are so very big in comparison with these wave lengths. To "see" the direction of anything with radio waves, therefore, we need very short wave lengths.

Looking more closely at shadows, we can observe easily that they are not geometrically perfect. A little light bends into a shadow, even if the body casting the shadow is fairly large. With a small body, such as a piece of wire, and a very small source of light, it is easy to find streaks of light within the shadow and to connect them logically with the wave theory of light. Likewise a microscope cannot show perfectly sharp images, even if the shapes of its lenses are perfect, because the wave lengths of light are not infinitely small in comparison with its dimensions. These phenomena and many others like them show two qualitative aspects of directional effects in light waves: (1) no directional effect in waves is geometrically perfect; but (2) waves go fairly straight if their wave lengths are small in comparison with any obstacles or bodies directing the waves.

A radio searchlight cannot be perfect, therefore; but it can be fairly good if the wave length of the radio waves used in it is small in comparison with the concave mirror that is used for directing the waves. A wave length of only an inch or two, with a mirror a few feet in diameter, is a good combination.

It is easy to build the mirror, much easier for these waves than for light because this mirror does not have to be polished but merely made of fairly smooth metal. But to get much power at this wave length is not easy. No purely electrostatic tube, such as a triode, can work well unless an electron can get through its vacuum in a time small in comparison with the cycle. An electron, nimble as it is, cannot go so fast as the waves; so it cannot go

through a tube big enough to carry much power while the waves are going only a small fraction of an inch. The big problem of radar, therefore, was not the basic idea of using short waves but the development of a new type of tube (called the magnetron) to furnish the power. (*See* also RADAR.)

So much for the radio searchlight, and with it for the problem of direction; now for the problem of range. At a wave length of three centimetres, or 1.2 inches, 10,000 waves go out in a millionth of a second. This number is large enough to satisfy all needs, even those of resonant circuits that need many cycles in which to start resonating; and a millionth of a second is plenty of time for the current that feeds power to the magnetron to be turned on and off by tubes of more ordinary types. It is plenty of time also for the magnetron to get hot, so as to need a period of rest and cooling off. So the waves go out in flashes, each flash lasting only about a microsecond, with relatively long rest periods between flashes.

If such a flash is directed by the mirror at an airplane at a distance of 30 km., or 19 mi., it reaches the plane in a ten-thousandth of a second. Then some of its energy is reflected by the plane, and in another ten-thousandth of a second a little of this energy—very little, but enough—gets back to the radar apparatus. Two ten-thousandths of a second is very little time by some standards, but it is long in comparison with the duration of the flash. Electronic circuits can readily measure the time between outgoing and incoming flashes and direct a beam of cathode rays to an appropriate point on a screen so as to tell an observer how far away the plane is. This is how the radar answers the question of range.

In directing these waves with a concave mirror scientists went even further than Hertz had when he said, "Starting with purely electrical phenomena we have gone on step by step until we find ourselves in the region of purely optical phenomena." And in timing the return of the waves, which he so rightly called optical, with electrons to mark the time, science had come back to purely electrical phenomena in a way to reaffirm his statement about "the great domain of electricity," that it "has become a mighty kingdom."

D. Waves and Quanta

1. The Photoelectric Effect.—A need for some radically new laws appeared early in the 20th century on the extension of electrical theory to include light; and this need became especially urgent in relation to phenomena coming under the name "photoelectric effect." One of these, though not the easiest to measure, occurs in the television camera. There, as described above, the particles in the mosaic on its plate emit electrons when light falls on them. Electrons emitted in this way are called photoelectrons. In a vacuum tube designed for measurements on them, photoelectrons are found not to be merely released from the plate but to be ejected from it with positive, measurable kinetic energy. If X-rays are used instead of ordinary light, this energy is high; many X-ray photoelectrons move as if they had fallen through potential differences of tens of thousands of volts. Yet no law of waves could be expected to entitle them to any such energy.

On the contrary, such laws could be expected to deny them this energy. X-rays do not stop at the surface of a plate, even if the plate is made of metal. In fact their ability to penetrate into things is their most familiar property. So only a very small fraction of the energy of an X-ray beam is absorbed by atoms near enough to its surface to eject electrons from the metal. Moreover, the rays from an X-ray tube of any common type carry less energy altogether than the light from an ordinary lamp. With easily obtainable data one may calculate the energy that any one atom can be expected to absorb in any assigned time, or, conversely, the time required for any one atom to absorb any assigned amount of energy. This time, for the energy of one X-ray photoelectron, is many years. Yet photoelectrons jump out instantly, as soon as the rays strike the plate.

But they do not jump out from every atom; on the contrary, they jump out only from just enough atoms to make the energy of all the photoelectrons, as a group, agree with calculations in which proper deductions have been made for the work done in getting them out of the atoms. For the whole X-ray beam, and the whole group of atoms and photoelectrons, the law of conservation of energy still holds.

2. Quanta.—This example and others of its kind led some physicists to discard the wave theory and try their luck with a corpuscular theory; that is, to postulate that a beam of light is a shower of corpuscles, like little bullets except that they are composed of pure energy. (There is a semantic difficulty in postulating that something like a bullet is composed of pure "ability to do work," but perhaps that might be cleared up.)

These corpuscles were called quanta. The amount of energy in a quantum was not assumed to be the same for all kinds of light, but to vary from one colour to another in proportion to the frequency of vibration. X-rays, having frequencies about 10,000 times those of visible light, were given quanta 10,000 times as energetic. So this postulate went well with the high energy of an X-ray photoelectron. For some phenomena, especially the photoelectric effect and some closely related effects, this corpuscular theory seemed to do very well. Nevertheless it failed when applied to phenomena of the type that had led to the adoption of the wave theory a hundred years earlier. There is a great variety of these phenomena, called interference.

Moreover the corpuscular theory failed to account for the phenomena of refraction of light. On these, Maxwell had found some discrepancies between calculated and measured values, but he had expressed the hope that improvements in "our theories of the structure of bodies" would bring the calculated values into line with the measured. And so they did: because of refraction, as well as interference, the wave theory had to be retained.

It had to be retained even for X-rays, though these rays were the stronghold of the corpuscular theory. In the first decade of the 20th century many of the leading investigators of X-rays regarded them as corpuscular, though not as casting any doubt on the wave theory of light. Interference and refraction had not then been found in X-rays, though many tests for these phenomena had been made. This fact seemed to set X-rays apart from light; the X in their name had been put there to mean "the unknown," and it was still appropriate. Nevertheless the difference between light and X-rays is not actually in kind, but only in degree.

The reason for the failure of the tests for interference had been only that the apparatus was too big. Interference was found in X-rays in 1912, when Laue, Friedrich and Knipping used apparatus with one atomic dimension, the distance between layers of atoms in a crystal. Each of these layers treated X-rays much as each surface of a soap-bubble film treats white light when the soap bubble selects colours by interference. Wave lengths of visible light are very short (not far from thicknesses of soap-bubble films), and wave lengths of X-rays are still shorter, by a factor of about 10,000. For refraction the corresponding factor is about 1,000,000. These numbers certainly show that the differences in degree are very great, but they do not indicate any differences in kind. X-rays are a kind of light, after all; so they must be given a place in the scale, along with light and radio waves.

Recalling how Hertz's waves were located in this scale, between 18 and 23 octaves below red light, and how violet light is only about one octave above red, we may now give X-rays their place: from 10 to 14 octaves above violet light. In this scale, therefore, Hertz's waves are even farther than X-rays from visible light.

In visible light the phenomena that suggest a corpuscular theory are not at all conspicuous. They are not absent, to be sure, so they must be covered by the theory. To cover them the wave theory has to be supplemented with new laws. The wave theory requires the electric field in a light wave to be nearly uniform over thousands of atoms which therefore must treat their electrons alike. Nevertheless, only a minute fraction of these atoms, if any, eject photoelectrons. Evidently the electric field does not tell every electron positively and unequivocally to jump out as a photoelectron. Instead, the field must deal with the electrons in terms of mathematical probabilities; it must give each electron a certain probability of suddenly acquiring a quantum of energy. (*See* QUANTUM MECHANICS.)

A change from visible light to X-rays makes a 10,000-fold increase in the quantum of energy, and therefore a corresponding

reduction in any one electron's probability of acquiring a quantum. A change from light to Hertzian waves has the opposite effects; in these waves no electron does anything very violent, and the probability of its doing what it does amounts almost to a certainty.

Quanta and their probabilities play practically no part, therefore, in the theory of radio waves, even those that we call microwaves. Neglecting quanta we have what is called classical electromagnetic theory; and as stated under *The Accuracy of the Fundamental Laws*, above, we apply this theory, within its proper boundaries, as if its accuracy were strictly perfect.

See ELECTROMAGNETIC WAVES; *see* also references under "Electricity" in the Index volume.

BIBLIOGRAPHY.—M. Abraham, *The Classical Theory of Electricity and Magnetism*, rev. by R. Becker, tr. by J. Dougall (1932); J. A. Stratton, *Electromagnetic Theory* (1941); W. R. Smythe, *Static and Dynamic Electricity* (1939); G. P. Harnwell, *Principles of Electricity and Electromagnetism* (1938); L. Page and N. I. Adams, *Principles of Electricity*, 2nd ed. (1949); J. C. Slater and N. H. Frank, *Electromagnetism* (1947); M. Mason and W. Weaver, *The Electromagnetic Field* (1929); H. H. Skilling, *Fundamentals of Electric Waves*, 2nd ed. (1948); S. Ramo and J. R. Whinnery, *Fields and Waves in Modern Radio* (1944); F. E. Terman, *Radio Engineering*, 2nd ed. (1937); R. M. Bozorth, *Ferromagnetism* (1951); R. A. Millikan, *The Electron*, 2nd ed. (1924); K. K. Darrow, *Introduction to Contemporary Physics*, 2nd ed. (1939); F. K. Richtmyer, *Introduction to Modern Physics* (1928); H. H. Skilling, *Exploring Electricity* (1948); H. A. Lorentz, *The Theory of Electrons* (1909); J. C. Maxwell, *A Treatise on Electricity and Magnetism* (1873); H. R. Hertz, *Miscellaneous Papers*, tr. by D. E. Jones and G. H. Schott (1896). (D. L. WR.)

ELECTRICITY, ATMOSPHERIC. Atmospheric electricity is that branch of the physics of the earth which deals with electric fields, currents and charges in the atmosphere, with atmospheric conductivity, and with atmospheric processes which produce electrical effects. Traditionally, the workers in the field of atmospheric electricity have confined their attention to electrical phenomena between the surface of the earth and the ionosphere, that region of the earth's atmosphere above 40 mi. responsible for the reflection of radio waves. Electrical processes within the ionosphere have been investigated as branches of other geophysical sciences.

This article deals with the general aspects of atmospheric electricity. After a general survey and a historical account of the investigation of atmospheric electricity, there is a detailed discussion of problems and results of study of the atmosphere as a conducting medium, of the distribution of electric fields and currents in the atmosphere, and of the physical processes which may produce electrical effects. For information concerning extensive special investigations of the lightning discharge, *see* the article LIGHTNING.

GENERAL SURVEY

Scientific studies of electrical phenomena in the earth's atmosphere have been carried on since the first half of the 18th century. During this period, and particularly during the 20th century, electrical measurements have been made at the earth's surface on every continent and every ocean at many latitudes from the polar regions to the tropics. Balloons and aircraft have been extensively employed for making electrical observations in the free atmosphere from sea level to 30 km. (100,000 ft.). The various electrical investigations have encompassed all types of meteorological conditions from clear, calm weather to violent storm. In spite of the extensive nature of atmospheric electrical data only a few broad generalizations have emerged, a result due, in part, to the great complexity of the atmosphere, and, in part, to the lack of experimental controls which have proved so fruitful in laboratory studies of physical phenomena. It is not the purpose of this article to provide a comprehensive catalogue of the wide variety of observations but rather to select those observations which provide a basis for understanding the nature of the atmosphere as an electrical medium, the atmospheric processes which produce electrical effects and the characteristic distributions of electrical fields, currents and charges.

The lower atmosphere is a very poor conductor of electricity but it is nevertheless rendered slightly conducting because of its ionization by cosmic rays. Only in a shallow layer of air over the continents does the ionization due to radioactive materials from the earth's crust exceed that due to cosmic rays. Above an altitude of about 70 km. copious ionization is produced by solar ultraviolet and soft X-radiation and, to a lesser extent, by meteors and solar particle beams. The highly ionized region of the upper atmosphere, appropriately called the ionosphere, and the surface of the earth are both enormously better conductors than the lower atmosphere and may, for the purpose of atmospheric electrical studies, be considered as regions of uniform potential. Thus the earth's atmosphere tends to resemble a large spherical condenser with the ionosphere and earth serving as the upper and lower plates, respectively, and with the atmosphere between as a leaky dielectric. Since the earth and ionosphere are effective electrostatic shields, the electrical phenomena in between must be produced primarily by electrical generating processes which lie within the atmosphere itself and which derive their energy directly or indirectly from atmospheric motions.

Atmospheric electrical phenomena can be divided for convenience into those associated with fair weather and those associated with storms. The area covered by storms at any time is a small fraction of the total area of the earth, hence the fair-weather phenomena may be considered as normal even though electrical phenomena of disturbed weather are of great importance. The fair-weather electrical field is directed toward the earth and has an average magnitude of 130 v. per metre at the surface. The vertical, electric conduction current associated with the fair-weather field has a mean density of 3×10^{-12} amp. per square metre, and the total current flowing to the earth over all regions of fair weather is approximately 1,500 amp. Since the effective resistance of the whole atmosphere to vertical current flow is 200 ohms, the potential of the ionosphere is about 300,000 v. higher than the earth. The source of the continuous air-earth current was long the central problem of the atmospheric electricity. However, a series of investigations carried out during the second quarter of the 20th century demonstrated that the strength of the current at any hour of the day was nearly proportional to the total number of thunderstorms in progress, that thunderstorms produced conduction currents which proceeded upward toward the ionosphere, and that the estimated average magnitude of the total upward current from all thunderstorms in progress at any time was roughly equal to the total downward air-earth current. In the absence of any alternate explanation with comparable supporting evidence from observation, it is generally agreed that thunderstorms serve as generators for an upward current which spreads laterally near the base of the ionosphere, descends to the earth in all regions remote from storm and returns through the earth to the storm regions. This current system accounts for the normal electric field of the earth and the potential difference between earth and ionosphere.

The conduction currents in the air, such as the air-earth current, are carried by molecular ions which drift through the air in the presence of an electrical field. There are, in addition, electrical convection currents in which charge is transported by the motion of the air itself or by the motion of precipitation forms falling through the air under the influence of gravity. The convection currents, particularly those in clouds and storms, are of basic importance since they provide the means of separating charges of opposite signs within the air, thus creating atmospheric electrical fields and establishing conduction currents. They may be described appropriately as electrical generating processes.

A majority of electrical generators in the air produce only local effects. The generators exist primarily in the lower troposphere, close to the lower plate of the spherical atmospheric condenser, and do not appear to produce significant fields in the higher atmosphere. The thunderstorm (*q.v.*), the largest and most spectacular of the atmospheric electrical generators, appears to be the major exception to this statement. The main charges generated in the thunderstorm are usually a few tens of coulombs. These charges are found at altitudes where the temperature is 0° C. or below with the positive charge characteristically at greater elevation than the negative charge.

The magnitude of the convection current necessary to build up

and maintain the charges in the presence of conduction losses is estimated to be a few amperes. The potentials of the charges with respect to ground are of the order of 10^8 to 10^9 v., so that the potential of the uppermost positive charge is much higher than the potential of the ionosphere. The resistance of the conducting paths from the positive charge to the ionosphere is substantially less than the resistance of the conducting paths to the earth, a situation favourable for supplying a current to the high atmosphere. The average upward conduction current over a single storm is of the order of one ampere.

While the main features of the fair-weather field appear to be understood, there remain many important unsolved problems concerning the electrical phenomena in storms and the character of the charge-generating processes.

HISTORY

18th Century.—Even primitive man was impressed by Jupiter's thunderbolts but the first serious studies of atmospheric electric phenomena were made in 1746–52 by Benjamin Franklin and some of his Philadelphia friends. Their experiments dealt chiefly with the very intense and sporadic electrical effects accompanying storms and these culminated in the invention of the lightning rod. So great was the interest in this device that many European scientists were immediately stimulated to begin experiments with electrical phenomena in the atmosphere. As early as 1752 P. C. Lemonnier in France discovered that a "state of electricity" exists in the atmosphere even in fair weather and in a series of systematic observations prior to 1775 Giovanni Battista Beccaria in Italy verified Lemonnier's conclusions and discovered a diurnal variation of the fair-weather electric field.

In Switzerland Horace Benedict de Saussure began experiments in 1766 with what was perhaps the first electrometer and he was the first to report an annual variation in the fair-weather field. His method of measuring the field consisted of raising a mobile conductor away from the earth by a standard distance and noting the resulting change in its potential by means of an electrometer connected to it.

Franklin had collected charge from the atmosphere by means of a sharp point and during storm conditions had found that the charges were sometimes positive and sometimes negative. De Saussure and Beccaria, on the other hand, found that charges collected in the same way during fair weather were always positive. These facts remained unexplained, however, until, toward the end of the 19th century, the ionic conduction of electricity through gases became understood and it was realized that the effect of the sharp point was to increase the field strength in its vicinity and thus to accelerate the motion of the ions in the air. Under normal conditions the sharp point if connected to the ground would produce an attracting field for the positive ions.

19th Century.—The idea of electrostatic induction was introduced in 1840 by the Frenchman J. C. A. Peltier to explain the action of the mobile conductor. When an object such as a metal sphere is in contact with the earth and is in a shielded box it will have no charge induced upon it. If it is insulated and then raised out of its shield to a point above the earth, a positive charge will be induced on its lower surface and an equal negative charge on its upper surface by the negative charge on the earth. An electrometer connected between the earth and the sphere will indicate the tendency of positive charge to flow toward the earth.

New interest in the investigation of atmospheric electric phenomena was aroused in 1850 when Sir William Thomson (later Lord Kelvin) put forth an interpretation of all known facts on the basis of a sound theory of electrostatics involving the concept of potential. He also developed new instruments and inaugurated programs of continuous recordings at the Kew observatory in England. With portable instruments he carried out measurements at a number of other places on the earth's surface and standardized his readings for the first time in terms of the potential of the Daniell cell. At the International Electrical congress in 1883 Kelvin and Henry A. Rowland introduced important resolutions on atmospheric electric investigations and shortly thereafter work was begun by many other physicists. Air potentials were recorded continuously at several land stations and observations were made at sea. Observers in Austria, France and Germany also used balloons to explore the electric fields aloft. As a result of this burst of activity it was soon discovered that the atmosphere was positive with respect to the earth at such widely distributed places that it could be regarded as a world-wide characteristic. The electric field near the surface was shown to be greater in the winter than in the summer and a regular diurnal variation in field strength was found at all stations, although the phase and amplitude of this variation changed from place to place in a puzzling manner. The balloon observations showed a rapid decrease in field strength with height above the ground and from this it was concluded that the lower atmosphere has a positive space charge. It was also realized during this period that the earth has a negative charge, as De Saussure and Peltier had postulated, and that at least a part of the lines of force from its surface charge terminate on the positive charges in the lower atmosphere. At first these positive charges were erroneously regarded as stationary in the atmosphere in spite of the fact that Charles A. de Coulomb had concluded in 1785 that the air is not a perfect insulator.

The first realization of the existence of an atmospheric electric current came in 1887 as the result of the work of W. Linss, a German, who, following the method of Coulomb, observed the rate at which a charge on an insulated conductor is dissipated when exposed to the free atmosphere. His studies revealed a diurnal and an annual variation of the conductivity, and Julius Elster and Hans F. Geitel showed that the conductivity thus measured was inversely proportional to the field strength as would be expected if a constant air-to-earth current were flowing. Near the end of the 19th century Elster and Geitel made more extensive measurements of the rate of dissipation of charge from insulated conductors and showed that fog or smoke reduces the conductivity of the atmosphere, a fact which is explained by the attachment of the small ions to the smoke particles or fog droplets with a consequent loss of mobility. They also found that negatively charged bodies lose their charge more rapidly than positively charged bodies, especially in the pure air of the mountainous regions where positive space charge is more in evidence. But the greatest contribution of these investigators was to give a full explanation of the conductivity of the atmosphere in terms of Sir J. J. Thomson's theory of ions in gases and thus to open up the modern epoch in the study of atmospheric electricity.

20th Century.—With the realization that a current must be flowing continuously to the earth under the influence of the ever-present electric field, attention began to focus on the problem of the supply current required to maintain the negative charge of the earth against its constant dissipation by the air-to-earth current. When extensive measurements were made in regions remote from sources of pollution, over the surface of the oceans and in the polar regions, it was noted that the electric field executed a simple diurnal oscillation. In 1923 K. Hoffman and S. J. Mauchly independently pointed out that the maximum and minimum values of the field strength occurred at about the same hours, universal time, independent of local time. Following a suggestion by C. T. R. Wilson, F. J. W. Whipple demonstrated that there was a correlation between the number of thunderstorms in progress over the world and the strength of the fair-weather field. This presumptive evidence that thunderstorms were responsible for the fair-weather field and the air-to-earth current was supported in 1950 by O. H. Gish and G. R. Wait, who measured the current flowing upward from the tops of the thunderstorms. Their average value of the current from a single thunderstorm cell together with a reasonable estimate of the number of thunderstorms in progress at any time yields a total upward current in approximate agreement with the total downward current in regions remote from storm.

During the 20th century the thunderstorm has been the subject of many investigations because of its importance to atmospheric electrical problems. Between 1915 and 1920 C. T. R. Wilson observed rapid field changes associated with lightning discharges and concluded that the thundercloud was essentially bipolar with the positive charge uppermost. Sir George Simpson found evidence for a small positive charge near the base of thunderstorm

clouds and at first maintained that the polarity of the storm was the reverse of that found by Wilson. Later work by Simpson, B. F. J. Schonland and others has demonstrated that the Wilson polarity is correct for the main storm charges and that the positive charge at the cloud base is smaller and probably caused by a different process than the one responsible for the main charges higher in the cloud. Much work has also been done on the magnitude of the thunderstorm charges, the lightning flash and the physical process responsible for production of charges.

The military requirements of World War II accelerated the pace of atmospheric electrical investigations. Among the problems studied were electrical hazards to aircraft in thunderstorms and "precipitation static," the interference to radio communications on airplanes flying through certain types of clouds.

In the postwar period the availability of aircraft, high-level balloons and rockets provided the incentive for extensive electrical measurements in the free air and the use of newer electronic techniques significantly improved instrumentation.

The search for the fundamental mechanisms of charge separation in precipitating clouds, particularly thunderstorms, continued to be one of the most important lines of investigation. Studies related to this problem included both measurements in the field and laboratory experiments on the production of charge on ice and water.

The growing recognition of the fact that atmospheric electricity is a sensitive indicator of some meteorological phenomena has encouraged a number of studies designed to identify local atmospheric electrical effects with specific meteorological conditions. Thus the local effects, which retarded the progress of early investigators in their efforts to explain the fair-weather field, have become a subject of interest in themselves.

PROBLEMS AND RESULTS

In the following section all electrical quantities are given in the rationalized metre-kilogram-second (or Giorgi) system of units (*see* PHYSICAL UNITS). The permittivity of free space, ϵ_0, in these units has the value of 8.85×10^{-12} farads per metre. In most problems the value of ϵ for air is essentially the same as ϵ_0.

Ionization of the Atmosphere.—Any high-speed electrically charged particle passing through the atmosphere will produce ionization. Rays of this type occurring naturally include the alpha and beta rays from radioactive elements in the soil and air, electrons ejected from air molecules by gamma rays from the same sources, and cosmic rays. Since 1945 ionization from naturally occurring radioactive substances has been augmented slightly, although not uniformly over the earth, by the radioactive isotopes from large nuclear explosions. The intensity of ionizing radiations is usually measured in terms of the number of ion pairs produced per cubic centimetre per second in a standard gas such as air at standard density. Unit intensity, one ion pair per cubic centimetre per second, is represented by the symbol I. The ionization over land surfaces varies with the concentration of radioactive materials. At the surface the average value is estimated at 9.5 I, one-half from radioactive matter in the air, one-third from radioactive matter in the soil and one-sixth from cosmic rays. Over the surface of the oceans, about 95% of the ionization is due to cosmic rays and the larger part of the remaining 5% is believed to be due to radioactive materials of land origin.

The radioactive atoms in the air, principally radon, thoron and their decay products, enter the atmosphere through pores in the soil. These substances are carried upward from surfaces by eddy diffusion, but as they become ionized during successive decay processes they probably serve as condensation centres for moisture and are precipitated to the earth. In any case, the concentration of radioactive materials decreases rapidly with altitude. Above the exchange layer, the turbulent atmospheric layer nearest the surface, the effect of natural radioactive materials becomes very small by comparison with cosmic rays. Thus, throughout most of the atmosphere below the ionosphere cosmic rays are the dominant cause of atmospheric ionization.

Cosmic-ray primaries are principally high-energy protons, the nuclei of hydrogen atoms and, to a small extent, the nuclei of heavier atoms. As they approach the earth they are deflected by the earth's magnetic field so that only the more energetic reach the earth near the equator while the less energetic enter the atmosphere at higher geomagnetic latitudes. The cosmic-ray primaries suffer collisions with atmospheric atomic nuclei producing a complex of secondary particles principally responsible for cosmic-ray ionization. The production of secondaries is infrequent in the high atmosphere because of the low atmospheric density. Production increases as the primaries descend into the denser atmosphere and once more falls off as the rays approach the earth because of their absorption in the atmosphere. Cosmic-ray ionization at the surface of the earth produced by the deeply penetrating hard component varies only slightly with latitude, being about 1.5 I in equatorial regions and 1.8 I in the higher latitudes. The ionization increases with altitude, reaching a maximum near 13 km. where about 45 ion pairs are produced per cubic centimetre per second at high latitudes and about 20 at the equator. The marked increase in ionization with latitude at 13 km. is associated with soft component which is more easily deflected in the magnetic field and which does not penetrate to the surface of the earth.

The rate of ionization of the atmosphere by short-wave-length radiation from the sun above 70 km. is several orders by magnitude greater than that due to cosmic rays and radioactivity in the lower atmosphere but precise values are of little significance in the problems of the lower atmosphere.

The radioactive material from large nuclear explosions did not produce any important world-wide effects in atmospheric ionization during the decade following the first explosion. The ionization rates in radioactive clouds within a few days after the explosion are much greater than normal. As radioactive decay and atmospheric diffusion reduce concentrations, the effect on ion density ceases to be easily detectable. There have been a number of reported cases in which the ionization rate at the surface has increased several fold following a heavy rainstorm in air carried away from a large explosion. In these cases the precipitation is effective in sweeping out a large volume of air and concentrating the radioactive material on the surface.

Ionic Motions and Conductivity.—The primary ionization process discussed in the preceding section involves the removal of an electron from a neutral molecule, thus creating positive ion and a free electron. In the lower atmosphere the electron attaches to an oxygen molecule creating a negative molecular ion within less than a microsecond. Accordingly, the number of free electrons in this part of the atmosphere is negligible.

Experimental evidence derived from the study of ionic motions indicates that atmospheric ions may form clusters by attaching a small number of neutral molecules. The group of individual and cluster ions are usually referred to as "small" ions. In the presence of minute suspended particles of salt, dust, water or other substances, the small ions may attach to suspended particles forming large or Langevin ions.

All ions partake of the random molecular motion and, if an electric field is established in the air, it superimposes on the random motion a drift velocity in the direction of the field if the charge is positive and counter to the field if negative. The average ionic drift velocity in a unit field is called the mobility, k; thus, the ionic velocity may be written kE where E is the strength of the field. The current density, i, defined as the charge which flows in one second across a unit area perpendicular to the direction of flow, depends on the number of ions per unit volume, n, the charge on each ion, e, and the mean ionic velocity. The current produced by the motion of one type of ion is, $i=nekE$. If there are several groups of ions present, the total current density will be the sum of their separate contributions, $i=\sum_r n_r e_r k_r E$.

Finally, since the conductivity, λ, is defined as the current density per unit field, $\lambda = i/E = \sum_r n_r e_r k_r$. Because most of the ions have a single electronic charge, the conductivity of the air varies with the number of ions of each type per unit volume and their respective mobilities.

Because of their relatively high mobilities only the small ions

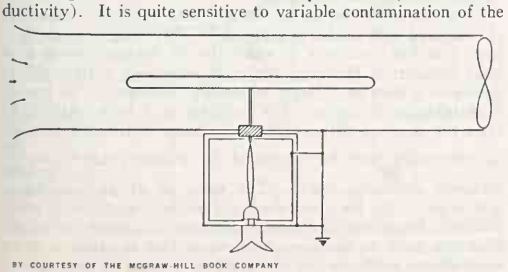
make a significant contribution to atmospheric conductivity so that practically $\lambda = n_1 e k_1 + n_2 e k_2$ where the subscripts refer to positive and negative small ions, respectively. The two expressions on the right are called the polar conductivities.

The mobilities of atmospheric ions differ somewhat from one measurement to the next, presumably depending upon the precise atmospheric composition and the age of the ions. The observed mobility of small ions lies between 1 and 2×10^{-4} m. per second per volt per metre, the mobility of the negative ions characteristically exceeding that of the positive ions. That is, in normal fields of 100 v. per metre the small ions drift at speeds of one to two centimetres per second. The mobilities of large ions are about 500 times less than those of small ions. Some observers have found groups of intermediate ions with mobilities about 100 times smaller than those for small ions. Above the surface the mobilities of all ions increase as the density of the air diminishes.

The number of any ion group per unit volume under equilibrium conditions depends upon a balance between their rate of formation and their rate of destruction. Well above the surface layers where the density of particulate matter is small the number of ions produced in each unit volume per unit time, q, is equal to their rate of destruction by recombination with each other, $\alpha n_1 n_2$, where α is the recombination coefficient. Since n_1 is nearly equal to n_2 at altitudes well above the surface, $q = \alpha n^2$ and $n = \sqrt{q/\alpha}$. In the lower layers where particle concentrations are high, the rate of loss of small ions due to attachment to neutral particles or recombinations with large ions may greatly exceed αn^2. The equilibrium equations are somewhat complex and for the present discussion it is sufficient to note that an increase in particle concentration reduces the concentration of small ions and, hence, the conductivity. Near the surface of the earth n is highly variable and even average values range from about 200 ions per cubic centimetre in large industrial cities to as much as 1,000 per cubic centimetre in open country under favourable meteorological conditions. Above the exchange layer the particle concentration is usually so low that computed values of conductivity are in good agreement with observations if it is assumed that q is due solely to cosmic radiation and if due account is taken of the increase of k and decrease of α with altitude.

Conductivity Observations.—Atmospheric conductivity is usually measured with equipment basically similar to that of H. Gerdien shown in fig. 1. Air is drawn continuously through a pair of coaxial cylinders insulated from each other so that a potential difference may be established between them. If the ion current to the inner cylinder, I, is measured in that range of air velocities in which the current is proportional to the potential difference, V, the conductivity is $\lambda = \epsilon I/CV$ where C is the capacitance of the coaxial cylinders. Each of the polar conductivities is measured separately. Such equipment has been used over land and sea surfaces, in airplanes and in balloons to heights of 30 km. Conductivity may also be measured by observing the rate at which charge leaks off of an insulated conductor exposed to the atmosphere. The time, τ, required for the charge to diminish to $1/e$ of its initial value, where e is the base of the natural logarithms, is given by $\tau = \epsilon/\lambda$.

The conductivity measured over land surfaces near sea level has an average value of 1.8×10^{-14} mhos per metre (units of conductivity). It is quite sensitive to variable contamination of the atmosphere, and near cities and inhabited places the variation of conductivity is often determined largely by the pattern of human activity. In the vicinity of large industrial cities it is much lower than the general average; for example, at Kew observatory, near London, it has a mean value of only 3×10^{-15} mhos per metre. Remote from habitation and sources of industrial pollution, the conductivity executes a relatively simple daily oscillation with a maximum value near dawn and a minimum value in mid-afternoon. The oscillation is controlled by the concentration of particulate matter in the air and the distribution of radioactive gases, which are in turn affected by the local daily cycle of turbulence.



BY COURTESY OF THE MCGRAW-HILL BOOK COMPANY

FIG. 1.—SKETCH OF GERDIEN CONDUCTIVITY APPARATUS

The conductivity over sea surfaces has a mean value of 2.8×10^{-14} mhos per metre, surprisingly close to the average for land. This is due to the fact that while the particle concentration is much lower over the ocean so also is the ionization rate. In contrast with observations over land, the diurnal variation of conductivity is small, usually about 4%. Larger variations of conductivity occur in regions within a few hundred kilometres of continents when swept by offshore winds. The variation of conductivity with latitude can be observed more easily over the ocean surface than over land. Consistent with variation of cosmic-ray intensity, the conductivity is about 20% lower in the tropics than in higher latitudes.

In any discussion of the variation of λ with altitude it is convenient to divide the atmosphere into the exchange layer and the region above. The exchange layer commonly extends from the surface of the earth to an altitude usually between one and three kilometres depending on meteorological conditions and surface heating. It is characterized by a high particle concentration which usually decreases sharply at the top of the layer. Within the exchange layer the conductivity may exhibit relatively large fluctuations from place to place and with time. In the average λ increases with elevation through the layer, although the fluctuations in any one set of observations is large. At the top of the layer λ commonly increases by 50% or more in a rise of 100 m. Above the layer λ is subject to small, relatively slow fluctuations and increases monotonically with altitude. Above the exchange layer λ increases rapidly with altitude. At 5, 18 and 30 km. respectively, the conductivity is 10, 100 and nearly 500 times the value at the surface of the earth. In one rocket flight Gerdien-type instruments were used to measure conductivity between 30 km. and the base of the ionosphere. The increase of conductivity was approximately that expected, assuming ionization to be produced solely by cosmic rays. Above 50 km., there was some evidence of a small concentration of particulate matter, possibly meteoric dust.

At the surface of the earth the positive conductivity exceeds the negative conductivity by 10% to 20% due to the electrode effect, described later. In the high atmosphere the negative conductivity would be expected to exceed the positive conductivity by 20% because of the higher mobility of negative ions. This excess has been reported by some observers but others find the positive and negative conductivities to be nearly equal.

The values of λ given above apply to fair weather but have been found to be substantially the same in the vicinity of thunderstorms when the measurement is made outside the thundercloud itself. Measurements within clouds are difficult and uncertain; however, the existing observations suggest that λ within clouds is about three times smaller than outside clouds at the same altitude.

Measurement of Electric Fields.—The characteristics of the atmosphere as a conducting medium have been briefly outlined in the preceding sections. The actual conduction currents flowing in the medium depend upon the electric fields established. The electric field intensity, or its negative, the potential gradient, has been observed more extensively than any other atmosphere electrical variable. A very large number of devices have been used to measure the field but the most common equipment may be divided into two types: those employing probes or collectors and those depending upon induction of charge.

In the first type of measurement an insulated metal probe is located at a convenient position in the atmosphere. If its potential initially differs from the surrounding atmosphere, it will collect

ions of the appropriate sign until it reaches the air potential. Because of the low conductivity of the air, the process may require many minutes and if the field varies rapidly, the probe will generally have a potential different from the surrounding air. One of the methods employed to bring the probe quickly to air potential is to place a small amount of radioactive material such as polonium on its surface so that the surrounding air is strongly ionized. Such a probe is called a collector. The measurement of the potential difference between ground and a collector at known height above ground with an electrometer serves to determine the average value of the potential gradient or field intensity in the height interval above ground. The measurement of the potential difference between two collectors suspended from a balloon at constant vertical separation can be used to determine the potential gradient in the free atmosphere.

The several induction instruments for measuring field intensity differ in details of construction and operation but all depend upon the basic fact that when a conductor is placed in an electric field, the density of charge induced on its surface at any point is proportional to the field intensity at that point, that is, $\sigma = \epsilon E$ coulombs per square metre. One very simple induction instrument will serve as an illustration. An insulated plate may be mounted flush with the surface of the earth and provided with a grounded shield plate which may be placed directly above it or removed. When the insulated plate is exposed to the atmospheric electric field and grounded, a charge of magnitude $Q=\sigma A=\epsilon EA$ coulombs will be induced on its exposed surface of A sq.m. The plate is then insulated and shielded. If the potential difference between the grounded shield and insulated plate measured by an electrometer is V, then $V=Q/C$ where C is the capacitance to ground of the insulated plate. Thus, $E=CV/\epsilon A$. In one commonly used form of the induction device called the field mill, several insulated plates are mounted with their centres on a horizontal circle and the shields in a form resembling a windmill are rapidly rotated above them. Induction devices like the field mill are convenient to use both on the ground and on the surfaces of airplanes.

The device shown in fig. 2 may be used for measuring field as described above or it may be used for direct air-earth current measurement if the uncharged insulated plate is exposed to the normal air-earth current for a definite time while its potential is kept at ground by means of a capacitance-coupled potentiometer. At the end of the interval, collected charge is measured as described above and the average current during the interval is computed.

The so-called "point discharge" is also used for measuring electric field. It is well known that corona discharges develop in sufficiently high fields at surfaces of high curvature. The current from an exposed point is a quadratic function of the intensity of the field. Thus, the current which passes through a point indicates both the direction and magnitude of the field after proper calibration. The advantage of the method is the simplicity of the equipment needed.

Electric Fields and Conduction Currents.—The conduction current density at any position in the atmosphere may be computed as the product of the measured values of the electric field intensity and conductivity. The study of the distribution of conduction currents within the atmosphere is, in general, very complex unless the electric field is constant or changing slowly with time and the conduction current is correspondingly steady. When a steady state condition exists and the field is suddenly changed, it requires an appreciable time for the conduction current system to approach a new steady state value. The relaxation time of the atmosphere, τ (see *Conductivity Observations*, above), is a measure of this period of readjustment. Only when significant field changes occur in intervals substantially greater than τ can it be assumed that a steady state is approximated. The value of τ usually lies between a few minutes and half an hour in surface layers of the atmosphere but decreases with altitude as the conductivity increases and achieves a value of about four seconds at 20 km. Fortunately, from the viewpoint of simplicity in calculation, the important case of the normal or fair-weather field can be treated approximately as a steady state condition. This problem will be examined in some detail.

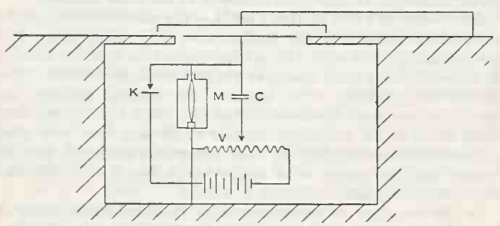


FIG. 2.—C. T. R. WILSON'S TEST-PLATE APPARATUS FOR MEASURING POTENTIAL GRADIENT AND CURRENT DENSITY

The average value of the fair-weather field intensity, E, at the surface of the earth is 130 v. per metre. It is directed toward the earth implying a negative charge density, σ, of 10^{-9} coulombs per square metre ($\sigma=\epsilon E$). The total negative charge on the surface in regions of fair weather is thus about 500,000 coulombs. There is no estimate of comparable accuracy for the surface charge under stormy regions where the field intensity deviates markedly from its fair-weather value but it appears probable that the total surface charge under storms is much less than the charge in regions of fair weather. Accordingly, the 500,000 coulombs is often referred to as the earth's negative charge.

The air-earth conduction current which flows in the presence of the fair-weather field has an average density of 3×10^{-12} amp. per square metre (3 microamperes per square kilometre) and the total current flowing to the surface in all regions of fair weather is about 1,500 amp. It is evident that an electrical generating process must be acting continuously to maintain the charge on the earth's surface and the approximately steady air-to-earth current. In the absence of such a generator, the fair-weather field would disappear in a matter of minutes.

As far upward as measurements have been made, the field intensity, E, decreases with altitude, varying nearly inversely with the conductivity. Thus, the conduction current density, $i(=\lambda E)$, must be substantially independent of altitude in the lower atmosphere and flowing vertically downward from the highly conducting regions of the upper atmosphere. The potential difference between the surface of the earth and any point in the atmosphere can be calculated by an application of Georg S. Ohm's law. Consider a vertical atmospheric column one square metre in cross section extending from the surface of the earth to any height z in the atmosphere. The resistance of such a column is $R_z=\int_0^z dz/\lambda$ and the potential at z is $V_z=iR_z$. When z is the altitude of the ionosphere the total resistance of the column (the columnar resistance) is about 10^{17} ohms, hence the potential difference between the earth and ionosphere is approximately 300,000 v. It is interesting to note that a major part of both the difference of potential and the columnar resistance is found in the first few kilometres above the surface. At 10 km. the potential is already about 90% of the ionospheric potential and above about 40 or 50 km. the atmosphere may be considered an equipotential region in atmospheric electrical problems.

Since the earth and ionosphere are both excellent conductors as compared with the lower atmosphere, they tend to resemble a large spherical condenser in which the air between serves as a leaky dielectric. However, the earth-ionosphere system differs significantly from an ordinary laboratory condenser in the matter of distribution of charge. The variation of E with height indicates the presence of positive space charge distributed through the air which may be computed by Poisson's law: $\rho=\epsilon\frac{dE}{dz}$ coulombs per cubic metre. The value of dE/dz, and hence ρ, is largest in the low atmosphere and becomes negligible at great altitude. Integration of Poisson's equation with respect to height from the earth to the ionosphere shows that the total positive space charge in the air (in fair weather) is equal to the negative charge on the surface of the earth. There is no charge on the

inner surface of the ionosphere and most of the positive charge resides in the lowest layers of the atmosphere. Further, since the ionosphere and the earth are effective electrostatic shields the total net charge on their inner surfaces and within the atmosphere must be zero at all times. A second consequence of the shielding effect of the ionosphere is that any measurement made beneath the ionosphere can give no information concerning a net charge on the earth as a whole. Whether or not a charge resides on the outer ionosphere can be determined only by measurements outside the earth's atmosphere.

In the absence of any local generators of electric fields, the current density, i, flowing to any point on the earth's surface in fair-weather regions is determined by the potential difference, V, between the earth and the highly conducting layers of the upper atmosphere and the columnar resistance, R, through the relation $i=V/R$. Over the oceans, on the polar icecaps and on mountaintops (especially in winter) the current density shows the same type of daily variation when measured on universal (Greenwich) time rather than local time. This implies that the observed variation in i is proportional to V and that R is essentially constant in time at each observing station, but not necessarily the same at all observing stations. Over the oceans where λ is nearly constant in time, the electric field intensity or potential gradient is also proportional to V. $E=V/\lambda R$. Fig. 3 shows the average variation of the potential gradient over the oceans and, hence, presumably the average of the relative daily variation of V.

The situation over land surfaces in temperate and tropical climates is complicated by important local variations. As noted above a large fraction of the columnar resistance lies in the lowest portion of the atmospheric column which is most susceptible to change. To the extent that the resistance particularly in the exchange layer is controlled by the local cycle of turbulence, R tends to be least around dawn and greatest in the afternoon. In addition, R is affected by local meteorological conditions, such as the character of the air mass which overlays the observing station. Variations in R superimpose a local effect upon the world-wide diurnal pattern of i and at times may be sufficiently large to mask the world-wide effect completely.

The effect of nonprecipitating clouds may be considered as a perturbation of the fair-weather current system. If the cloud is stratoform and of considerable horizontal extent it will tend to increase R and decrease i. A fair-weather cumulus of limited horizontal extent will produce a distortion in the vertical flow pattern, decreasing the current under the cloud and increasing it slightly in regions around the cloud. The vertical current flow is also distorted by surface irregularities such as trees, buildings, valleys and mountains. For example, the current density on a somewhat rounded mountaintop is usually more than three times as great as on a level plain.

Atmospheric Electrical Generators.—In the preceding discussion of atmospheric electricity, attention has been focused on the electrical properties of the medium and the normal distribution of electric fields and currents. Knowledge of the processes which produce electric fields in the atmosphere is clearly fundamental to any coherent interpretation of atmospheric electrical phenomena. Such processes are called atmospheric electrical generators. In spite of their importance and the study devoted to them by many investigators, knowledge concerning generators was still very limited in the mid-20th century because of the complexity of the phenomena involved and the great difficulty in making definitive measurements.

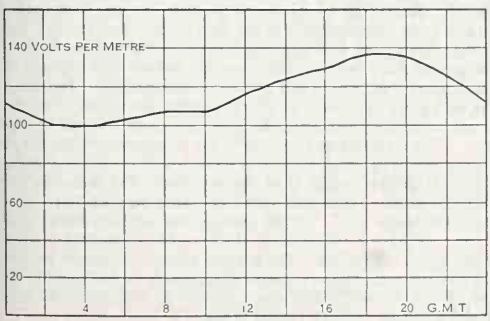


FIG. 3.—MEAN DIURNAL VARIATION OF POTENTIAL GRADIENT AS DETERMINED FROM 47 DAYS OF RECORDING AT SEA ON BOARD THE "CARNEGIE," 1915–21

All of the known generators which produce significant effects in the lower atmosphere derive their energy directly or indirectly from atmospheric motions. Those which are capable of creating an electric field where none existed before may be called primary generators while those depending for their operation upon a previously existing field produced by another generator may be referred to as secondary generators. Most generators produce local effects; only the largest are capable of affecting the whole atmospheric system.

Two generating processes of importance in other branches of electricity deserve brief consideration. In the ordinary dynamo, electric fields are produced by the motion of conductors across a magnetic field. Similarly, motion of the air across the magnetic field of the earth creates an electric field in the atmosphere. Within the highly conducting regions of the lower ionosphere this process, called the dynamo effect, is responsible for large currents which are studied at the surface of the earth by means of the magnetic fields they produce. Within the lower atmosphere, however, the fields created by the dynamo effect are several orders of magnitude less than the normal field and may be disregarded. Also, the process of electromagnetic induction is negligible in the lower atmosphere because of the very small time rate of change of the magnetic field produced either by terrestrial or extraterrestrial sources. For the transient local effects associated with intense electrical discharges, *see* LIGHTNING.

The generators of importance in the lower atmosphere depend upon the creation of positive and negative charges and their physical separation on a macroscopic scale. The actual separation of charge normally involves the transport of one sign by motion of the air itself or by the motion of charged particles or precipitation forms through the air. Such transport of charge by mass motion constitutes an electrical convection current. A number of specific examples of charge separation are given in the succeeding sections.

Fair-Weather Generators.—The generating processes which produce local fields during fair weather may be viewed as perturbing influences on the normal field and conduction current pattern described above. Dust storms in drier climates produce surface fields many times greater than the normal field and, characteristically, in the opposite direction. Snow blown from the surface, particularly in extremely cold regions, produces large fields at the surface usually in the same direction as the normal field. Modern industrial activities are also frequently the source of charge separation. Fields have been observed at the earth's surface produced by smoke or steam from locomotives, by smoke from diesel engines and factory stacks, and by various other means. The charge separation presumably occurs near the point where the particulate matter enters the air and the charge is often carried with the wind for one or more kilometres until it has been dissipated by conduction currents.

Another type of field-generating process which occurs in fair weather may be described as a turbulence generator. It is known that under normal circumstances a positive space charge develops within the first few metres above the surface (the electrode effect), since in the normal field positive ions move into the region from above on their way to the earth while negative ions created in this layer move upward with no compensating supply of negative ions from the earth. Under turbulent conditions, especially those associated with surface heating, upward moving air carries more positive charge than downward moving air. Hence, there is a net upward convection current which enhances the existing field and thereby increases the downward conduction current. It has been suggested that the turbulence generator is responsible for the observed rise of the conduction current density after sun-

rise on clear days. If further investigation supports the suggestion, this secondary generator will be significant in interpreting all fair-weather records over land areas where surface heating is important.

Precipitation Generators.—The most important charge-separating processes are to be found in regions of storm. There have been many suggestions and detailed theories concerning mechanisms which may be responsible for separation of charge in the presence of precipitation, some based on field observation, others on laboratory experiments, still others on known or postulated processes in the atmosphere. While some of the theories appear to offer much greater hope of explaining precipitation generators than others, the very great difficulty of making reliable measurements within precipitating clouds together with the complexity of the phenomena have retarded any final conclusions with respect to the relative importance of various precipitation generators. Accordingly, a few of the proposed mechanisms will be described briefly by way of illustration. A more exhaustive tabulation and a critique of the charge separation processes may be found in *Atmospheric Electricity* by J. A. Chalmers (see *Bibliography*).

The basic requirement of a precipitation generator is that the heavier precipitation forms (water drops, snow, hail) systematically acquire one sign of charge while the lighter forms (minute water droplets, ice crystals) or the surrounding atmosphere acquire a charge of the opposite sign. Since the heavier particles fall relative to the light particles or the atmosphere, separation of charge is accomplished.

A form such as a water droplet may acquire a charge of one sign and the surrounding air the opposite sign if the positive and negative atmospheric ions have different diffusion coefficients. The droplet gains a net negative charge if the negative ions diffuse more rapidly. A similar result obtains if the surface of the droplet selectively captures ions of one sign through chemical adsorption or other type of surface interaction. Any of these processes is limited by the electric field which builds up around the drop as it charges. As the larger droplets fall, however, they grow by accretion of smaller droplets. Accordingly the total charge on a precipitation form can be larger than on the initial small droplets.

A raindrop may also acquire a charge as it falls in the presence of an electric field. Such a drop is polarized in the field and, if the drop velocity exceeds the ionic velocities in its vicinity, it will accumulate an excess of ions of the sign which is attracted to the lower surface. While this process is not a primary generator it is probably of considerable importance in altering charge on drops or other precipitation forms as they fall to the earth from the region where they acquired their initial charge.

There is some evidence to indicate that the most important charge generators, particularly in thunderstorms, are those in which two hydrometeors with different physical or chemical properties collide with each other. Laboratory experiments have shown that when two ice particles at different temperatures make contact, the warmer acquires a negative charge. Also, if the saline contamination of the two ice particles is different, the more contaminated particle will tend to become negatively charged. Applying the results to atmospheric processes, an ice form (graupel) growing by accretion of supercooled water droplets which freeze on contact will be warmer than minute ice crystals growing by sublimations because of the more rapid release of latent heat in the former case. Further, the contamination of the graupel will exceed that of the small crystal. Thus, as the graupel falls, colliding with small ice crystals, it will acquire a negative charge. On a large scale, the process would produce a vertical charge dipole with positive charge uppermost. A second process in this category involves an ice form (hail) accumulating water droplets which form a water film over the ice before freezing. Since a potential difference exists between the ice and water during freezing, if water droplets are blown off the water film as the particle falls, the ice particle and water droplets will have charges of different signs and will fall at different rates.

It is also known that when large water drops splash or are broken up in a strong air blast, the droplets resulting from the breakup are positively charged and the surrounding air negatively charged. This result, then, may also serve as a process of charge separation in the atmosphere under special conditions.

It is quite likely that more than one charging process is operative within a storm cloud at any one time but it is also probable that in any one region of such a cloud, determined by the character of the precipitation forms, one process is dominant. It is generally found that large charge separation occurs only in those clouds from which precipitation is falling.

Observed Precipitation Charges.—Extensive observations of the charge on precipitation forms have been made at the surface of the earth and some within clouds. The results of various observers show rather large differences. Some of the disparity can be attributed to variability in meteorological conditions and some to differences in measuring techniques. Further, since most measurements have been made in the middle latitudes, the results may not be representative of all parts of the earth. With these important limitations, a few tentative general statements may be made.

The charge carried by drops of quiet, continuous rain is more frequently positive than negative. The charge per drop is less than in storm rain and the drops tend to be smaller. Quiet rain is more likely to be predominantly of one sign. By contrast, thunderstorm rain is characterized by large drops and large charges per drop. Its characteristics are variable and the charges of both signs are found, often on consecutively measured drops. More observers have found negatively charged drops predominating than the reverse. Shower rain appears to partake of some of the characteristics of both quiet rain and thunderstorm rain. Charge on quietly falling snow is more frequently negative than positive but under turbulent conditions the reverse appears to be true.

Charges on individual precipitation forms have been observed to range between about 10^{-15} and 10^{-10} coulombs, the charges measured within clouds tending to be larger than those measured at the ground. The electrical convection currents to the ground carried by precipitation are found to vary between 10^{-12} and more than 10^{-8} amp. per square metre, the currents carried by quiet rain being smaller than for showers and thunderstorms.

Observations on the sign of precipitation charge made when the electric field is strong enough to produce corona discharges from surfaces of high curvature at the ground show that the direction of the precipitation current is opposite to the conduction current. The result has been explained in terms of the selective capture of upward moving ions by the falling precipitation in an external field (see *Precipitation Generators*, above). Similar but less conclusive results have been obtained with smaller fields. A significant result of the observation is that one can draw no reliable conclusion concerning the charge of precipitation forms in a cloud from measurement as the precipitation reaches the surface since processes occurring between the cloud and ground may alter the charge magnitudes and even reverse the sign.

Thunderstorms.—The thunderstorm deserves special consideration as the largest and most important generator of electricity in the atmosphere and as a central subject in atmospheric electrical theory. Knowledge of its electrical structure has been derived largely from measurements of the electric field and of rapid field changes associated with lightning strokes. Observations made on the ground have been supplemented by balloon and airplane observations below, within and above the thundercloud. In spite of the very great complexity of charge distribution within a thunderstorm most storms observed have certain gross features in common. The main charge structure of the storm, responsible for the production of lightning strokes, is bipolar with the positive charge region at greater height than the negative. The main negative region is usually found well above the cloud base where temperatures are below 0° C. In the average the negative charge is located near the −15° C. isothermal surface at an elevation of about 5 km. above the surface. The positive charge, commonly between 0.5 and 1 km. above the negative charge, may be horizontally displaced with respect to it in some storms. In addition to the main dipole a smaller positive charge is found in most storms near the 0° C. isothermal surface, below the negative region.

The thunderstorm is a dynamic entity and the charges are con-

tinuously changing. The principal dipole is produced by a convection current, presumably carried by precipitation. As the dipole charges grow, there is a corresponding increase in the conduction currents tending to dissipate the charges. If electrical breakdown of the air did not occur, the charges would increase until the convection current was balanced by the conduction loss, a situation approached in very weak storms with small convection currents. In most intense storms, the charge growth is limited by a lightning discharge between two cloud charges, between a cloud charge and ground or between a cloud charge and space charge of opposite sign outside of the cloud (usually below). More than 90% of the cloud-ground strokes transport negative charge to ground while the strokes between clouds and strokes to the air outside the cloud destroy equal charges of opposite sign.

The charges in the main dipole, which are dissipated by lightning strokes, have an average value of about 25 coulombs, and the smaller positive charge nearer the cloud base is estimated to be about 5 coulombs. The fact that a majority of lightning flashes to ground consist of a series of discharges over the same channel succeeding each other by a few hundredths of a second suggests that the cloud charge regions do not have a uniform charge density but consist of several cells or pockets of charge with a few coulombs each. Observations of the charge dissipated by a lightning stroke and of the time required for the electric field to recover after the stroke lead to estimates of the intensity of the charging current, characteristically of the order of 3 amp.

There is no agreement among investigators concerning the principal mechanism or mechanisms responsible for production of thunderstorm charges. Since the cloud charges develop in regions where the precipitation and cloud particles must be ice forms or both ice and water forms, most theories invoke a mechanism which provides for systematic charging of the precipitation forms of ice, ice and water or water. Examples of processes which can produce the observed polarity in the cloud and which appear to be capable of maintaining an adequate rate of charge separation are described above.

The distribution of currents and fields within and around the thunderstorm is of considerable interest. Fig. 4(A) is a schematic diagram showing the directions of the main charging current, the principal conduction currents and fields for a bipolar thunderstorm model in a medium in which the conductivity increases with altitude. The diagram purports to show only gross features of the electrical system and no complex detail.

At ground level the field under the principal dipole is opposite the normal field and over flat terrain may reach values of the order of 10,000 v. per metre just before a lightning discharge. The field diminishes with horizontal distance from the dipole and reverses sign about 10 km. from it. Finally, at about 30 km., the thunderstorm field becomes small compared with the fair-weather field. Under the thundercloud the field intensity increases with altitude reaching values a few times greater than the surface field near the cloud base. Fields measured within the cloud are normally less than 10^5 v. per metre, about 15 times less than the breakdown field at the level of the principal dipole. It appears, therefore, that fields large enough to initiate a lightning stroke must be confined to very small volumes in the immediate vicinity of the larger charge concentrations. Estimates of the potential difference between the large charge concentrations and ground, based upon the distribution of electric fields, yield values between 10^8 and 10^9 v.

An examination of fig. 4(A) reveals that there are large volumes below the negative charge and above the positive charge in which the field produces an upward conduction current. The convection current between the charges, also directed upward, must exceed the downward conduction current in this region. Thus, the net effect of the thunderstorm is to transport positive charge from the earth toward the high atmosphere. After a cloud-ground lightning stroke which discharges the lower negative charge, the upward current to the ionosphere must be enhanced as shown in fig. 4(B). It will be noted that more current must flow toward the ionosphere than toward the earth under these circumstances, for while the potential difference between the positive charge and the ionosphere is slightly less than between charge and ground, the resistance of the upward conducting paths is less than that of the downward paths.

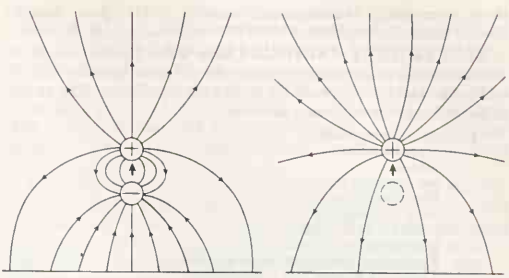


FIG. 4.—SCHEMATIC DIAGRAM OF CURRENTS IN A THUNDERSTORM: (A) BEFORE LIGHTNING STROKE; (B) AFTER STROKE TO GROUND. HEAVY ARROWS INDICATE CONVECTION CURRENT CHARGING MAIN DIPOLE

Measurement of the average magnitude of the total upward current is one of the crucial tests in determining the world-wide importance of thunderstorms. Measurements at the surface are complicated by the presence of corona (or point) discharges which develop in high fields at surfaces of high curvature and by falling charged precipitation which varies significantly with position under the cloud. Two attempts have been made to measure the upward current under the more favourable conditions over cloud tops, one using an airplane and the other using balloons. The respective average currents were 0.5 amp. and 1.3 amp. per storm, and it could be concluded that the average upward current was of the order of magnitude of 1 amp.

Climatological data indicate that on the average there are 1,800 thunderstorms in progress over the world at any one time. Thus, the total upward current supplied by thunderstorms must be about 1,800 amp. The agreement between this estimate and the estimated total air-earth current of about 1,500 amp. is closer than could be expected in view of the accuracy of several measurements involved. However, it does give strong support to the hypothesis that thunderstorms are the principal generators of the air-earth current.

Further evidence on this point may be found in the fact that, throughout its average daily variation, the air-earth current density is approximately proportional to the number of thunderstorms in progress over the world. Climatological records show that thunderstorms are dominantly tropical and subtropical phenomena. Over land areas storms are much more frequent in the afternoon hours than at other times of the day, while over the sea there is no pronounced daily variation in thunderstorm frequency. Thus, the large tropical thunderstorm centres over land make the greatest contribution to the daily variation in world thunderstorm activity. Quantitative estimates of the number of storms in progress at each hour of the day show that the maximum number occurs between 15 and 20 hours, universal time, when it is respectively mid-afternoon over central Africa and the Amazon valley. The minimum number occurs about 4 hours, universal time, when it is late evening or early morning over the same centres. The estimated daily variation in thunderstorm activity is strikingly similar in relative amplitude of oscillation and phase to the daily variation in the potential gradient (and hence to the air-earth current density) over the oceans, shown in fig. 3.

While the observations cited are not so precise as to exclude possible small contributions from other generators, there is little doubt that world thunderstorm activity is the primary cause of the earth's normal electric field and air-to-earth current.

See also METEOROLOGY; ATMOSPHERE; ELECTRICITY.

BIBLIOGRAPHY.—J. A. Chalmers, *Atmospheric Electricity* (1957); J. A. Fleming (ed.), *Terrestrial Magnetism and Electricity*, ch. iv by O. H. Gish, ch. v by O. W. Torreson, ch. xiii by M. D. Harradon (1949); "Atmospheric Electricity" in T. F. Malone (ed.), *Compendium of Meteorology* (1951); B. F. J. Schonland, *Atmospheric Electricity* (1953);

H. R. Byers (ed.), *Thunderstorm Electricity* (1954); J. A. Ratcliffe (ed.), *Physics of the Upper Atmosphere* (1960). (R. E. Hr.)

ELECTRICITY, CONDUCTION OF. The mechanism of the conduction of electricity is markedly different in solids, liquids and gases, which are accordingly dealt with separately. The article is divided into the following sections:

I. Conduction in Solids
1. Introduction
2. Metals
3. Semiconductors
4. Photoconductors
5. Ionic Conduction
II. Conduction in Liquids
1. Introduction
2. Electrolyte Resistance Measurements
3. Theory of Electrolytic Conductance
III. Conduction in Gases
1. Introduction
2. Basic Processes
3. Processes Leading to Voltage Breakdown
4. The Glow Discharge
5. The Arc
6. The Gas Plasma

The conduction of electric current in a vacuum is due to electrons that are controlled by electric and magnetic fields, and is more appropriately treated elsewhere (*see* ACCELERATORS, PARTICLE; ELECTRON TUBE; MICROSCOPE: *The Electron Microscope*).

I. CONDUCTION IN SOLIDS

1. Introduction.—The property of transmitting electricity is common to almost all states of matter, though the ease with which a current can pass varies enormously from one substance to another. Metals are the most effective conductors and dielectrics the least effective. The latter will only transmit charge under extreme conditions. The earliest recorded observation of electrical conduction appears to be due to Stephen Gray who, in 1729, before the discovery of the galvanic cell, observed the transmission of electrical effects over a considerable distance along a packthread suspended by silken hoops.

One general law, known as Ohm's law, is applicable to all true conductors. This law states that the current transmitted by a particular substance is directly proportional to the applied potential difference. If the electrical current is denoted by I and the potential difference by V, the constant ratio V/I is, by definition, the resistance of the specimen. If the specimen is in the form of a cube whose edges are 1 cm. in length, and a uniform current flows between two opposite faces, the resistance of this specimen is known as the resistivity of the material, which may be denoted by ρ. The reciprocal of the resistivity is called the conductivity and is denoted by σ. Resistivity can be measured in the practical units (ohm-centimetre) or in absolute electrostatic (E.S.U.) or electromagnetic (E.M.U.) units. The conversion factors for units of resistance are

$$1 \text{ E.S.U.} = 9 \times 10^{11} \text{ ohms}$$
$$1 \text{ E.M.U.} = 10^{-9} \text{ ohms}$$

Typical numerical values of the conductivity for common materials in ohm^{-1} cm.$^{-1}$ and 0° C. are:

Copper	640,000	Glass	10^{-12}
Iron	110,000	Quartz	10^{-14}
Graphite	300	Rubber	10^{-16}

Solid conductors of electricity may be classified conveniently into the following main groups: (1) metals, characterized by a high conductivity which decreases as the temperature increases; (2) semiconductors, which are roughly intermediate in their electrical properties between metals and insulators, and in which the conductivity rises with increasing temperature; (3) photoconductors; (4) ionic solids.

2. Metals.—Metals are characterized by their high electrical and thermal conductivities; that is to say, they transmit large electrical currents under the influence of a small difference of potential, and large quantities of heat under the influence of a small difference of temperature. Both effects are known to be caused by the presence of free electrons. The resistivity of metals depends on the temperature, being roughly proportional to the absolute temperature; for copper, for instance, the resistivities at the temperatures of liquid air, melting ice and boiling water have the relative values

0.184 1.0 1.431

The resistivity is also sensitive to purity, increasing with increasing impurity content. Copper has the lowest resistivity of the common metals (only silver has a lower value), and is therefore extensively used for wires and cables carrying current. Table I gives the values of the resistivity at 0° C. for most of the metallic elements. The second column of the table gives the temperature coefficient of the resistivity, which is defined as the increase in resistivity per degree divided by the resistivity at 0° C.

TABLE I.—*Resistivity and Temperature Coefficient of Resistivity of Some Metallic Elements at 0° C.*

Element	ρ microhms cm.	α
Aluminum	2.5	4.67×10^{-3}
Antimony	32.1	5.1
Arsenic	26.0	..
Barium	30.0	6.1
Beryllium	2.78	7.5
Bismuth	107.0	4.45
Cadmium	6.73	4.26
Calcium	3.6	4.0
Cesium	19.0	5.0
Chromium	15.0	..
Cobalt	5.2	6.58
Copper	1.55	4.33
Gadolinium	140.0	3.45
Gallium	13.7	4.1
Gold	2.04	3.98
Indium	8.2	5.1
Iridium	4.74	4.33
Iron	8.71	6.57
Lead	19.3	4.22
Lithium	8.5	4.37
Magnesium	3.94	4.2
Molybdenum	5.03	4.7
Mercury	21.2	..
Nickel	6.58	6.75
Palladium	9.77	3.8
Platinum	9.81	3.92
Potassium	6.3	5.4
Rhenium	18.9	3.1
Rhodium	4.33	4.57
Rubidium	11.6	5.3
Silver	1.5	4.1
Sodium	4.27	5.5
Strontium	20.0	5.0
Tantalum	12.4	3.6
Thorium	10.1	3.3
Tin	10.1	4.03
Tungsten	4.89	4.83
Uranium	25.0	..
Vanadium	18.2	..
Zinc	5.45	4.2
Zirconium	40.5	4.0

The resistivity is given by the formula $\rho = \rho_0(1+\alpha T)$, where ρ_0 is the value at 0° C., and T the temperature in degrees centigrade.

Effect of Temperature and of Impurities.—For nearly all metals over the whole temperature range, the resistivity increases with increasing temperature; certain special alloys such as constantan are exceptions. At high temperatures the resistivity is relatively insensitive to the purity of the metal, but as the temperature is lowered toward absolute zero, the resistivity tends toward a value (the residual resistivity ρ_r) which depends entirely on the purity of the metal. For very pure metals it is very small, while for alloys it may be a large fraction of the total resistivity at room temperature. The types of curve for pure and impure metals are shown in fig. 1. It is a good approximation for most metals and alloys to write for the resistivity

$$\rho = \rho_r + \rho(T)$$

where ρ_r depends on the nature and the amount of foreign metal introduced as impurity, and is independent of the temperature. $\rho(T)$ depends on the temperature but does not depend on the impurity content. This division of the resistivity into two parts is known as Matthiessen's rule.

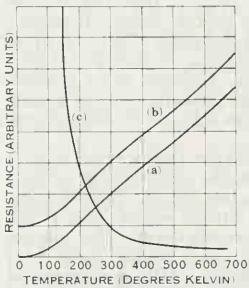


FIG. 1.—DEPENDENCE OF RESISTANCE OF TEMPERATURE

(a) pure metal (copper) (b) impure metal or alloy (c) semiconductor (on different scale)

The temperature variation of the resistivity has been examined over a wide range of temperatures for many metals. It is found that at very low temperatures, of the order of 1° K., $\rho(T)$ varies as a high power of T. In some cases, *e.g.*, gold, the low temperature variation is according to T^5, which is the power law suggested by existing theory. In other metals the effective power is sometimes smaller and sometimes greater than 5; the reasons for this are not yet fully understood.

A curious discovery of recent years is that, in this low-temperature region, certain metals show a minimum of the resistivity when this is plotted against temperature. Thus in silver with small traces of manganese, for example, the resistivity falls to very low values at a few degrees Kelvin; and thereafter, as the temperature is still further lowered, the resistivity rises. No satisfactory theory of this effect has yet been put forward.

For magnetic metals such as iron, nickel and their alloys, the resistivity increases with abnormal rapidity in the neighbourhood of the Curie point; that is to say, as the temperature is raised to the point where the magnetic properties disappear.

The resistivity ρ_r depends on the amount of impurity which has gone into solid solution in the metal and which is therefore dispersed in atomic form. If an alloy consists of a two-phase system, *i.e.*, a mixture of two different metals, the conductivity of the alloy is a weighted mean of those of the two phases. When one metal enters into solid solution in another, however, a remarkable rise in the residual resistivity ρ_r always occurs. The rise is greater, the greater the dissimilarity of the two elements in both their chemical characteristics and their atomic sizes. Thus the effect of adding silver or gold to copper is rather small, while the effect of adding a comparable amount of antimony is some 30 times as great. The effect of alloying small amounts of other metals was extensively studied in the Technical high school at Stockholm, Swed., where the remarkable result was obtained that the increase of resistivity caused by the addition of one atomic per cent of a large number of elements to copper, silver or gold is proportional to $(Z-1)^2$, where Z is the number of electrons outside a closed shell in the element added or, in other words, the valence. Thus $Z=2$ for zinc, $Z=4$ for tin, $Z=5$ for antimony and so on.

Effect of Pressure.—This was extensively studied by P. W. Bridgman at Harvard. Usually the conductivity increases with increasing pressure, although the percentage change is small, and is of the same order as the volume change. There are certain exceptions to this rule, such as lithium, calcium and strontium.

Effect of Melting.—The resistivity of most metals increases on melting by a factor of the order two, although that of antimony and bismuth decreases.

Effect of a Magnetic Field.—A magnetic field increases the resistivity of all metals except iron, nickel and cobalt (the ferromagnetic metals) for which the resistivity decreases. The effect is in general quite small; thus a field of 200 kilogauss increases the resistivity of copper by 0.25%. For bismuth, a comparatively enormous effect is observed, so much so that the resistivity of spirals of thin bismuth wire is used for measuring magnetic fields. The effect is particularly large at low temperatures. Thus a field of 20 kilogauss will about double the resistivity of bismuth at room temperature, but will increase it by a factor 250 at the temperature of liquid air. For small fields the change of a resistivity is proportional to the square of the field, but for strong fields a linear law is obtained.

Other measurements frequently made on conductors in a magnetic field are classified under the general heading of Hall effect determinations. (For an explanation of this effect *see* below under *Semiconductors.*) Measurements of the Hall coefficient have been made for many metals, and these have played a significant part in the development of the modern electron theory, particularly for semimetals such as bismuth.

Thermal Conductivity.—A very important property possessed by all good electrical conductors is the ability to transmit heat. Thermal conductivity is possessed in a greater or less degree by all substances. It is measured by the quantity of heat transferred across unit cross section in unit time for unit temperature gradient. In solids, three different types of thermal conductivity may be distinguished. Amorphous solids such as glass possess a conductivity, usually small, which does not vary much with temperature. Crystals such as rock salt possess a conductivity roughly inversely proportional to the temperature. At ordinary temperatures it is usually of the same order as the conductivity of an amorphous glass, but at low temperatures it becomes much larger. The expression "low and high temperatures" refers to the characteristic temperature of the substance, familiar from the theory of the specific heats.

Electrical conductors possess in addition a thermal conductivity which would in a nonconductor be considered large and which does not vary greatly with temperature. The mechanism which enables heat to be transported in conductors is closely allied to that which comes into play in electrical conduction. This is shown by the fact that there is a very distinct proportionality between the electrical and the thermal conductivity of the metals. The so-called Wiedemann-Franz constant is the ratio of thermal to electrical conductivity. As shown by Table II, it is approximately constant for a large number of metals, whose conductivities diverge by a factor of no less than 27. It will be noted that there is a steady increase in the constant as the conductivity decreases. This is in part because of the fact that the metals possess, in addition to electronic conductivity, the ordinary atomic conductivity of every solid. It is clear that this will become more important as the electronic conductivity diminishes. Its amount may be estimated in the case of single crystals, where it is known that the atomic conductivity varies inversely as the temperature, and a corrected Wiedemann-Franz constant can be obtained by subtracting the atomic conductivity from the total conductivity and dividing the electronic conductivity thus obtained by the electrical conductivity. This is found, in the cases examined, to give a result even more constant than that obtained in Table II. In the last column of the table may be found the temperature coefficient of the ratio of the conductivities. As will be seen, this is approximately 0.36%; *i.e.*, the Wiedemann constant appears to be proportional to the absolute temperature.

Thermoelectric Effects.—One cannot consider the electrical conductivity of metals without taking into account secondary effects connected with the mechanism by which electricity is transported from one part to another. The best known of these is the thermoelectric effect, also well known as the Seebeck effect. If two different metals are brought into electrical contact at two points, thus forming a circuit, and the two junctions are kept at different temperatures, a current flows round the circuit. The voltage difference for any given pair of metals depends solely upon the two temperatures, the current, of course, being determined by this voltage difference and the resistance of the circuit. At any given temperature the metals may be arranged in a definite order according to the potential difference which is produced when a given temperature difference is applied between each metal and any one of them arbitrarily chosen as the zero. The potential difference produced by forming a circuit between any two of them is then simply the difference between the voltages produced against the standard metal. Table III gives a list of the commoner metals and the potential differences arising from a temperature difference of 1° C. at 0° C. It is evident that these thermoelectric constants are small, usually less than 10^{-4} v. per degree. Their magnitude depends upon the temperature, becoming smaller as the absolute

TABLE II.—*Heat and Electrical Conductivity of Metals*

Item	Heat conductivity divided by electrical conductivity at 18° C. × 10 ^{8} E.M.U.	Temperature coefficient of heat conductivity divided by electrical conductivity × 10^{-4}
Theoretical value	517	36.6
Aluminum	637	42
Cadmium	707	38
Copper	668	40
Gold	710	..
Iron	802	43
Lead	715	40
Nickel	698	40
Silver	686	38
Tin	735	34
Zinc	670	38
Manganin	912	25
Steel	890	36

TABLE III.—*Thermoelectric Effect in Metals*

Metal	Thermoelectric effect against platinum at 0° C. for 1° diff. in temp. $\times 10^{-6}$ v.
Aluminum	3.8
Antimony	47.0
Bismuth	−65.0
Copper	7.4
Iron	16.0
Palladium	−5.6
Silver	7.1
Constantan	−34.4

zero is approached. Small though they are, the thermoelectric effects form the basis of a large class of instruments for measuring temperature. It is evident from what has been said that all that is required is to solder or weld wires of two metals together and measure the voltage difference between the two other ends when the junction is brought to the temperature it is desired to measure. An inversion of the thermoelectric effect is the Peltier effect. If two dissimilar metals are brought into electrical contact and a current is passed through the junction this is heated or cooled according to the direction of the current. The amount of heating or cooling for a given current at a given temperature depends solely upon the metals used.

The Thomson effect is of somewhat similar character. If a current is passed through a metal, parts of which are at different temperatures, a certain transport of heat takes place, the amount of which depends upon the strength of the current.

Conductivity at High Frequencies.—At high frequencies an electrical current in a metal confines itself to a shallow zone on the surface as a result of mutual inductance of the elements of the current. This effect is known as the skin effect, and the depth in the metal to which the field penetrates is called the skin depth. It is possible to define an effective surface resistivity which may be shown, in accordance with classical electrodynamics, to be equal to $2\pi\sqrt{(\gamma/c^2\sigma)}$, where γ is the frequency of the applied electromagnetic field, σ the bulk conductivity of the metal for steady currents and c the velocity of light.

An interesting phenomenon known as the anomalous skin effect has been the subject of many investigations in recent years. This phenomenon consists of the tendency of the surface resistivity to become independent of σ at low temperatures and high frequencies. The effect arises when the mean free path of the electrons in the metal becomes comparable with the skin depth as given by the above expression.

The effects on metals of visible light and radiation, in both the ultraviolet and infrared regions of the spectrum, may be regarded as special cases of conduction under the influence of oscillatory electromagnetic fields. From this point of view, satisfactory theories have been constructed of the optical properties of metals.

Superconductivity.—This remarkable phenomenon was discovered by H. Kamerlingh Onnes in 1911. The most striking feature of a metal in the superconductive state is the complete absence of electrical resistance. In all pure metals the electrical resistivity falls to very small values as the temperature approaches absolute zero, but the distinction between a small and an actually zero resistance is readily demonstrated. Only when the resistance is truly zero is it possible to observe a persistent current. The existence of a persistent current may be shown as follows: A ring of lead is placed in a Dewar flask in the presence of a magnetic field such as is produced, for example, by a bar magnet. Liquid helium is then poured into the flask, thereby reducing the temperature of the lead to approximately 4° K., and the external magnetic field is removed by withdrawing the bar magnet. The change in magnetic flux through the lead ring induces an electromotive force which gives rise to a circulating current. This current is made manifest by its magnetic field, which persists indefinitely so long as the temperature remains below 7.26° K. When the temperature is raised above this value, the current and its magnetic field vanish. The lead is then in its normal state, in which, though the resistivity is still small, the current is dissipated almost immediately.

There are 23 pure metals which become superconductive when cooled to sufficiently low temperature. Transition temperatures for pure metals vary from less than 0.5° K. for titanium and hafnium to 8.7° K. for niobium, and 11.2° K. for technetium. It appears that not all metals become superconductive. Experiments were carried out at temperatures less than 0.1° K. on monovalent metals such as copper, silver and the alkalis, with the result that no indication of superconductivity was found.

Superconductivity is not restricted to the pure metals only. Many alloys and chemical compounds show this effect. In some cases transition temperatures are found which are appreciably higher than those of the pure metals; *e.g.*, niobium nitride, NbN, in which it is 14.7° K.; and the intermetallic compound of niobium and tin, Nb_3Sn, which has the highest transition temperature yet recorded, namely, 18° K.

An important early discovery concerning the nature of superconductivity was made by Kamerlingh Onnes himself in 1913, when he observed that the superconductive state of a wire could be destroyed by passing through it a sufficiently large current. It was shown later that this was due to the magnetic field generated by the current, and from this observation grew the large body of experimental data on threshold curves. These are graphs of the critical magnetic field at which superconductivity disappears, plotted against the absolute temperature.

A notable step forward in the understanding of superconductivity was taken in 1933 with a discovery of W. Meissner and R. Ochsenfeld which can be understood as follows: Imagine a long cylindrical wire in a uniform magnetic field which is parallel to the direction of the wire. Initially, let the intensity of the magnetic field be small and the temperature of the wire below the critical temperature. Now let the intensity of the field be increased to some value below the threshold for the temperature concerned. The changing field induces persistent currents whose magnetic effect just balances the external field, leaving the magnetic induction within the metal zero. This is in accordance with the laws of electromagnetic induction applied to a metal with zero resistivity, and it does not, therefore, lead to new information. Now, however, let the experiment be performed in a somewhat different way. The magnetic field is first applied to the metal in the normal state and maintained constant. The metal is then cooled through the transition temperature, and again it is found that the magnetic flux is forced out of the metal, leaving the interior in zero field. This result goes beyond what could have been expected on the basis of merely a zero resistance in the superconductive state, since in this experiment the external magnetic field is not changed. The superconductive state is thus characterized by zero magnetic induction or, equivalently, it may be said to be a perfect diamagnetic with susceptibility equal to $-\frac{1}{4}\pi$. The transition between the normal and superconductive state has been shown to be thermodynamically reversible, which implies that when the persistent currents are destroyed by an increasing magnetic field, this occurs reversibly and not as a result of Joule heating.

In 1950, a significant step forward was taken by the discovery of the isotope effect in the superconductivity of mercury, tin and lead. It has proved possible to separate the isotopes of these metals and to measure the superconductive properties of samples of different atomic mass. The transition temperature in the absence of a magnetic field is found to vary very nearly as the inverse square root of the atomic mass. Expressed symbolically, this may be written

$$M^n T = \text{constant}$$

where M denotes the isotopic mass, T the transition temperature and n is nearly equal to 0.5. Exact measurements for tin give $n = 0.46$, and for mercury $n = 0.504$.

The theoretical significance of this effect is that it shows clearly that atomic vibrations normally associated with thermal motion are intimately connected with the existence of the superconductive state. Based on this experimental discovery, the mathematical work of H. Fröhlich, J. Bardeen and others resulted in a far more satisfactory understanding of the phenomenon of superconductivity than had hitherto been obtained; and it appeared that a solution of this problem, which was one of the outstanding problems of metal physics, was in sight. For additional information on this,

as well as other low-temperature phenomena, *see* LOW-TEMPERATURE PHYSICS (CRYOGENICS).

Theory of Metallic Conduction.—Every major advance in atomic physics has led to an advance in the theory of conduction. For instance, the discovery of the electron led to the suggestion that the electrons might be the charge-carrying element in metals. This was actually proved by the experimental work of Richard Tolman and his colleagues, who set a coil of wire into rapid rotation about its axis and then suddenly stopped it. The electrons in the metal continued moving, giving an observable current, the magnitude of which enabled a value of e/m, the ratio of charge to mass of the elementary carrier of electricity, to be determined. This agreed sufficiently closely with the value determined for the electron in other experiments.

In the hands of H. A. Lorentz and Paul Drude a theory of metallic conduction was built up in the first years of the 20th century, in which it was supposed that a dense gas of free electrons, interacting strongly with each other, existed in a metal. In order to account for the Hall effect the number of free electrons was assumed to be of the same order as the number of atoms. The theory had certain successes, among which was the explanation of the relation between electrical and thermal conductivity, and the theoretical derivation of the Wiedemann-Franz number. Until the introduction of quantum mechanics in 1926, however, the theory was very incomplete and in some respects gave definitely the wrong answer, notably in the following respects:

1. So large a number of free electrons should make a large contribution to the specific heat of a metal, which was not in fact observed.

2. Since the electrons would undergo frequent collisions with the atoms of the metal, it would be difficult to understand a mean free path much greater than the interatomic distance; in order to explain the high values of the electrical conductivity, however, especially at low temperatures, a value of the mean free path several thousand times greater than this had to be assumed.

Numerous theories were introduced to try to explain these facts, associated with the names of Wilhelm Wien, J. J. Thomson, F. A. Lindemann, J. Frenkel and many others. With the introduction of quantum mechanics in 1926, however, it was found that the difficulties in the theory could be resolved without any special assumptions. Arnold Sommerfeld, Felix Bloch, Rudolf Peierls and others were the first to apply the new theories to the subject, and by 1930 the main features were understood.

The achievements of the new theory were as follows: It shows that the electrons in the metal obey the so-called "Fermi-Dirac statistics," which means that they have only a very small specific heat. Then the application of wave mechanics shows that electrons would not be scattered at all in a metal if the atoms were arranged in a perfectly regular array; the finite mean free path and with it the finite resistivity arise only when the atoms are displaced from their mean positions. This accounts for the approximate proportionality between the electrical resistivity and the thermal energy, because the thermal energy is a measure of the mean displacement of the atoms. It also accounts for the high resistivity of alloys; a foreign atom in solid solution has a high scattering power, especially if its size is very different from that of the atoms of the solvent.

In the theory of so complicated a phenomenon as conduction many approximations have to be made, especially with regard to the interaction between the electrons; and for many years some difficulties of principle also remained. However, after the discovery of plasma oscillations in the electron gas of a metal by D. Bohm and D. Pines in 1951, many of the difficulties associated with electron interaction were eliminated.

3. Semiconductors.—The semiconductive materials which were most widely used in the late 1950s were the elements germanium and silicon and, to a smaller extent, selenium and tellurium. Several intermetallic compounds (*q.v.*), such as indium antimonide, InSb, and magnesium stannide, Mg_2Sn, possess semiconductive properties; as do also some oxides, notably cuprous oxide, and certain sulfides.

The main characteristic of all semiconductors is that the resistivity decreases as the temperature increases (*cf.* fig. 1). Broadly speaking, the relation between resistivity and temperature is of the form

$$\log \rho = \frac{A}{T} + B$$

where A and B are constants. The conductivity is very sensitive to the degree of purity of the specimen. Very pure germanium has a high resistivity of the order of 20 ohm-cm. Small additions of indium or antimony in solid solution will reduce the resistivity by many orders of magnitude.

Electrical conduction in semiconductors can only be understood, even qualitatively, on the basis of concepts introduced by wave mechanics. In metals, the essentials of conduction can be appreciated in terms of concepts such as that of the free electron, and that of the mean free path for collision between an electron and the lattice ions. In semiconductors the essentials of the phenomenon turn on the nature of the electronic motion through the crystal lattice. The result of applying wave mechanics to the motion of an electron in a space-periodic field, such as that arising from a crystal of positive ions, is that the energy states are grouped into so-called bands. Thus, unlike a particle in a gas, an electron in a crystal can only exist with an energy that lies in certain specified ranges. States with energies between these ranges are excluded. The result is that the energy spectrum possesses a band structure. When all the states in an energy band are completely occupied by electrons, there is no resultant current; and the application of an electric field of ordinary magnitude is unable to produce a flow of charge because no electron can receive a small increment of energy. A crystal in such a state is a nonconductor. If the energy gap between the top of the occupied band, often known as the valence band, and the bottom of the next unoccupied one is small, it is possible for some electrons to be excited thermally into this band, which is referred to as the conduction band. Here the electrons are free to move through the crystal and, therefore, under the influence of an external field, will give rise to an electrical current. A crystal in which the energy states are of this kind is known as an intrinsic semiconductor. Because the number of electrons in the conduction band increases with temperature, the conductivity also increases with temperature, in contrast to the behaviour of metals, in which the conductivity falls as the temperature rises.

When electrons are thermally excited from the valence band into the conduction band, the incomplete valence band itself can then sustain a current. It may be shown that the electrons in an almost completely filled valence band can give rise to an electric current which is exactly similar to the current which would arise were the vacant levels filled by electrons with positive charge. These hypothetical carriers of currents are termed positive holes. It is found that the complex phenomena of conduction in various semiconductor devices, such as rectifiers and transistors (*q.v.*), can be understood in terms of these two concepts: conduction electrons and mobile positive holes.

Most industrially important semiconductors are of the extrinsic type. These usually consist either of pure germanium or pure silicon, into which small amounts of impurity are deliberately introduced. In ideally pure germanium, the conduction band is completely filled when there are exactly four valence electrons per atom. If an atom of antimony, which has five valence electrons, replaces an atom of germanium in the crystal, the extra electron is readily excited by thermal motion into the conduction band, and gives rise to conductive properties. When, as in this case, the impurity has a higher valency than the germanium solvent, the resulting semiconductor is known as an n-type semiconductor, which means that the current is carried by electrons in the conduction band. If, on the other hand, the impurity is, for example, indium, which is trivalent, the result is to introduce positive holes in the valence band and no electrons in the conduction band. In such a crystal, known as a p-type semiconductor, the current is carried by positive holes.

It is possible to distinguish experimentally between n-type and p-type semiconductors through a phenomenon known as the Hall effect. Fig. 2 shows a plate of thickness t of conducting material

(in this case a semiconductor) with a magnetic field H applied normally to its surface. If a current I is passed through the plate in the direction shown, it is found that a potential difference is established between the points A and B. This potential difference is known as the Hall voltage and may be denoted by V. The ratio Vt/IH is constant for a given material and is known as the Hall coefficient. The sign of this coefficient depends upon the sign of the carriers which give rise to the current. Hence it is negative for n-type and positive for p-type semiconductors. It may be shown, moreover, that in certain simple cases the Hall coefficient is inversely proportional to the number density of carriers. Thus, for example, in purely n-type material it is inversely proportional to the number density of electrons, and in the purely p-type inversely proportional to the number of holes. When both holes and electrons are present, as in an intrinsic semiconductor, the relationship is more complicated.

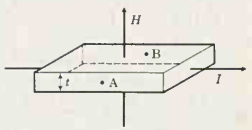


FIG. 2.—THE HALL EFFECT (SEE TEXT FOR FURTHER EXPLANATION)

The importance of this effect for unraveling the mechanism of conduction in semiconductors is that, by combining measurements of the Hall coefficient with measurements of the conductivity, it is possible to determine separately the number of carriers and the mobility, the latter being the average drift velocity for unit electromotive force.

Transistors.—A direct result of the knowledge of the mechanism of semiconduction has been the invention and development of the transistor. This is a device which will perform many of the functions of the triode thermionic valve. It consists of a small n-type piece of germanium to which three electrodes are attached as indicated in fig. 3. A and B denote two fine metallic wires separated at the points of contact by only about 0.1 mm. C is a large area metal contact. The potentials in the circuit are such that electrons are driven from B to C, and a positive current from A to C. The potential difference between B and C is greater than that between A and C. The action of the device depends upon the fact that positive holes are injected into the germanium at A, and diffuse toward the point contact B, thereby reducing the resistance between B and C. Thus an oscillating potential applied between A and C produces an oscillating resistance between B and C, and therefore an oscillating current in the outer circuit connected to these points. Just as with a triode valve, it is possible to arrange that the power in the circuit containing BC is greater than that in the circuit containing AC. Hence the device will operate as an amplifier. Advantages over the thermionic valve are smallness of size and elimination of filament current.

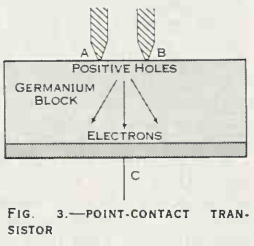


FIG. 3.—POINT-CONTACT TRANSISTOR

4. Photoconductors.—A number of substances, insulators or poor semiconductors in the dark, show a marked increase in conductivity on exposure to light. This property, known as photoconductivity, has many technical applications in the construction of photocells. Selenium and thallous sulfide are much used for the purpose. A large number of semiconductors, however, show the effect, and its explanation is similar to that of the former property: electrons are held on impurity atoms, from which they are released by the action of light.

A large amount of systematic work on the properties of certain semiconductors was carried out by R. Pohl's school in Göttingen, Ger., in the period 1920–39. It was shown that one must distinguish between the primary conductivity, which is the motion of the electrons actually released by the light; and the secondary conductivity, a lowering of the resistance of the material which follows on the motion of the primary current. It is the secondary current which is important in selenium cells.

Photoconductivity is an important process in photochemical reactions in solids. For instance, the primary action of light on the silver bromide grains of a photographic emulsion seems to be to make them conducting, the electrons released by the light setting in operation the reduction to metallic silver.

5. Ionic Conduction.—Many metallic salts and oxides show ionic conduction if the temperature is high enough. Thus silver bromide is a conductor at room temperature, and the alkali halides conduct at about 600° C. The conductivity is small and very sensitive to temperature. An explanation of the phenomenon was first given by C. Wagner and J. Frenkel, and is now well established. In silver bromide, for instance, as the temperature is raised an increasing number of silver ions leave their normal positions in the crystal lattice and move into the so-called interstitial positions, where they are squashed into the small space between the ions of the crystal. These interstitial ions can move about through the solid, as can also the holes from which they have come. Both processes lead to a transport of silver ions.

Ionic conduction is an important process in the oxidation of metals. Thus in the oxidation of aluminum or copper it is believed that metal ions move through the oxide layer by a process of the type described above, and combine with oxygen at the surface.

BIBLIOGRAPHY.—Articles by N. Gerritsen (experimental) and H. Jones (theoretical) in *Handbuch der Physik*, vol. xix (1956); A. H. Wilson, *Theory of Metals*, 2nd ed. (1953); E. Grüneisen, "Metallische Leitfähigkeit," *Handbuch der Physik*, vol. 13 (1928); N. F. Mott and H. Jones, *Theory of the Properties of Metals and Alloys* (1936); N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals*, 2nd ed. (1948); P. W. Bridgman, *Physics of High Pressure*, 2nd ed. (1950); D. Shoenberg, *Superconductivity* (1952); W. Shockley, *Electrons and Holes in Semiconductors* (1950); C. Kittel, *Introduction to Solid State Physics*, 2nd ed. (1956); J. Eisenstein, *Rev. Mod. Phys.*, vol. 26 (1954).
(N. F. M.; Hy. Js.)

II. CONDUCTION IN LIQUIDS

1. Introduction.—The transport of electricity through liquids, as through any matter, is effected by the movement of electric charges which may be electrons, as in metallic conductors such as mercury; positive and negative ions, as in electrolytic conductors such as salt, acid and base solutions, or fused salts; or larger charged particles as well as ions, as in colloidal conductors such as soap solutions.

The extent to which electric current flows through an electrolytic conductor depends on the applied voltage and on the number of ions or other charged particles which are present and their current-carrying capacity in the solvent at a given temperature and pressure. Electric current can be led into and out of an electrolytic conductor through solid electrodes, other electrolytic conductors or gaseous conductors. The circuit as a whole, however, generally contains two metallic electrodes, an anode and a cathode, at which chemical reactions always take place as the conductance changes from ionic to electronic transport or vice versa. Negative ions (anions) move toward the anode, where electrons are released to the metal and oxidation occurs; positive ions (cations) move toward the cathode, where electrons are absorbed from the metal and reduction occurs. The nature of these electrochemical changes (electrode reactions) is discussed in the article ELECTROCHEMISTRY, whereas in this section only the flow of electric current through the liquid electrolyte will be considered.

In metallic conduction it is found that the current is proportional to the applied electromotive force. According to this relation, known as Ohm's law, $I = V/R$ where I is the current, V the electromotive force and R the resistance of the circuit. Ohm's law has been confirmed in the case of metallic conduction to a very high degree of accuracy. It was first shown by Friedrich Kohlrausch that an electrolyte possesses a definite resistance which has a constant value when measured with different currents and by different methods. It has also been established that if effects caused by changes at the electrodes and by variations in concentration, leading to a counter electromotive force, are eliminated or corrected, Ohm's law applies to electrolytes as well as to metals.

2. Electrolyte Resistance Measurements.—The usual and most satisfactory method of measuring the resistance of electrolytes is that of Kohlrausch and L. Holborn, which consists in

eliminating the effect of polarization by the use of platinum electrodes with alternating currents; *i.e.*, currents that are very rapidly reversed in direction. The electrode reactions produced (polarization) thus tend to be eliminated by this rapid reversal of current. The polarization is also diminished by increasing the area of the electrodes, which can be brought about to a very high degree by coating the plates with platinum black. The coating is effected by passing an electric current first one way and then the other between the two platinum electrodes immersed in a 3% solution of platinum chloride to which a trace of lead acetate has been added. The platinized plates thus obtained, while satisfactory for the investigation of concentrated solutions, have the power, however, of absorbing a certain amount of salt from the solutions and of giving it up again when water or more dilute solution is placed in contact with them. The measurement of very dilute solutions is thus made difficult, but if the plates are heated to redness after being platinized, a gray surface is obtained which possesses sufficient area for use with dilute solutions and yet does not absorb an appreciable quantity of salt.

For most purposes any convenient source of alternating current may be used. The currents from the secondary circuit of a small induction coil from which the condenser has been removed are satisfactory for measurements not requiring a high degree of accuracy.

With this arrangement the usual method of measuring resistance by means of Wheatstone's bridge may be adapted to the case of electrolytes. With alternating currents, however, it is impossible to use a galvanometer in the usual way. The galvanometer was therefore replaced by Kohlrausch with a telephone, which gives a sound when an alternating current passes through it. The most common plan of the apparatus is shown diagrammatically in fig. 4.

The electrolytic cell R_1 and a resistance box R_2 form two arms of a bridge. The current from the secondary circuit of a small induction coil is applied at the connecting points of R_1 and R_2 and at a sliding contact C which can be moved along the meter wire AB, whereby the resistances given by the lengths of wire r_1 and r_2 form the other two arms. Adjustment of the position of contact of C along the wire is made until no sound or only a minimum sound is heard in the telephone. This position is more sharply attained by connecting a variable air condenser parallel with the resistance box so as to balance the cell's residual polarization capacity. The resistance of the electrolyte is then given by the relation

$$R_1 = R_2 \frac{r_1}{r_2}$$

Although occasionally still used for approximate measurements, this older form of apparatus has been replaced by alternating-current Wheatstone bridges with fixed ratio arms and more elaborate reactance balancing means. The source of A.C. power is now usually some form of electronic oscillator and, to obtain greater sensitivity in the measurements, an electronic amplifier is used with a telephone or oscilloscope detector. The theory and

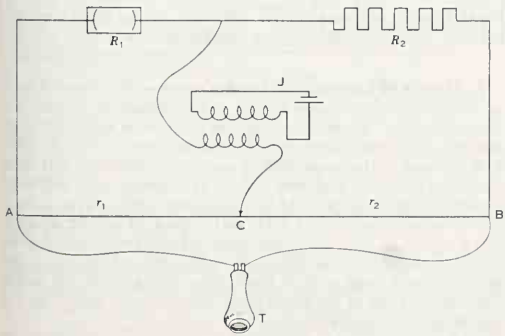


FIG. 4.—MEASUREMENT OF RESISTANCE BY WHEATSTONE'S BRIDGE AND TELEPHONE

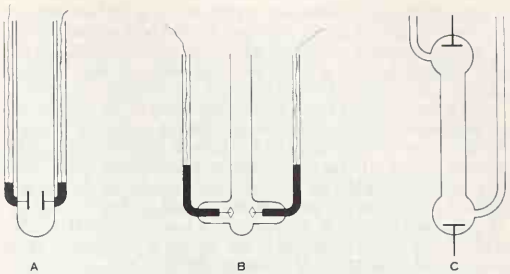


FIG. 5.—CELLS FOR MEASUREMENT OF RESISTANCE OF SOLUTIONS

design of precision bridges for measuring electrolytic solutions have been discussed by G. Jones and R. C. Josephs and by T. Shedlovsky.

Many forms of cell have been proposed: of those shown in fig. 5, (A) is a simple form suitable for dilute solutions, and is convenient for immersing in a thermostat. When this is used with currents of higher frequency, however, it has been found necessary for the highest accuracy to keep leads as far separated as possible, to avoid the capacity effect of parallel conductors. A type of cell which has been designed to meet this requirement is shown at (B). For measurements with stronger solutions requiring a greater distance between the electrodes, the modification illustrated by (C) is generally used.

Since the resistance changes about 2% for each degree of temperature change, the cell must always be immersed in a bath of known temperature and preferably thermostatically controlled. For accurate work, oil rather than water should be used so as to avoid A.C. current leakage errors as much as possible.

Alternative methods for measuring conductivity are based on the use of direct current with a view to overcoming errors introduced by self-induction, electrostatic capacity and the effect of polarization at the electrodes. In a procedure employed by E. Newbery, a modification was made of the usual method for measuring the resistance of a metallic conductor by determining the fall of potential between two points in a path of known length and cross section during the passage of a known current. The electrolyte is for this purpose contained in a tube of accurately determined dimensions, placed horizontally and connected at each end with vessels containing the respective electrodes. Standard electrodes (calomel, mercurous sulfate) make contact with the solution in the tube at two intermediate points and connect with an accurate dial potentiometer. The resistance of the electrolyte is given by Ohm's law from the measured potential difference and current. A method similar to that of Newbery was subsequently employed by E. D. Eastman. However, the direct current methods suffer from the limitation that they are applicable only to those electrolytes for which nonpolarizable electrodes can be obtained.

The units selected for expressing concentrations are conveniently represented by the electrochemical equivalent weight in grams in a litre of solution. This is defined as a normal solution (N). This equivalent concentration is identical with the molecular concentration for uni-univalent compounds. For compounds of bivalent and trivalent ions which possess two or three charges, the equivalent weight is one-half and one-third respectively of the molecular weight.

In order to avoid the necessity of an accurate measurement of the dimensions of conductivity cells, it is usual for determinations with different electrolytes to measure the resistance of a cell containing a solution of known conductivity; *e.g.*, 0.1 normal potassium chloride. Thus a factor, the cell constant, relating the resistance of the cell and the conductivity of the solution, is found and applied to all future measurements. The conductivity of the solvent used is subtracted from the observed conductivity of the solution. Conductivity water, made by twice distilling water in a block tin apparatus, has usually a specific conductance of about

0.1×10^{-6} reciprocal ohms which rises to 0.8×10^{-6} on exposure to clean air, caused mainly by traces of dissolved carbon dioxide. It has been usually considered advantageous to conduct the second distillation from alkaline permanganate solution. Ion exchange resins can also be used to obtain low-conductivity water.

For standardization purposes, measurements of the conductivity of potassium chloride solutions at different concentrations and temperatures were first made by Kohlrausch, Holborn and H. Diesselhorst in 1898. In the earlier determinations of the conductivity of potassium chloride, the normal concentration was defined as a solution containing 1 gram-molecular weight (taken as 74.59 g.) of potassium chloride in a volume of 1 l. of solution at 18° C.; while the litre itself is defined as the volume occupied by 1 kg. of water at the temperature of maximum density (4° C.).

The values obtained by some of the main workers in this field are given in Table IV, and represent the specific conductance of potassium chloride solutions in ohm^{-1} $cm.^{-1}$ at different concentrations and temperatures.

TABLE IV.—*Conductance of Potassium Chloride Solutions*

Concentration		Temp. ° C.	Specific conductance			
Normality	G. KCl in 1,000 g. water*		Kohlrausch, Holborn and Diesselhorst (1898)	Kohlrausch and Maltby (1900)	Newbery (1918)	Kraus and Parker (1922)
N	76.9247	18	0.09822	0.09827	..	..
..	..	25	0.11180	..	0.11247	..
N/10	7.4945	18	0.01119	0.011203	..	..
..	..	25	0.01288	..	..	0.0128988
N/100	0.74766	18	0.001225	0.0012243	..	..
..	..	25	0.001413	..	..	..

*Weights in air (density 0.0012) using brass weights.

For the higher degree of precision which has been attained by later workers it has been found necessary to define more accurately the conditions defining the units employed, and to express data in terms of a given weight of salt when dissolved in 1,000 g. of water (or solution), thus giving results which are unaffected by the value taken for the molecular weight and avoiding the complication arising in the definition and measurement of volume. The concentrations given in the second column expressed on this basis are calculated to correspond with the values employed by Kohlrausch.

In accordance with this distinction it has been found preferable to express normal concentration in terms of gram equivalents per cubic decimetre (cu. dm.) which is defined in terms of dimensions. Conversion from litres to cubic decimetres is made by multiplying by the factor 1.000027. When expressed on this corrected basis, the term demal is suggested to take the place of normal or molal concentrations.

The values given in the last column by C. A. Kraus and H. C. Parker apply to this new basis. Further accurate determinations with potassium chloride solutions have been made by H. C. Parker and E. W. Parker with the results given in Table V, in which the assumed molecular weight of the salt is taken as 74.553.

TABLE V.—*Conductance Values (Revised Concentrations)*

Concentration		Specific conductance at		
Normality	Weight (in air) of KCl in 1,000 g. water	0° C.	18° C.	25° C.
1.0 D at 0° C.	76.6276	0.065098	0.097700	0.111322
0.1 D at 0° C.	7.47896	0.0071295	0.0111036	0.0128524
0.01 D at 0° C.	0.746253	0.00077284	0.00122023	0.00140789

In determinations made by G. Jones and B. C. Bradshaw in 1933, it is preferable, in order to obtain greater accuracy in the preparation of the solutions, to express weights in terms of grams of KCl per 1,000 g. solution, also to correct weights for air buoyancy and record as weights in vacuum. The results for specific conductance on this basis are given in Table VI.

Applying corrections according to this new basis, Jones and Bradshaw calculated the concentrations employed by earlier workers to be those shown in Table VII.

The equivalent conductivity of a solute (Λ) is defined as the

TABLE VI.—*Conductance Values (Revised Concentrations)*

Concentration		Specific conductance at temperatures	
Normality on demal basis	G. KCl in 1,000 g. sol. in vacuum	18° C.	25° C.
D	71.1352	0.0097838	0.111342
0.1D	7.41913	0.0111067	0.0128560
0.01D	0.745263	0.0012205	0.00140877

specific conductance of the solution in ohm^{-1} $cm.^{-1}$ divided by its concentration in equivalents per cubic centimetre. As was shown by Kohlrausch, this quantity increases with decreasing concentration, reaching a limiting value Λ_0 at zero concentration; and from a study of the limiting values for a series of electrolytes, it was concluded that each ion makes its own characteristic contribution to the conductivity. According to this relation, known as the law of independent mobilities, $\Lambda_0 = \lambda_+^0 + \lambda_-^0$ where λ_+^0 and λ_-^0 are the contributions of the positive and negative ions respectively.

TABLE VII.—*Definition of Standard Solutions*

Normality on demal basis	Kohlrausch		Parker and Parker	
	G. KCl in 1,000 g. water (in air)	G. KCl in 1,000 g. solution (in vacuum)	G. KCl in 1,000 g. water (in air)	G. KCl in 1 cu. dm. solution (in air)
D	76.9153	71.3828	76.6276	74.5184
0.1D	7.4933	7.44037	7.47896	7.45184
0.01D	0.74756	0.746554	0.746253	0.745184

The limiting conductivity values are accordingly proportional to the mobilities (defined as velocities per volt per centimetre) of the ions, u_+ and u_-, in accordance with the relation $\Lambda_0 = \lambda_+^0 + \lambda_-^0 = k\ (u_+ + u_-)$. Since the mobility ratio $u^+/u^+ + u^-$ is equal to the ratio of the so-called ionic transport number, which can be determined experimentally, the individual ionic conductances can also be calculated by combining conductance and transference measurements.

The most accurate results which have been determined for the limiting equivalent conductances of some ions at 25° C. at infinite dilution are given in Table VIII together with the temperature coefficients.

TABLE VIII.—*Limiting Ion Conductances in Water at 25° C. and Approximate Temperature Coefficients*

$\lambda_t = \lambda_{25°}\ [1 + \alpha(t - 25)]$

Cation	$\lambda_{25°}$	α	Anion	$\lambda_{25°}$	α
K^+	73.52	0.0189	Br^-	78.4	0.0187
Na^+	50.11	0.0209	I^-	76.85	0.0186
H^+	349.82	0.0142	NO_3^-	71.44	0.0180
Ag^+	61.92	0.0197	HCO_3^-	44.48	..
Li^+	38.69	0.0226	OH^-	198	0.0160
NH_4^+	73.4	0.0192	$CH_3CO_2^-$	40.9	..
Tl^+	74.7	0.0187	$CH_3ClCO_2^-$	39.7	..
$\frac{1}{2}Ca^{++}$	59.50	0.0211	$CH_3CH_2CO_2^-$	35.81	..
$\frac{1}{2}Ba^{++}$	63.64	0.0206	$CH_3(CH_2)_2CO_2^-$	32.59	..
$\frac{1}{2}Sr^{++}$	59.46	0.0211	ClO_4^-	68.0	..
$\frac{1}{2}Mg^{++}$	53.06	0.0218	$C_6H_5CO_2^-$	32.3	..
$\frac{1}{3}La^{+++}$	69.6	..	$\frac{1}{2}SO_4^{--}$	79.8	0.0196
$\frac{1}{3}Co(NH_3)_6^{+++}$	102.3	..	$\frac{1}{3}Fe(CN)_6^{---}$	101.0	..
Cl^-	76.34	0.0188	$\frac{1}{4}Fe(CN)_6^{----}$	110.5	..

3. Theory of Electrolytic Conductance.—The classical ionic or electrolytic dissociation theory was proposed by Svante Arrhenius in 1884. He assumed that the decrease in equivalent conductance of an electrolyte with increasing concentration was due to a decrease in the proportion of conducting ions, but that their mobilities remained unaltered. Thus, he expressed the "degree of ionization," x, as the ratio of the equivalent conductance, Λ, at a given concentration, to the value extrapolated at infinite dilution, Λ_0; $x = \Lambda/\Lambda_0$. This theory appeared to hold fairly well for weak (poorly conducting) electrolytes such as acetic acid in water, but not for strong ones such as most salts, strong acids and bases. The difficulty lay in the assumption that the ionic mobilities are independent of concentration.

S. R. Milner (1913) attempted a quantitative mathematical calculation for the interionic, electrostatic force effects of the

ions on themselves, but his theoretical treatment led to serious mathematical difficulties. In 1923 P. Debye and E. Hückel formulated an interionic attraction theory in more useful form. A further improvement in the theory, as applied to electrical conductance, was presented by L. Onsager in 1927.

In the interionic attraction theory of electrolytes, the properties of the solutions are considered to be due to the interplay of electrostatic forces and thermal motion. The first of these tends to give the ions a definite arrangement, and the second tends to produce a random distribution. With these ideas it is possible to account for varying ion mobilities with changing concentration. The conductance variations of dilute solutions of strong as well as weak electrolytes can thus be satisfactorily explained.

As a result of electrostatic (coulomb) forces and thermal motion, any selected ion in a solution, a positive one for instance, will, on a time average, have more negative ions near it than if the distribution were purely random. This is known as the ion atmosphere of the selected ion. Its effects on electrolytic conductance are twofold, both tending to decrease the ionic mobilities with increasing ion concentration. They are the so-called "electrophoretic effect" and "time of relaxation effect."

When an electric potential is applied, the ion atmosphere, which has a net charge equal but opposite in sign to that of the central ion, tends to move with its associated solvent molecules in a direction opposite to that of the central ion. This produces a slowing up of the central ion, which, in effect, moves against a countercurrent of solvent amounting to an increase of viscous drag. Because of its analogy to the retarding effect of the solvent on the motion of colloidal particles in an electric field, this phenomenon has been called the electrophoretic effect.

When no external electric field is applied on a solution, the ion atmosphere is spherically symmetrical with its electrical centre of gravity at the central ion. If, however, the ion is made to move by an external field, the ion atmosphere must also move to adjust for this charge. This takes place rapidly (during the short time of relaxation) but not instantaneously, so that there results a distortion of the ion atmosphere, whose electrical centre of gravity is shifted to a point behind the central ion, producing an electrostatic retarding force which slows up its motion. This phenomenon is called the time of relaxation effect.

In very dilute solutions the "thickness" of the ion atmosphere decreases with the square root of the ionic concentration and with ionic charge. The retarding effect on ionic mobility due to both the electrophoretic and time of relaxation effects turns out to be inversely proportional to this thickness, which is usually of the order of angstroms. Moreover, the former is proportional to the ionic mobility at infinite dilution, and the latter to the fluidity of the solvent.

The quantitative conductance equation for very dilute electrolytic solutions, derived by Onsager, based on these considerations is

$$\Lambda = \Lambda_0 - (\alpha\Lambda_0 + \beta)\sqrt{C}$$

where C is the ion concentration, and the constants α and β depend on the nature of the solvent and the temperature. For uni-univalent electrolytes:

$$\alpha = \frac{8.20 \times 10^5}{(DT)^{3/2}}$$

and

$$\beta = \frac{82.4}{(DT)^{1/2}\eta}$$

D being the dielectric constant at the absolute temperature T, and η being the viscosity.

It has been shown that the Onsager equation is the correct limiting expression for the conductance of strong electrolytes. However, positive deviations from the theoretical conductance values occur as the concentration is increased (fig. 6). These have been accounted for by an empirical extension of the Onsager equation, proposed by Shedlovsky (1934),

$$\Lambda_0 = \Lambda'_0 - AC - BC \log C + DC^2$$

in which $\Lambda'_0 = (\Lambda + \beta\sqrt{C})/(1 - \alpha\sqrt{C})$. In many cases, such as for most uni-univalent salts and strong acids in water up to a concentration of about 0.1 N, the last two terms can be omitted. Later (1957) R. M. Fuoss and Onsager accounted for these positive deviations on theoretical grounds, taking the size of the ions in consideration. R. H. Stokes and R. A. Robinson (1954) proposed a semiempirical equation which appears to be valid for conductance at quite high concentrations.

It was observed by M. Wien (1927) that the equivalent conductance of electrolytes appeared to be increased at high potential gradients. This effect is in accord with the interionic attraction theory. When the velocity of ions due to high field strengths becomes sufficiently great, the ion atmospheres hardly have time to form, with the result that both the electrophoretic and time of relaxation effects become correspondingly diminished. In accordance with the Wien effect observations, the theory thus predicts an asymptotic approach of Λ *v.* field strength curves to the Λ_0 values. Also, as would be expected, the relative increase in conductance is most marked in the cases of relatively concentrated solutions of high-valence electrolytes, in which the interionic effects are most marked.

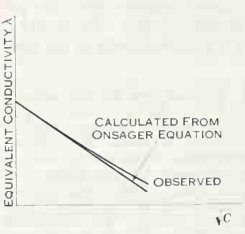


FIG. 6.—RESULTS OF MEASURED AND CALCULATED VALUES FOR POTASSIUM CHLORIDE SOLUTIONS IN WHICH THE GRAPH RANGES FROM CONCENTRATIONS OF 0 TO 0.005 N

An unexpected experimental result was the large Wien effect shown by weak acids and bases. It appears, therefore, that for such weak electrolytes, high field strengths tend to produce increased ionization.

In order to avoid complications arising from polarizing effects at the electrodes, it is customary to carry out conductance measurements with alternating current. It was shown by Debye and H. Falkenhagen (1928) that when sufficiently high frequencies of current (more than 5 m.c.) are used, the equivalent conductance is increased, approaching a limiting value somewhat lower than Λ_0. The theoretical explanation of this phenomenon lies in the fact that at high frequencies the ion atmosphere does not fail to form, as in the Wien effect, but has insufficient time to distort fully. Consequently, the time of relaxation effect tends to vanish, although the electrophoretic effect retains its full influence. Thus, as frequency is increased, Λ approaches $\Lambda_0 - \beta\sqrt{C}$ rather than Λ_0.

In solvents of high dielectric constant, such as water, strong electrolytes appear to be completely ionized. On the other hand, in solvents of low dielectric constant no strong electrolytes exist, and large negative deviations from the Onsager slopes of conductance are observed.

It is generally assumed that only a portion of a weak electrolyte is in the form of free ions, while the rest is undissociated; and that an equilibrium exists between the two forms of the solute in accordance with the law of mass action. However, the undissociated solute need not consist of stable molecules bound by chemical or quantum forces. They may be ion pairs, as suggested by N. Bjerrum. The electrostatic force between a pair of ions of opposite charge which have approached each other closely tends to hold them together as a dipole. Now, if the average kinetic energy of the solvent molecule is greater than the potential energy of the ion pair, it will soon be disrupted. This is probably the case for strong electrolytes in solvents with high dielectric constants.

If, on the other hand, the average kinetic energy of the solvent molecules is less than the mutual potential energy of the ion pair, it will exist until it is struck by an exceptionally fast solvent molecule. This viewpoint predicts that there should be relatively more ion pairs when the dielectric constant is low, when the ions are small, when the charges are large and when the temperature is low.

According to Arrhenius' dissociation theory, weak electrolytes were found to follow the Ostwald dilution law:

$$\Lambda = \Lambda_0 - \frac{cA^2}{K\Lambda_0}$$

an expression derived from the mass action law and the assumptions (1) that the ions are perfect solutes; and (2) that the ratio Λ/Λ_0 represents the degree of dissociation. However, the first assumption is true only as a very rough approximation in dilute solutions; and the second assumption can be valid only if the mobilities of the ions do not change with the concentration, which is contrary to fact. Corrections for nonideal behaviour of ions (activity coefficients) and for the decrease in ionic mobilities with increase in concentration have been made through the use of the interionic attraction theory of Debye, Hückel and Onsager. For example, MacInnes and Shedlovsky (1932) analyzed their conductance measurements on aqueous acetic acid solutions, and computed the ionization constant of this acid.

The strong and weak electrolyte theory briefly outlined above accounts quantitatively for ionic conductivity in relatively dilute aqueous and nonaqueous solutions. However, the quantitative theory of concentrated ionic solutions and of fused salts is not yet available, as is also the case with colloidal solutions. The electrical conductance in liquid metals, which is electronic, falls in the category of metallic conductance.

BIBLIOGRAPHY.—F. Kohlrausch, L. Holborn and H. Diesselhorst, "Leitvermögen von Elektrolyten," *Ann. Phys., Lpz.*, 64:417–440 (1898); F. Kohlrausch and M. E. Maltby, "Conductivities of Aqueous Solutions of Alkali Chlorides and Nitrates," *S. B. Preuss. Akad. Wiss.*, 36:665 (1899); *Wiss. Abh. Phys.-Tech. Reichsanst.*, 3:180–210 (1900); *J. Amer. Chem. Soc.*, 44:2426 (1922); E. Newbery, "A New Method for the Determination of Conductivity," *Trans. Chem. Soc.*, 113:701 (1918); E. D. Eastman, "Conductivity and Frequency," *J. Amer. Chem. Soc.*, 42:1648 (1920); C. A. Kraus and H. C. Parker, "The Calibration of Cells for Conductance Measurements," *J. Amer. Chem. Soc.*, 44:2422 (1922); D. A. MacInnes and E. R. Smith, "A Study of the Moving Boundary Method for Determining Transference Numbers," *J. Amer. Chem. Soc.*, 45:2246 (1923); P. Debye and E. Hückel, "Zur Theorie der Elektrolyte," *Phys. Z.*, 24:185 (1923); H. C. Parker and E. W. Parker, "The Calibration of Cells for Conductive Measurements," III, "Absolute Measurements on the Specific Conductance of Certain Potassium Chloride Solutions," *J. Amer. Chem. Soc.*, 46:313 (1924); L. Onsager, "Zur Theorie der Elektrolyte," *Phys. Z.*, 27:388 (1926), 28:277 (1927); P. Debye, "Conductivity of Strong Electrolytes in Dilute Solutions," *Trans. Faraday Soc.*, 23:334 (1927); L. Onsager, "A Revision of the Conductivity Theory," *Trans. Faraday Soc.*, 23:341 (1927); G. Jones and R. C. Josephs, "Electrolytic Conductivity Bridge," *J. Amer. Chem. Soc.*, 50:1049 (1928); T. Shedlovsky, "Bridge for Electrolytic Conductance," *J. Amer. Chem. Soc.*, 52:1793 (1930); D. A. MacInnes and L. G. Longsworth, "Transference Numbers by the Method of Moving Boundaries," *Chem. Rev.*, 11:171 (1932); T. Shedlovsky, "The Electrical Conductivity of some Uni-univalent Electrolytes in Water at 25°," *J. Amer. Chem. Soc.*, 54:1411 (1932); D. A. MacInnes and T. Shedlovsky, "The Ionization Constant of Acetic Acid from Conductance Measurements," *J. Amer. Chem. Soc.*, 54:1429 (1932); G. Jones and B. C. Bradshaw, "The Measurement of the Conductance of Electrolytes," *J. Amer. Chem. Soc.*, 55:1780 (1933); D. A. MacInnes, T. Shedlovsky and L. G. Longsworth, "The Conductance of Aqueous Solutions of Electrolytes and the Interionic Attraction Theory," *Chem. Rev.*, 13:29 (1933); G. Jones and D. M. Bollinger, "The Measurement of the Conductance of Electrolytes," VII, "On Platinization," *J. Amer. Chem. Soc.*, 57:280 (1935); R. M. Fuoss and L. Onsager, "Conductance of Unassociated Electrolytes," *J. Phys. Chem.*, 61:668 (1957); R. M. Fuoss, "Ionic Association," *J. Amer. Chem. Soc.*, 79:3301 (1957); "Ionic Conductivity," *Int. Crit. Tab.*, vol. vi, p. 230; H. J. M. Creighton, *Electrochemistry* (1924); W. M. Clark, *The Determination of Hydrogen Ions* (1928); F. H. Newman, *Electrolytic Conduction* (1930); C. W. Davies, *The Conductivity of Solutions* (1933); H. Falkenhagen, *Electrolytes* (1934); S. Glasstone, *Electrochemistry of Solutions*, 2nd ed. (1937); F. H. MacDougall, *Thermodynamics and Chemistry*, 3rd ed. (1939); D. A. MacInnes, *The Principles of Electrochemistry* (1939); S. Glasstone, *Text-book of Physical Chemistry* (1940); I. M. Kolthoff and H. A. Laitinen, *pH and Electro Titrations*, 2nd ed. (1941); H. T. S. Britton, *Hydrogen Ions* (1942); H. S. Harned and B. B. Owen, *The Physical Chemistry of Electrolytic Solutions* (1943); R. A. Robinson and R. H. Stokes, *Electrolyte Solutions* (1955). (J. N. P.; T. Sy.)

III. CONDUCTION IN GASES

1. Introduction.—Electricity can be conducted through a gas if, and only if, free electric charges exist in it. In a normal environment, gases are composed of neutral atoms and molecules, are not electrical conductors, and are sometimes used for their insulating properties. The important phenomena, therefore, are those associated with the appearance and motion of free charges; these phenomena are the subject of this section.

Most commonly, the charges arise from subjecting the gas to conditions which lead to ionization of some of its constituents. In this process, electrons are removed from a fraction of the atoms or molecules (hereinafter called atoms unless clarity requires otherwise), leaving the atoms as positively charged ions. The words "normal environment" mentioned above refer, of course, to conditions on the surface of the earth. From the point of view of the universe as a whole, however, such conditions are highly abnormal. In fact, all but a small fraction of the universe (the cool planets) consists of highly ionized gas in various states of gravitational contraction (stars) and rarefaction (the tenuous medium between them).

The electrical conductivity of a gas may range from practically zero to values much higher than those of metallic conductors. Such high conductivities exist in stars, and can be produced artificially under special laboratory conditions. A very useful intermediate degree of ionization is that wherein the conductivity is about 10^{-6} (one millionth) that of copper. Fluorescent lamps, "neon" signs and many other useful devices operate in this ionization range.

Electrical conduction in gases has been recognized almost as long as electricity itself. Probably most familiar are the experiments of Benjamin Franklin, who conducted currents from charged thunderclouds by means of a kite and conducting string. Some other experimenters were not so fortunate, having neglected to insulate themselves adequately. In the latter part of the 19th century, many phenomenological studies were made of electrical discharges taking place in glass tubes containing gas at low pressure; and many beautiful and more or less reproducible glowing regions of gas were observed. Such tubes were not fundamentally different from modern neon signs. These phenomena could be studied spectroscopically. Since the light arises from various processes involving individual atoms, conducting gases were powerful tools for exploring the physics of the atom. For just this reason, understanding of the conduction mechanisms in gases and of the nature of the atom developed hand in hand through the last part of the 19th and into the early 20th centuries.

In 1897, J. J. Thomson, performing experiments on a partly evacuated tube, confirmed the existence of electrons as light, negatively charged, freely moving particles. The nature of the atom itself, and just how the electrons and positive charges were distributed in it, was still something of a mystery. After much further work, Ernest Rutherford brilliantly demonstrated in 1911 that the atom consists of an exceedingly small and dense positive nucleus, surrounded by a cloud of electrons. Ionization of an atom amounts to removing one (or more) of these orbiting electrons, leaving a net positively charged atom, which is called the ion.

The electrons are bound to the atom by a certain amount of energy, which is very conveniently measured in terms of electron volts. This unit is the change of energy experienced by a particle having one electronic charge ($e = 1.602 \times 10^{-19}$ coulomb) in passing through a potential difference of one volt. Whether it is a gain (+) or loss (−) of energy depends on the sign of the product of the charge (+ or −) and the direction of motion (+ or −) with respect to that of decreasing potential, but the absolute magnitude is the same. One electron volt (ev) is equal to 1.6×10^{-12} erg. Depending on the atom, the energy required to remove the outermost electron (ionization potential) ranges from about 5 to almost 25 ev. The inner electrons are bound with considerably higher energies.

The light emitted by the glow discharges mentioned above arises from a process different from ionization, but related to it. An electron may be displaced from its normal atomic orbit by absorbing energy less than the ionization energy. The atom is then in a so-called excited energy state. The possible excited states of an atom are discrete and not continuous, as is shown by the quantum theory. The displaced electron then can and usually does return to its normal orbit in a very short time. It gives up its excess energy by emitting a photon of electromagnetic radiation (light, if it be within the visible spectrum) of fixed frequency depending

on the atom and excited state. Somewhat inverse to ionization is recombination, when a free electron falls into the empty electron orbit of a positive ion, making a neutral atom again. Here, there is a characteristic light emission also.

How is the energy transferred to the gas atoms, so that they may become excited or ionized? It is generally by collision with energetic particles, which in principle could be other gas atoms, ions, electrons or even photons. Heating the gas increases the energy of the atoms, but this is usually not sufficient, for one electron volt corresponds to a temperature of about 7,700° K. (nearly 14,000° F.). The effect, however, does take place under some circumstances, as will be mentioned below. Far more efficient is the application of voltage across the gas, which accelerates charged particles chancing to be in the gas, which then ionize neutral atoms by impact. Electrons are most efficient in this respect.

In continuing the subject, we will consider first a further number of basic processes, then deal with the breakdown of a gas into a conducting medium. Following that, the glow and arc discharges will be described. Finally, brief attention will be given to the physical nature of the ionized medium itself—commonly called the plasma.

2. Basic Processes.—The processes of atomic ionization and excitation by impact of other atoms, ions or (most important) electrons, and by recombinations were all described in the introduction. Some other important ones will now be taken up.

Residual Ionization.—Any appreciable volume of gas almost invariably contains a few ions and electrons, and exhibits an exceedingly small conductivity. The ionization arises from the passage of cosmic rays from outer space, gamma rays and a small number of other highly energetic particles from radioactive sources on the surface of the earth—uranium, thorium and so on. The effect is vividly demonstrable in a cloud chamber (*q.v.*), where the ionization along the tracks of the particles is made visible. In addition, light can release electrons by the photoelectric effect from metal surfaces adjacent to the gas. As a result one can, for example, draw the minute current of perhaps 10^{-12} amp. from an unlit fluorescent lamp by application of a few volts to its terminals. In spite of the small amount, this residual ionization is important, for it provides an initial source of electrons necessary for breakdown of the gas.

Particle Collisions.—Atoms collide with each other and with ions and electrons. Typical collision cross sections of atoms lie in the range 10^{-15} to 10^{-16} cm.^{2} As a result, in ordinary room air, for example, the mean free path of an atom between successive collisions is about 10^{-4} cm. It is convenient to measure gas pressure p in terms of millimetres of mercury, (symbol mm.Hg; atmospheric pressure = 760 mm.Hg). The mean free path l is inversely proportional to p, so that

$$l(\text{cm.}) = 1/P_c p \tag{1}$$

The constant of proportionality P_c is different for different gases, and depends also on the velocity of the colliding particles. Especially important are collisions of electrons with gas atoms. Many such electron-gas collisions were studied by R. B. Brode and others around 1930; and the values of P_c obtained generally lie in the range 5–100 (cm. × mm.Hg)$^{-1}$ for common gases, but are as much as ten times higher for metal vapours.

Velocity and Temperature of Particles.—At any temperature T, gas atoms, electrons and ions are in motion; and the average velocity v of each type of particle is characteristic of its mass m and the temperature measured from absolute zero. In the kinetic theory of gases, it is shown that, on the average,

$$\frac{1}{2}mv^2 = \frac{3}{2}kT \tag{2}$$

where k is Boltzmann's constant. Here the quantity $mv^2/2$ is the average kinetic energy, which is proportional to the temperature. Incidentally, this quantity divided by the electronic charge e is the average energy measured in units of electron volts, mentioned in the *Introduction*, above.

For mixtures of particles in equilibrium, the temperature is the same for each kind; therefore the energy is the same. The velocity v is higher for lighter particles and lower for heavy ones. For example, air molecules at room temperature travel at about 4×10^4 cm. per second; and electrons at the same temperature travel at about 10^7 cm. per second, or some 250 times faster.

While this is true on the average, no particle maintains the average velocity—at least not for very long. In temperature equilibrium, kinetic theory shows that a so-called maxwellian distribution of velocities exists. As an example, fig. 7 shows the distribution in velocity of electrons whose average energy is one electron volt (about 7,700° K.). The distribution is proportional to the quantity

$$(m/2kT)^{3/2}v^2e^{-\frac{mv^2}{2kT}}$$

The principal things to notice are that many particles have velocities near the average value, and that a small fraction go very much faster—*i.e.*, have much higher energy. In equilibrium, all directions of travel are equally likely.

The average time between collisions is the mean free path divided by the average velocity v. For atoms in room air, the time is about 2×10^{-9} sec. For electrons, the time would be much shorter, because the velocity is much higher. This collision time, t_c, is of course inversely proportional to the pressure, from equation (1).

Diffusion.—Consider a number of, for instance, electrons in a gas; let them be concentrated near a particular point in the gas; then release them so that they are free to move. Because of their thermal velocity and the effect of collisions with gas atoms, they travel in complicated paths; but on the average, the originally compact group will be smeared out in space after a little while. The process is called diffusion; if we denote by r the distance from the origin an electron has traveled, kinetic theory shows that the average value of r is

$$r = \sqrt{6Dt} \tag{3}$$

Here, t is the time, and D is the diffusion coefficient, where

$$D = \frac{1}{3}lv = v^2t_c/3 \tag{4}$$

and the quantity is an average for all electrons. From equation (3) one notes that the average distance to which a particle diffuses is proportional to the square root of the time. From equation (4) one notes that the longer the mean free path or the higher the velocity, the faster is the diffusion process.

Energy Gain From an Electric Field.—An electric field E is by definition the rate of change of voltage with distance. That is, in derivative form, $E = -dV/dx$, where x is linear distance. Note that E has a direction which is the same as that of decreasing voltage—implied by the minus sign. Thus, for example, for a voltage applied between two parallel metallic plates a distance d apart, $E = V/d$ in magnitude.

The force on a charge q in a field E is qE, and the particle moves

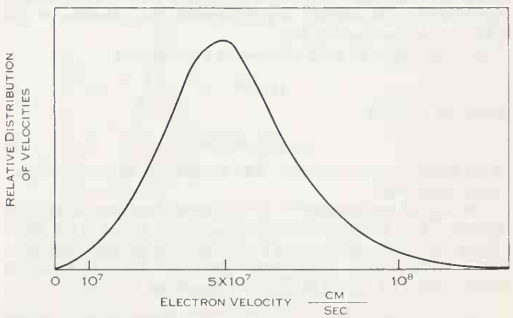


FIG. 7.—THE MAXWELLIAN DISTRIBUTION IN VELOCITY OF ELECTRONS WHOSE AVERAGE ENERGY IS ONE ELECTRON VOLT

with an acceleration qE/m, where m is again its mass. Now let some gas be present in the space. The collisions of the charge with gas atoms change the accelerated motion in the field direction into random motion. Thus after each collision, the charge is accelerated anew. To be sure, the particle gradually gains velocity this way, for between each collision it picks up a little extra velocity from the action of the field. From the discussion of *Velocity and Temperature of Particles*, above, it can be seen that this increase in random velocity is equivalent to raising the energy (or temperature) of the particle. Thus the application of an electric field raises the energy of the charged particles.

This heating effect does not go on indefinitely. Each time the charge collides with a gas atom, the atom recoils and takes away some of the charged particle's extra energy. Thus after a short time, a new quasi-equilibrium situation arises, in which the energy of the charged particles is some fixed amount higher than the energy of the neutral atoms. For light charges and heavy atoms, the effect is very pronounced, for the light particles cause very little recoil of the heavy ones.

The Variable E/p.—As discussed in the previous section, a charge moving in a field E loses its excess energy as it collides with gas atoms, the amount depending on the energy it has and on the nature of the collision. Since it gains energy only between collisions, it must reach an equilibrium energy which is determined solely by the amount of energy picked up per mean free path. But for any charge and gas combination, this unit of energy must be determined by the product El or, through equation (1), by the more conveniently measureable quantity E/p. Thus one may write

$$\text{average energy} = f(E/p) \tag{5}$$

where f is the particular functional dependence, and depends on the nature of the charge and gas under study. Since gas conduction phenomena depend on the energy of the charged particles, the result is a very important one. It states that the ratio of electric field and gas pressure is the important variable, rather than either one separately.

Mobility and Conductivity.—In addition to the effect of the applied field in raising the random velocity (or energy) of charged particles, as discussed above, there is also the effect of the field in giving the charges a drift velocity. Suppose that after each collision, the charge undergoes an acceleration qE/m in the field direction, which persists for the collision time t_c. It moves a distance $qEt_c^2/2m$, and its average velocity in the field direction would be this distance divided by the time, or $qEt_c/2m$. This simple derivation is almost correct, but neglects some of the physics; for example, the charge does not start from rest each time. The correct expression is twice as large: the drift velocity v_μ is given by

$$v_\mu = \frac{qt_cE}{m} = \frac{qlE}{mv} \tag{6}$$

where t_c is suitably averaged over the moving charges. The quantity v in (6) is the random velocity (a function of energy) and is generally much larger than v_μ. Thus the charges still travel in every direction, but with a slight preference for the forward one. Note that v_μ is a function of E/p, because the quantities El and v in (6) are functions of E/p.

It is convenient to write equation (6) in the form

$$v_\mu = \mu E \tag{7}$$

where the quantity

$$\mu = qt_c/m = ql/mv \tag{8}$$

is by definition the mobility. The mobility is inversely proportional to pressure.

The sign of the mobility (+ or −) arises from the charge q in equation (8). Thus positive ions (charge $+e$) travel in the direction of the field, and electrons ($-e$) travel in the opposite direction. An interesting result, first proved by Albert Einstein, is demonstrated by taking the ratio of equations (4) and (8):

$$\frac{D}{\mu} = \frac{mv^2}{3e} \tag{9}$$

From equation (2), and the sentences immediately following it, one sees that the ratio of diffusion coefficient to mobility is equal to two-thirds of the average energy. This result is strictly true for a maxwellian velocity distribution (fig. 7) and is approximately true for many other practical velocity distributions.

Suppose there are n such charged particles per unit volume of gas, each with charge q. The current per unit area, J, is the product of the charge density, charge per particle, and drift velocity, or

$$J = nqv_\mu \tag{10}$$

From equation (7), one can define the conductivity σ, which is

$$\sigma = \frac{J}{E} = nq\mu \tag{11}$$

Note that the conductivity is always positive because q and μ have the same sign. Usually, in an ionized gas, both positive ions and electrons are present. If the two types of particles do not interact with each other, the total conductivity is the sum of that due to each type:

$$\sigma = n_+q_+\mu_+ + n_-q_-\mu_- \tag{12}$$

Here, the plus and minus subscripts apply to the ions and electrons respectively.

A few illustrative numbers are now in order. The drift velocity of electrons in various gases was extensively measured by many experimenters, *e.g.*, R. A. Nielsen and N. E. Bradbury, particularly during the period 1935–40. Such experiments are generally done by timing the passage of a group of electrons which travel down a gas-filled tube under the action of the electric field. In air, for example, at $E/p = 15$ v./cm. × mm.Hg (field strength about 11,000 v./cm. in room air), the drift velocity is 7×10^6 cm./sec. This corresponds to a mobility $\mu_- = 4.7 \times 10^5$ (cm.²/v. sec.) at a pressure of one millimetre of mercury. In neon, at $E/p = 1$ v./cm. × mm.Hg, the electron drift velocity is about 1.4×10^6 cm./sec. By definition, this is also the mobility at $p = 1$ mm.Hg. For other gases, the numbers are similar. In most of these cases, the drift velocity is roughly proportional to E/p—that is, at constant pressure, proportional to E. The conductivity (equation [11] or [12]) is therefore almost independent of electric field, and the current in the gas follows Ohm's law. In general, more complicated conducting gas tubes do not follow Ohm's law, as will be illustrated later.

The drift velocity (or mobility) of ions is measured in much the same way as that of electrons, and by pulsed techniques first developed by J. A. Hornbeck. For nitrogen positive ions in nitrogen at $E/p = 30$ v./cm. × mm.Hg, $v_\mu = 6 \times 10^4$ cm./sec.; or $\mu_+ = 2 \times 10^3$ (cm.²/v. sec.) at $p = 1$ mm.Hg. In neon at $E/p = 10$, $v_\mu = 3 \times 10^4$ cm./sec.; or $\mu_+ = 3 \times 10^3$ at $p = 1$ mm.Hg. These mobilities are much lower than those for electrons. One concludes, then, that if there are approximately equal numbers of free electrons and ions present, nearly all the conductivity and current result from the electrons.

In this brief review, space considerations preclude mention of most of the individual contributors to knowledge of gaseous conduction. In ionic mobility, early theoretical work of P. Langevin (1905) and experiments of A. M. Tyndall and L. B. Loeb (about 1930) were noteworthy. Loeb also made important studies of negative ions (*see* below).

Attachment.—In many gases, free electrons tend to attach to neutral atoms, forming reasonably stable negative ions. The effect in gases commonly used for their conducting properties is slight (hydrogen, mercury, neon, etc.) and can be ignored. But in some gases (the halogens, oxygen and others), the effect is very pronounced. The mobility of negative ions is low, and is comparable to that for positive ions. It has been seen from the preceding discussion that the electrons and not the ions reach high energies when an electric field is applied; and that the free electrons make most of the contribution to the conductivity. Thus it can be expected that electron attaching gases will be very hard to break down electrically, and will have very low conductivity; this is so. The saturated sulfur-halogen gas compounds (*e.g.*, Freon, sulfur hexafluoride) attach electrons very well, and furthermore are not very

active chemically. They are therefore used as insulating media in high-power radar systems and the like.

Secondary Electron Emission.—This process is primarily one of surfaces, but is very important to the conduction mechanism of gas-filled devices. The photoelectric effect has already been mentioned in *Residual Ionization*, above, whereby light causes electron emission. The light can come from an external source, but may come from the conducting (luminous) gas itself.

In addition, and often more important, positive ions (and sometimes excited atoms) can also eject secondary electrons from metal surfaces which they strike. The energy for the process arises from the ionization energy of the ion, which it carries to the surface. The ion is neutralized by an electron from the metal, and there is a probability γ_i of a second electron being emitted into the gas. Since energy equal to the work function of the metal is required to remove each electron, the process is possible only if the ionization potential exceeds twice the metal work function. Thus all possible ion and metal combinations are not active in this way. Sodium ions, for example, will not emit electrons from most metals, but will do so from barium-covered surfaces. In the mid-1950s H. D. Hagstrum made a definitive study of γ_i for various clean metal surface-rare gas ion combinations. Hagstrum found that the highest values of γ_i are obtained from ions having the highest ionization potential ($\gamma_i \approx .3$ for helium ions on pure molybdenum; $\gamma_i \approx .022$ for xenon ions), and that foreign atoms on the metal surface produce many anomalous effects. In addition to this mechanism of secondary electron emission, very fast ions can also eject electrons from a metal by virtue of the kinetic energy they dissipate on the surface.

3. Processes Leading to Voltage Breakdown.—We are now in a position to describe the prebreakdown phenomena occurring in a gas. Take a specific example, consisting of a glass tube containing two metallic electrodes a distance d apart. Ideally, let them be disks whose diameter is large compared to the spacing d. Then a voltage V will be applied, giving a field $E=V/d$ between the plates. Let there be gas inside at pressure p mm.Hg.

Fig. 8 shows the voltage across such a tube as a function of current on a logarithmic scale. The specific numbers are typical of neon gas in a tube with plate spacing 2 cm. and pressure about 1 mm.Hg, but the results are similar in other situations. Note that the current scale extends over a very large range—a factor of ten to the sixteenth power (10^{16}). The phenomena occurring in each of the regions marked on the figure will be described.

The Ionization Chamber Region.—With no voltage applied, the few charges produced from residual gas ionization diffuse out to the solid surfaces. The numbers of negative and positive charges arriving at each point are on the average equal, and no current is recorded. Now let a very small voltage be applied (point A of fig. 8). A small excess of positive charges arrives by mobility at the negative electrode (the cathode), and a small excess of negative charges arrives at the positive electrode (the anode). As the voltage is increased, the current increases approximately proportional to voltage, as the effect of mobility tends to overcome the random diffusion (region A–B). Note that the curve A–B would be nearly straight if the current were plotted linearly instead of logarithmically. At point B, essentially all the charges are collected, and a further increase of voltage to point C does not affect the current appreciably. A current of this nature is sometimes called a dark discharge.

The device as operated at or near point B is an ionization chamber, for the current is a measure of ionization produced by some external source. For example, increasing the source intensity by a factor of 100 would yield the current A′–B′ . . . etc., 100 times greater. The device is useful as a radiation monitor. Sometimes the voltage is applied from a condenser which has been previously charged. It then slowly discharges through the tube at a rate depending on the ionization intensity, and the charge lost over a period of time is a measure of the total irradiation over that period.

The Townsend Region.—As the voltage is raised, E/p increases, and so does the energy of any electrons present. At point C, an important new effect appears—a few electrons reach the ionization potential of the gas, and can thereby multiply themselves by ionizing a few gas atoms. Suppose now that an electron leaves the cathode and drifts toward the anode under the influence of mobility, and has α ionizations per unit distance. Then the new electrons can gain energy in their travel toward the anode, ionize gas atoms, and so on. The effect is to create an electron avalanche. In a distance x one initiating electron leads to the formation of a total of $e^{\alpha x}$ electrons, where $e = 2.718\ldots$ is the base of natural logarithms. Thus an electron current i_0 leaving the cathode becomes a current i at the anode, where

$$i = i_0 e^{\alpha d} \tag{13}$$

The growth factor α is called the Townsend ionization coefficient, named for J. S. Townsend who made the first serious measurements of α around 1920. The phenomenon is called a Townsend discharge (region C–D).

Just as E/p is a more universal parameter than E or p alone, α/p (a measure of the chance of ionization on one collision or mean free path) is a more generally useful parameter than α itself: α/p is a function of E/p. This is shown in fig. 9 for hydrogen gas. One notes that at low E/p, α/p is very small, and its relative value changes rapidly. The principal reason for this is that at low E/p,

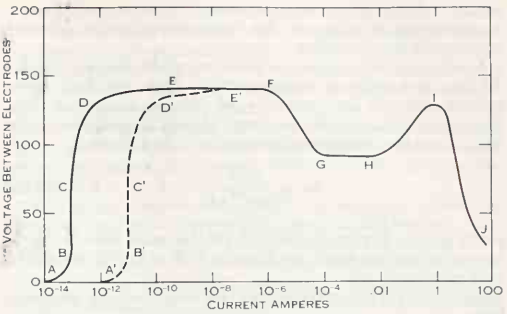


FIG. 8.—THE VOLTAGE *v.* CURRENT (ON LOGARITHMIC SCALE) OF A TYPICAL GAS TUBE A FEW CENTIMETRES IN SIZE AND FILLED WITH NEON AT A PRESSURE OF A FEW MILLIMETRES OF MERCURY. A-B-C = IONIZATION CHAMBER REGION; C-D-E = TOWNSEND REGION; E-F = BREAKDOWN; F-G = BREAKDOWN-GLOW TRANSITION; G-H = NORMAL GLOW; H-I = ABNORMAL GLOW; I-J = GLOW-ARC TRANSITION

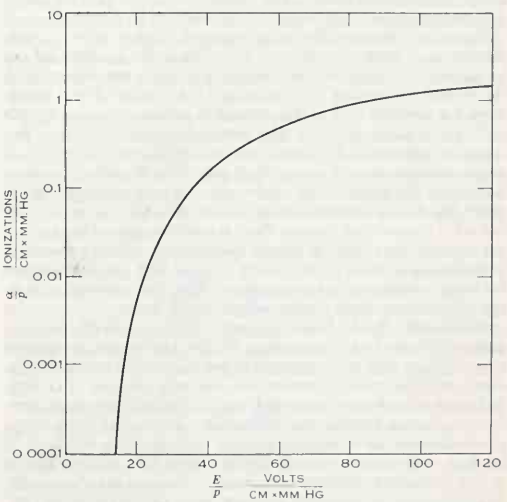


FIG. 9.—THE IONIZATION COEFFICIENT α/p PLOTTED LOGARITHMICALLY *v.* THE RATIO E/p OF ELECTRIC FIELD STRENGTH E TO PRESSURE p, FOR HYDROGEN GAS

very few electrons have sufficient energy to ionize (only those at the extreme right in fig. 7); but their relative number increases rapidly with increasing E/p.

Effect of Secondary Electron Emission, and Breakdown.—So far, the ions produced have been neglected. They drift back to the cathode, and each has a chance γ_i of releasing an electron there. This increases the current even more. An initial photoelectric current i_0 from the cathode can be shown to yield a current

$$i = \frac{i_0 e^{\alpha d}}{1-\gamma_i(e^{\alpha d}-1)} \qquad (14)$$

at the anode. For small values of $e^{\alpha d}$, the extra term in the denominator has little effect, but in region D–E of fig. 8 it becomes increasingly important.

Equation (14) illustrates the fact that breakdown can take place. When

$$\gamma_i(e^{\alpha d}-1)=1 \qquad (15)$$

(point E), the denominator tends to zero, and the current can increase without limit (region E–F). The discharge is self-sustaining, and there is a simple and important physical interpretation. Each electron leaving the cathode makes $e^{\alpha d}-1$ ions, each of which has a chance γ_i of re-creating the original electron at the cathode. Equation (15) (and all other breakdown criteria) then expresses the fact that each charge creates just one new one before being lost from the tube. At point F new effects appear, to be described in *The Glow Discharge*, below. The constant voltage of region E–F is the breakdown voltage and cannot be exceeded, provided any electrons are present in the gas to start the avalanche process.

In this description, the ions were presumed to be (and usually are) the secondary agents. In fact, excited atoms as well as ions are also created, and the radiation from them (usually ultraviolet) can act photoelectrically at the cathode in the same way. Equations similar to (14) and (15) can be developed for such cases.

The common Geiger counter tube works on this breakdown principle. The voltage is usually kept slightly above breakdown, but the tube is stable because no initiating electrons are present. If an ionizing particle (*e.g.*, from a radioactive source) passes through the tube, it breaks down rapidly and gives a pulse of current. The associated electrical circuit is arranged to decrease the voltage on the tube immediately, turn off the incipient discharge and sweep out all the charges, then reapply the high voltage. The tube then awaits another such event. The current pulses can each be amplified, and make the characteristic clicks heard in an earphone.

Generally, the quantity γ_i in equation (15) is either a constant or is at most a function of E/p. Since the quantity αd can be written as $(\alpha/p)(pd)$, and since α/p is also a function of E/p, the breakdown relation (15) states that the value of E/p at breakdown is a function only of the product of pressure and gap spacing (pd) for a given gas type and geometrical configuration. But since the voltage is E/p times pd, the breakdown voltage V_b is also a function only of the product pd. This is called Paschen's law, and is illustrated in fig. 10 for air. At high pd, V_b is high in order that E/p be high enough for ionization to take place. At very low pd, it is again high because there is too little gas in the gap, and the ionization per electron impact must be very efficient if enough ionization is to take place. Short gaps at high pressure behave the same as long gaps at low pressure. Note that a minimum voltage exists, below which the gas will not break down.

Breakdown With High-Frequency Voltage.—With low-frequency voltage (*e.g.*, 60 cycles), all the phenomena considered above happen very fast compared to one cycle. Therefore the discharge acts like a direct current one, but with the role of cathode and anode interchanged every half cycle. At high frequency, however, there is a fundamental difference. Instead of being swept across the tube by mobility, electrons oscillate back and forth in the field, and convert this oscillating energy into random kinetic energy by making collisions with the gas. Because the charges diffuse out of the gas rather than leave under the influence of D.C. mobility, the breakdown condition at high frequency is that one new charge must be made for each one diffusing to the surrounding walls. No secondary or cathode mechanism is required, so that breakdown can be obtained inside a totally insulating glass vessel, for instance. The curve of breakdown voltage against pd is much the same as shown in fig. 10, and for much the same reasons as given for the D.C. case. However, the distance d in this case is the average distance from the ionization region to the absorbing boundaries. High-frequency breakdown was extensively studied about 1950 by a group under W. P. Allis and S. C. Brown at the Massachusetts Institute of Technology.

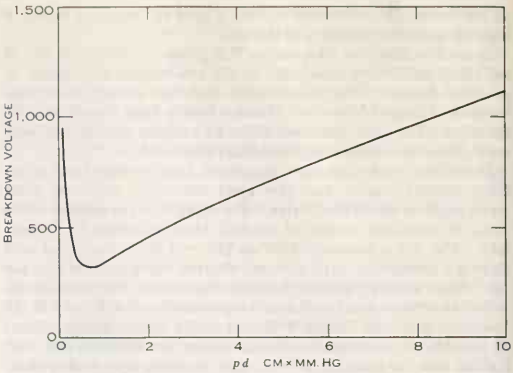


FIG. 10.—THE PASCHEN CURVE OF BREAKDOWN VOLTAGE FOR AIR BETWEEN PARALLEL PLANE ELECTRODES. BREAKDOWN VOLTAGE IS PLOTTED *v.* PRODUCT OF GAS PRESSURE AND INTERELECTRODE DISTANCE

Spark Breakdown.—Generally, one may designate as "sparking" all electrical gas-breakdown phenomena. But here we mean more particularly the class of filamentary discharges observed at large values of the parameter pd. These include the class of sparks observed in spark plugs, sparks picked off metal objects by the hand on dry days, corona from high-voltage power lines, and lightning. Many of the smaller of these sparks no doubt arise from the mechanisms described above, and are localized as filaments because of the effects of space charge (described below). But many of them operate differently—for lightning between clouds, for example, there is no cathode at all. Besides, the breakdown is found to occur too fast for ions to travel by mobility across the gap.

One important scheme is that of the streamer, proposed by Loeb and J. M. Meek about 1940. Here, an electron creates an initial avalanche as before. But the value of pd is high, and the avalanche has grown very large by the time it has crossed the gap. Two new effects come into play—space charge and photo-ionization of the gas. The electrons near the anode are drawn into it, leaving many slow positive ions. A space charge electric field (*see* below) arises, which enhances the applied field on the cathode side of the ion cloud. As mentioned before, ultraviolet light as well as ions are produced in the avalanche. In a gas mixture such as air, some of the light is capable of ionizing some of the air molecules—the process of photo-ionization, which is the exact inverse of recombination mentioned in the introduction. The light is therefore self-absorbed in the gas, and some electrons are produced on the negative side of the avalanche ions. These electrons find an enhanced field toward the anode, arising from the ion space charge. Thus they ionize strongly, feed into the ion cloud and leave behind a further excess of ions. In this way, a streamer of positive ions grows from the anode, being continually fed by photoinduced avalanches from its tip. In a short time, the streamer reaches the negative terminal (if there is one), and a conducting path is set up.

Corona from positive wires may arise from the same cause. Corona from negative wires can arise from the normal Townsend mechanism. In each case, the corona dies away at some distance from the wire, because the electric field intensity decreases sharply with distance from it.

4. The Glow Discharge.—Discussion of the phenomena of fig. 8 has so far extended to the current region E–F, which defines breakdown. The new phenomenon entering at F is space charge.

The Space Charge Electric Field.—Poisson's equation of electrostatics is

$$\frac{dE}{dx} = e(n_+ - n_-)/\epsilon_0 \qquad (16)$$

for a plane parallel system; dE/dx is the rate of change of E with distance. A similar, more general equation holds for more complicated geometries. The quantity $\epsilon_0 = 8.85 \times 10^{-12}$ farads per metre is the permittivity of free space. By equation (16), E cannot be uniform when a net space charge density $n_+ - n_- \neq 0$ exists; a space charge electric field arises, which varies with position. Its sign is such as to tend to expel the type of charge in excess and to attract the type of charge in the minority.

Let us apply this to the transition F–G of fig. 8. At F, ions are returning to the cathode; and since their mobility is low compared with that of electrons, they form a positive space charge. From equation (16), the field is no longer uniform—in fact, it is increased near the cathode, reduced elsewhere, and the total voltage stays the same. But the ionization does not stay the same. Note fig. 9—a doubling of E/p from 30 to 60 means an increase in α/p by a factor of 10. If this increase happens in one place, it more than counterbalances the reduction in ionization caused by a corresponding decrease in E/p at some other place. Thus, because of this nonlinear effect, the ionization has become too efficient and the tube is unstable. The current then increases with time as each electron avalanche more than reproduces itself by secondary electrons at the cathode. Alternately, if the tube current is controlled by some external means so that it cannot build up uncontrollably, the tube voltage decreases (region F–G). The effect of space charge becomes more pronounced as one proceeds from point F, for increasing current means more space charge, more field distortion and so on. The transition F–G extends over the approximate range 1–100 microamperes, but electrode size and configuration will change these values somewhat.

The Normal Glow Discharge.—At G of fig. 8 the space charge distortion has reached a limit—the region of high E/p is entirely confined to a region near the cathode, and a low value of E/p is found everywhere else. Electrons leave the cathode and are accelerated through the positive space charge region, ionizing partly in this region and partly at the end of it. The discharge now observed is the normal glow, and the voltages and currents near the cathode have fixed values, characteristic of the cathode surface (because of its secondary emission properties), type of gas and gas pressure.

The voltage drop through the space charge region (called the cathode fall) ranges from about 60 v. for a neon-argon gas mixture and barium oxide cathode, to about 275 v. for hydrogen and a molybdenum cathode. The current density per unit area is fixed, and is proportional to the square of the gas pressure. Thus a high-pressure gas tube will carry much more current than a low-pressure one. If the current is low enough, the glow discharge covers only a fraction of the cathode area; because the current density is constant, this fraction increases or decreases as the current is varied at will. The voltage stays the same, and region G–H is traced out. Because the voltage is constant with changing current, gas tubes operating in this region find much use as voltage regulators.

Appearance of the Glow Discharge.—Fig. 11 shows a typical glow discharge in a glass tube. Shaded areas indicate relatively bright regions of the discharge. For a tube of the size shown, the gas pressure would be about 1.0 mm.Hg.

The cathode fall, in which most of the tube voltage is now concentrated, is relatively dark, and is followed by the intense negative glow, where the electrons accelerated through the cathode fall region excite gas atoms copiously. From this region to the anode, relatively little voltage is available, and E/p is low. The negative glow provides a copious supply of low-energy electrons, which gain energy in passing through the Faraday dark space on their way to the anode. They can again excite and ionize by the time they reach the positive column, which therefore is also quite luminous. This region extends to the anode; often it appears striated, and looks like repetition of the negative glow at intervals all the way to the anode.

The phenomena near the cathode—the cathode fall and negative glow—are the most important sustaining ones of the discharge. The Faraday dark space and positive column provide the conduction mechanism by which current reaches the anode. Indeed, if the anode could be pushed in toward the cathode, the positive column would shorten to suit (or even disappear if necessary), the rest of the discharge staying the same. Similarly, if the anode be pulled out, the positive column extends; since there is a small longitudinal electric field present there, the tube voltage would rise somewhat.

The Positive Column.—In spite of its seemingly superfluous character, the positive column is of great practical interest. Consider a positive column in a long tube, so that end effects can be neglected. It contains a density n_+ of ions and n_- of electrons in almost equal amount. The flow of charges is a combined axial one along the tube, by mobility from the applied field; and a complicated radial one to the cylindrical walls. The fast electrons tend to diffuse more rapidly to the walls than the slow ions, creating a positive space charge within the column. By Poisson's equation (16), an outward radial space charge electric field E_s is therefore set up. The small radial ion diffusion current is aided by the outward mobility current $\mu_+ n_+ E_s$; and the electron diffusion current is partly canceled by the inward mobility current $\mu_- n_- E_s$. Thus the two flows are brought in balance, the net space charge adjusting itself to maintain the situation. This particular combination of diffusion and mobility flow is called ambipolar diffusion, because the combined flow can be described mathematically by an adjusted diffusion coefficient, which applies to both ions and electrons. It is approximately equal to $D_- \mu_+ / \mu_-$, which is a small fraction of the electron free diffusion coefficient D_-, but is several times the ion free diffusion coefficient D_+.

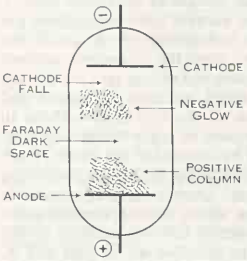


FIG. 11.—SKETCH OF THE PRINCIPAL VISUAL FEATURES OF A GLOW DISCHARGE. THE GLOW COVERS ABOUT HALF THE CATHODE AREA

Balance of production and loss of charge in the positive column demand that the total ionization (by electrons) just equal the ambipolar loss. Since the electron loss rate is reduced by space charge, the ionization rate (per electron) must be correspondingly reduced. The applied longitudinal E/p is therefore considerably lower than in the absence of space charge, and the column can be maintained with a relatively small voltage difference between its ends.

This type of ionized medium is one of the simplest so-called gaseous electronic plasmas, the term plasma being assigned by Irving Langmuir, who made many investigations of the medium. Since a space charge field exists in the positive column plasma, a plasma potential must also exist: the axis of the column is positive with respect to the walls by a voltage equal to several times the average electron energy.

Gas Lamps and Thyratrons.—From the point of view of the positive column, the cathodic parts of the glow discharge (*i.e.*, negative glow, etc.) merely provide the necessary copious source of electrons at its negative end. One can therefore replace everything up to the Faraday dark space by a thermionically emitting cathode, and save the high voltage drop (and power dissipation) associated with the cathode fall. An important class of devices operates this way.

The fluorescent lamp contains a thermionic emitter at each end, alternately serving as cathode and anode (on alternating current). The normal gas filling is a few millimetres of argon, which acts merely to inhibit charge diffusion to the walls. The current is carried by ionization of a much smaller addition of mercury, whose ionization potential is much lower, and which therefore is preferentially ionized. At turn-on, the cathode is heated directly through a separate starter connection, and line voltage is applied

to the ends of the tube. Electrons are emitted, and a positive column type plasma fills the tube. The cathode is designed also to be heated by ion bombardment from the plasma. This effect replaces the direct cathode heating, which is switched out. The lamp now is in normal operation. Light from the plasma is mostly ultraviolet. It is converted with good efficiency to visible light by a fluorescent coating on the inner surface of the tube walls—hence the name "fluorescent" lamp.

Many sodium-vapour and mercury-vapour lamps work on the same principle; they run very hot, so that the vapour pressure is appreciable. Some have thermionically heated cathodes. Light output in the visible spectrum is high. The mercury-vapour lamp in a quartz tube also passes a large amount of ultraviolet light.

A thyratron is similar to a triode vacuum tube, but contains gas (usually mercury vapour). Even with voltage on the anode, the control grid can prevent electrons from the thermionic cathode from penetrating the grid-anode space and setting up a conducting plasma. If the grid potential is raised, electrons are passed, and a highly conducting plasma forms all the way from cathode to anode. The grid loses control, because of space charge shielding, and the tube must be extinguished by removal of the anode potential. Thyratrons are therefore particularly useful for controlling large amounts of alternating current. Current-carrying capacity is much greater than that of vacuum tubes, because electron space charge—a limiting factor in the vacuum tube—is canceled by the presence of the ions. High charge densities n_+ of ions and n_- of electrons are present; the conductivity is high; and the net space charge $n_+ - n_-$ is low.

The Abnormal Glow Discharge.—At point H of fig. 8, the normal glow has covered all available cathode area. Further increase in current is accompanied by a corresponding increase in current density per unit area of cathode. The increased positive ion space charge near the cathode results in both a shortening of the cathode fall region (fig. 11) and an increase in voltage across it. The characteristic H–I of fig. 8 results. The glow runs less efficiently in this mode, for many of the ions produced in the negative glow are unable to diffuse back to the narrow high field cathode fall. Thus they do not produce secondary electrons, and more must be supplied to keep up the ionic return to the cathode.

The relatively high current density of the abnormal glow can be achieved without corresponding voltage rise if the cathode is made in the form of a cavity—for example, a hollow cylinder. In this situation, the glow operates on the inside surface; ions and radiation have virtually no place to go except back to the cathode surface. The current density can rise to very high values and transform to an arc (*see* below), without the potential rising above that of point H.

Impedance of the Glow Discharge.—The ratio V/i of voltage to current has dimensions of impedance, *i.e.*, ohms. It is evident from fig. 8 that the value ranges over many orders of magnitude, and that Ohm's law, which implies relative constancy of resistance, is not followed.

Of more interest in many applications is the dynamic or A.C. impedance, which is the incremental change in voltage resulting from an incremental current change. At very low frequency, it is just the slope of the voltage-current characteristic. In the figure, the region F–G displays high negative resistance (perhaps as high as -10^6 ohms). In order to operate the glow in this region, a positive resistance greater than this must be put in series with it to make the circuit stable. The region H–I displays positive resistance, typically of the order of a few hundred ohms. The region G–H has practically zero dynamic impedance.

The situation is different at higher frequencies, principally because of the effects of the slow ions. In the normal glow G–H, for example, the voltage is constant only if the glow can expand or contract on the cathode to accommodate a changing current. Ion mobilities are low, so that at a few kilocycles these excursions cannot occur in time. The glow therefore operates with fixed area, and has dynamic characteristics of the abnormal glow H–I, with positive impedance. The cavity cathode type discharge mentioned in the preceding section, having no abnormal region H–I, does maintain suitably low (sometimes slightly negative) impedance at kilocycle frequencies, and can be used in various combined switching and signal transmission applications.

Gas tubes of such types almost invariably have high reactive impedance. This is also an ionic effect—if the voltage is raised, the electrons ionize more efficiently forthwith, but ions can return to the cathode to enhance the secondary electron current only after their ion transit time. Thus a component of current lags the voltage, giving an inductive effect. The slowest part of the ionic transit is ambipolar diffusion from the negative glow to the cathode fall, and may take about 15 μsec. (1 μsec. = 10^{-6} sec.), even in the smallest electrode configurations. This inductive effect will persist up to frequencies of, typically, several tens of kilocycles. At certain higher frequencies, the current cycle may lag the voltage cycle so far that it appears to lead the next ensuing voltage cycle; the device then appears capacitive. These effects are presently more of a nuisance than of practical value. The frequency limitations are to be compared to those of vacuum tubes (no ions present), which operate into the kilomegacycle (10^9 cycles per second) frequency region.

5. The Arc.—At point I, the final transition of fig. 8 commences. The heating of gas and electrodes is intense. If the cathode becomes hot enough, it emits electrons thermionically: one then has the situation of the thermionic cathode glows mentioned in *Gas Lamps and Thyratrons*, above, but at much higher current—and the higher the current, the higher the cathode temperature becomes and the better is the emission. The situation is unstable and leads to the falling characteristic I–J.

Besides this effect, two others come into play. The gas is so hot (several thousand degrees) that thermal ionization, mentioned in the introduction, comes into play. The gas is partly ionized by collisions with its own fast (*i.e.*, hot) atoms. In addition, the density of excited gas atoms is very high, and these are much more easily ionized than unexcited ones. All of these effects become more efficient as the current density increases. Thus the arc naturally tends to concentrate its current in one intense channel, and the arc voltage drops (I–J) to a value near the ionization potential of the gas.

Arcs, of course, have very common usage. They are very luminous and make excellent high-intensity light sources. The gas plasma is intensely hot and, with various electrode and gas combinations, arcs are used for welding and for inducing difficult chemical reactions in the gas phase. The mercury arc is widely used as a rectifier. The cathode is a mercury pool, whose low boiling temperature evidently precludes thermionic emission. The arc mechanism is therefore presumably associated entirely with the gas itself, but its mode of operation is not completely understood. Current densities as high as a million amperes per square centimetre have been measured from small mercury cathode arc spots.

6. The Gas Plasma.—This section presents, in somewhat more detail, some elementary considerations of plasma physics. The behaviour of the plasma, alluded to in earlier sections, is sufficiently unlike the normal gaseous state that it qualifies as a separate state of matter.

Energy Densities.—It is convenient to define three energy densities associated with the plasma. The first is the kinetic energy per unit volume nkT, where n is the density of the particles in question, k is Boltzmann's constant and T the temperature. By the thermodynamic equation of state, energy density per unit volume is the pressure. The second is the electrostatic energy per unit volume $\epsilon_0 E^2/2$ (in m.k.s. units). The third is the magnetic energy per unit volume $B^2/2\mu_0$; where B is the magnetic induction, and $\mu_0 = 4\pi \times 10^{-7}$ henries/metre is the inductivity of free space (B/μ_0 is the magnetic field). Many plasma effects arise because of balances of pressure or energies per unit volume.

Electrostatic Containment.—From Poisson's equation (16), one may obtain the space charge field. Neglecting factors whose numerical values are nearly equal to unity, we find from (16)

$$E = en\Lambda/\epsilon_0 \qquad (17)$$

where n is the difference density ($n_+ - n_-$), and Λ is a characteristic length over which n is appreciably different from zero. The space charge voltage is approximately equal to $E\Lambda$; and it is ex-

pected that charges will be contained if this voltage is equal to or exceeds the average energy of the charges, or if

$$en\Lambda^2/\epsilon_0 = kT/e \tag{18}$$

But this is the same as writing

$$(en\Lambda)^2/\epsilon_0 = nkT \tag{19}$$

Equation (19) states the approximate equality of field energy E^2/ϵ_0 (from equation [17]) and kinetic energy. There exists, then, a net charge density which can be derived from (19), above which charges of the opposite sign tend to be electrostatically confined by the electric "pressure."

Let us apply this, for example, to the space charge transition at point F of fig. 8. There the charges are mostly ions, so that $n=n_+$; and those to be held in are electrons, so that $T=T_-$. The current from equation (10) is $n_+ev_{\mu_+}$ per unit area, and the total current is this quantity multiplied by the area, which is proportional to Λ^2. The total current, then, is

$$i = \epsilon_0 kT_- v_{\mu_+}/e \tag{20}$$

independent of the size of the system (but depending somewhat on its shape, through factors of the order unity neglected here). For reasonable values of T_- and v_{μ_+} (1 ev and 6×10^4 cm./sec.), equation (20) gives about 10^{-6} amp., as shown on fig. 8.

As n becomes very large, (19) shows that Λ must become very small. In the well-developed positive column or arc column, for example, a thin space charge "sheath" forms on all boundaries of the plasma, and over this sheath the containing voltage is built up.

Magnetic Containment.—Up to now, magnetic effects have been completely ignored. The effects of magnetic fields are often profound, but are more subtle and difficult to explain. For these reasons, and perhaps others, simple gas conduction devices using magnetic fields have not been widely developed; and the so-called hydromagnetic (or "magnetohydrodynamic") theory describing the processes has been slower in evolution. Strong interest in certain branches of astrophysics, ionospheric phenomena and controlled thermonuclear power provided great stimulus in this subject during the period following 1950.

Two basic effects are to be considered—the magnetic induction B arising from a current, and the motion of a charge in the presence of B. Fig. 12 shows the first effect in a somewhat oblique view. A conductor carrying current i is surrounded all along its length by a magnetic induction B, whose direction forms circles about the conductor. The induction is strongest near the conductor, and drops off with distance. Precisely,

$$B=\mu_0 i/2\pi r \tag{21}$$

for a long conductor, where r is radial distance. Fig. 13 shows the second effect, also in slightly oblique view. A magnetic induction B is presumed to be present in a region of space and is stronger to the right, as shown by the concentration of field lines in that direction. A charge moving in the field experiences a force equal to the product of its charge, its velocity component $v_\perp$ at right angles to the field, and the intensity of the field B; *i.e.*, $F=qv_\perp B$. The force is at right angles both to the field and to the direction of motion. While the motion is complicated, two limiting cases are simple: a charge traveling in the direction of lines of B is unaffected, and a charge traveling at right angles to the lines of B (in a uniform field) moves in circles. A charge initially moving at some oblique angle moves in a spiral, as shown in fig. 13. If it is a positive ion, it moves into the page at the top and out at the bottom.

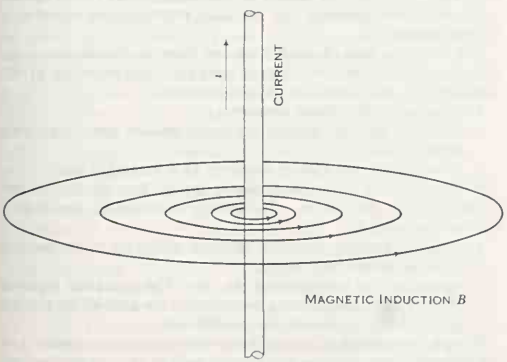


FIG. 12.—LINES OF MAGNETIC INDUCTION (B) SET UP AROUND A CONDUCTOR CARRYING A CURRENT (i). THE INTENSITY OF B IS HIGHEST NEAR THE CONDUCTOR, AND DECREASES PROPORTIONAL TO INCREASING RADIAL DISTANCE

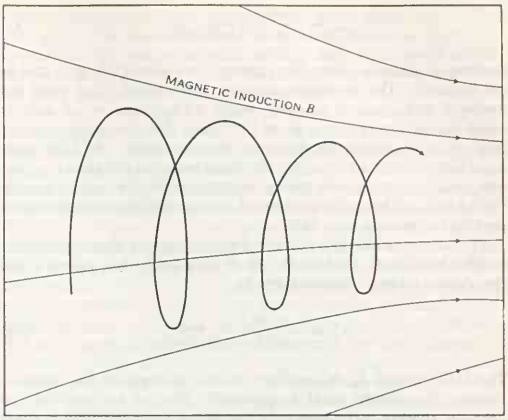


FIG. 13.—MOTION OF A CHARGE IN A MAGNETIC INDUCTION (B). THE CHARGE HAS A COMPONENT OF VELOCITY IN THE DIRECTION OF B, HENCE THE DRIFT TO THE RIGHT; AND A COMPONENT AT RIGHT ANGLES TO B WHICH CAUSES IT TO MOVE APPROXIMATELY IN CIRCLES. THE RESULTANT MOTION IS A SPIRAL

The frequency of rotation is the cyclotron frequency $f_b = qB/2\pi m$, independent of velocity; and the radius of the circle is $mv_\perp/qB$. Ions (large mass m) spiral slowly in large circles, and electrons (small m) spiral rapidly in small circles. If the magnetic field does not change with time, the charge neither gains nor loses energy from the field.

It is evident that the field B inhibits charge drift in all directions except in the direction of B. Indeed, if other effects are not present, a charge is trapped in the magnetic field lines, as shown in fig. 13. The only way that it can escape is by making a collision with (for example) a neutral atom, which changes its direction of motion and starts it off on a new spiral, offset from the previous one. In the presence of a magnetic field, the diffusion coefficient in the direction of B is given by equation (4), but at right angles it is the smaller value

$$D_\perp = \frac{l_v}{3\left(1+\left[\frac{qBl_c}{m}\right]^2\right)} \tag{22}$$

Similar relations hold for the mobility. Since these quantities have different magnitudes in different directions, the medium is not isotropic. Note that equation (22) shows that collisions with gas assist sideways diffusion if a strong magnetic field exists, consistent with the prior discussion.

The modification of the diffusion by the magnetic field is accompanied by a similar modification in the conductivity σ of the plasma. Parallel to the magnetic field, the conductivity is unaltered, and is properly described by equation (12) with insertion of the appropriate mobility for electrons and ions from equation (8).

If an electric field $E_\perp$ is applied perpendicular to the direction of B, one can distinguish two conductivities. First, in the direction of the electric field, the conductivity is

$$\sigma_{\perp} = \frac{nq^2 t_c}{m(1 + [qBt_c/m]^2)} \tag{23}$$

for ions or electrons when the appropriate values of n, q, t_c and m are inserted. The expression reduces to the usual value when the magnetic induction B becomes small. The mass m_i of ions is much larger than the mass m_e of electrons; thus the magnetic term in $\sigma_{\perp}$ is much larger for electrons than for ions. At high magnetic fields, therefore, $\sigma_{\perp}$ for electrons becomes less than $\sigma_{\perp}$ for ions, and currents across strong magnetic fields are carried mostly by the ions. This is the reverse of the normal situation obtaining parallel to the magnetic field.

The other conductivity arises from a current flowing perpendicular both to $E_{\perp}$ and to B. It is called the Hall current, and the corresponding conductivity is

$$\sigma_H = \frac{nq^2 B t_c^2}{m^2(1 + [qBt_c/m]^2)} \tag{24}$$

The Hall current J_H is peculiar: since q is negative for negative charges, the current itself is oppositely directed for positive and negative charges. It arises only because a magnetic field is present; if B approaches zero, σ_H is larger than $\sigma_{\perp}$, although both become small. Thus in this limit the small current carried across the magnetic field arises from the Hall conductivity.

The effect of an applied magnetic induction on gas breakdown or low current glows is relatively easy to see, in a qualitative way. Motion at right angles to B is inhibited, and any physical boundaries in those directions are that much more remote, as far as the charges are concerned. A long thin cylindrical tube with electrodes at each end (*e.g.*, a fluorescent lamp), and axially applied B, will behave like a much fatter tube of the same length. Since radial diffusion is inhibited, the tube breaks down more easily, and sustains a plasma column with lower applied voltage.

An interesting type of containment arises when the plasma current is high and generates its own magnetic field. Consider again fig. 12. Positive charges traveling up the conductor (or electrons traveling down, which give an upward conventional current) are acted upon by the magnetic field arising from their own current. The force in each case is toward the centre of the conductor. The conductor is therefore squeezed by the self-generated field; if it is a gaseous conductor, it may collapse. Let us calculate this current, on the presumption that the collapse occurs when the magnetic pressure $B^2/2\mu_0$ exceeds the kinetic pressure nkT at the surface of the conductor, whose radius is R. From equation (21)

$$\frac{\mu_0 i^2}{8\pi^2 R^2} = nkT \tag{25}$$

But the charge density n per unit volume is $N/\pi R^2$, where N is the number per unit length of conductor. Thus

$$\frac{\mu_0 i^2}{8\pi} = NkT \tag{26}$$

For a given charge density per unit length (N) and given energy (kT), a current exists for which a plasma column is unstable and pinches itself into a thin string. The effect was first described theoretically by W. H. Bennett in 1934. Generally, the current required is large—discharges can be made to pinch at currents of 10,000–100,000 amp., and much higher currents are often employed in studies of the process.

Questions of spatial stability of such plasmas in magnetic fields will not be discussed in any detail. The system of ionized particles and field has some of the properties of two fluids in contact—one composed of real matter and the other composed of pure energy. The dynamic behaviour of such a system involves the phenomena of mixing and turbulence, as well as others peculiar to this special medium.

Such plasmas, magnetically contained, began to be intensively studied during the late 1950s as a means of achieving controlled thermonuclear power. The problem is to heat a gas (probably deuterium) to about $3 \times 10^{8\circ}$ K. (30 kv. average particle energy), so that thermonuclear reactions can occur. The gas must, of course, be isolated from any confining material walls. Magnetic fields, either self-generated or externally applied, appeared to be most promising in this respect. A field of about 5,000 gauss is numerically equal to atmospheric pressure. Much higher fields can be produced, and the magnetic pressure increases as B^2.

Plasma Resonance.—Suppose electromagnetic energy W (watts per square centimetre) at frequency f (cycles per second) is incident on a plasma of electron density n. A fraction of the energy is transmitted, and the rest is scattered by the electrons, which oscillate at frequency f because of the field. As n is increased, the scattered part increases; and some critical density exists at which all is scattered and none is transmitted if the plasma is many wave lengths long. The critical density is a function of f. It is more customary to consider instead a fixed n, and and to find the corresponding critical plasma frequency f_p. Theory gives

$$f_p = \frac{1}{2\pi}\sqrt{\frac{ne^2}{m\epsilon_0}}\ \text{cycles/sec.} = 8970\sqrt{n} \tag{27}$$

where the latter form applies to electrons, with n measured in units of cm.$^{-3}$ The effect exists also for ions, but the large mass appearing in (27) indicates a much lower ionic frequency, and the influence of electrons is generally predominant. For a fixed density n, frequencies less than f_p will not propagate through the plasma. If the collision time t_c is low enough that little energy is transferred to neutral atoms, the plasma is highly reflecting.

These facts are paramount in utilization of the ionosphere for radio transmission. The electron density in the ionosphere, which is in the order of 10^6/cm.^{3} ($f_p \approx 9$ mc./sec.) changes somewhat according to solar conditions, and from day to night. For oblique incidence of radio waves, appreciably higher frequencies are tolerably well reflected; thus short-wave transmission up to about 40 mc. is possible over long distances, for the waves are trapped between the ionosphere and the earth and are guided around the earth's curvature. Higher frequencies (*FM*, television, radar, etc.) penetrate the ionosphere and must be received on line of sight from the transmitter. The detailed behaviour of the ionosphere is very complex. Magnetic interactions are very important, for the magnetic energy density (from the earth's field) far exceeds the kinetic energy density.

At the plasma resonant frequency, the energy of oscillation of the electrons and the energy density of the electromagnetic radiation per unit volume are equal. The two forms of energy—electromechanical and electromagnetic—may under some circumstances be coupled together, and coherent wave motion of the plasma can result.

Since the plasma transmission properties depend on frequency, electromagnetic techniques are widely used in the laboratory for probing the state of ionization of plasmas.

Factors Influencing Plasma Conductivity.—In this last section, some factors influencing plasma conductivity will be reviewed, and a few additional remarks will be made regarding the completely ionized plasma.

For the usual case of weakly ionized plasmas (fluorescent lamp columns, normal arcs, etc.), equation (12), with insertion of the mobility (8), indicates that the conductivity

1. increases with charge density n;
2. increases with decreasing pressure (smaller density of scattering atoms);
3. decreases in directions transverse to a magnetic field.

The existence of a finite conductivity implies that the system has internal energy dissipation. The energy imparted by the electric field is dissipated in

1. recoil of neutral atoms on collision with fast electrons, and to some extent with ions;
2. excitation and ionization of the gas. The ionization replaces charges lost through being swept out of the ends of the plasma by the field, or through lateral diffusion.

Positive ions are much more effective in scattering electrons than are neutral atoms, because the electric field of the ion reaches out to long distances from it. Thus if even a small percentage of the gas is ionized, the collisions which scatter electrons most effectively are those with ions. Increasing the charge density in this case no

longer increases the conductivity significantly, because increasing the density of the principal carriers (electrons) implies a corresponding increase in density of scattering centres (ions). Ionization and excitation of the neutral gas still goes on, however, and the luminous radiation from the plasma is intense.

Now consider the completely ionized plasma. Ionization and excitation are no longer energy loss mechanisms; and no light arises on this account, because the plasma consists of bare nuclei and electrons. Collisions are between charged particles entirely. The plasma is generally isothermal, each type of particle having the same average energy u.

The conductivity is high, and is given by

$$\sigma = \sigma_0 u^{3/2} \tag{28}$$

For u measured in kilovolts, σ_0 is approximately the value for metallic copper, if the nuclei are protons. For heavier elements, σ_0 is reduced by the atomic number, since nuclei with higher charge are more effective in scattering.

The energy transferred from an applied electric field to such a plasma goes into heating the plasma (through the collisions) and into so-called bremsstrahlung radiation. This form of radiation arises when free electrons change direction of motion, as in scattering collisions with ions. The bremsstrahlung is relatively weak, and a fully ionized plasma does not radiate intensely. The radiation per unit volume of plasma is proportional to the product $n^2 u^{1/2}$. The dependence on n^2 arises from the scattering rate, which is proportional to both the electron density n and the nuclear density n_+ $(=n)$. The scattering rate is also proportional to the relative velocity of electrons and nuclei, hence the $u^{1/2}$ (proportional to velocity) dependence also.

See DIFFUSION; ELECTRICITY; ELECTROCHEMISTRY; ELECTROMAGNETIC WAVES; LUMINESCENCE; MAGNETISM; PHOTOELECTRICITY; RADIO; *see* also references under "Electricity, Conduction of" in the Index volume.

BIBLIOGRAPHY.—J. D. Cobine, *Gaseous Conductors* (1941); M. J. Druyvesteyn and F. M. Penning, "Mechanism of Electrical Discharges in Gases of Low Pressure," *Rev. Mod. Phys.*, 12:87–174 (1940); F. Llewellyn Jones, *Ionization and Breakdown in Gases* (1957); *Encyclopedia of Physics*, vol. xxi and xxii, "Gas Discharges" (1957); L. B. Loeb, *Fundamental Processes of Electrical Discharge in Gases* (1939); S. C. Brown, *Basic Data of Plasma Physics* (1959). (D. J. RE.)

ELECTRIC LIGHTING: *see* LIGHTING.

ELECTRIC POWER. Public supplies of electricity in several of the more advanced countries appeared in the 1880s. These were at first quite small in both capacity and range. They were used mainly for electric lighting and most often supplied direct current at low voltages. Urban areas were the first to receive supplies. As the power demands and the distances of consumers from the power stations increased alternating current was employed because of its greater transmission efficiency through the use of transformers to increase transmission voltages.

As industry began to appreciate the advantages of electricity for power, private electric power stations were established by large manufacturing firms and at such places as mines where abundant cheap coal made generating costs low. With the rise in capacity of public power stations and the spreading of supply networks, private stations tended to close down or to be used only as stand-by plants, the main industrial demands being met from public networks. Although the generation of electric power by industrial concerns for their own use continued, in the years following World War II such generation was only a small proportion of the total in most countries.

Patterns of electricity supply have been influenced chiefly by the stage of industrial development of a given country when electricity was being introduced, the distribution of the population, the form of the government, the general level of prosperity and the energy resources available for power production. In Great Britain and France public supplies developed in thickly populated urban areas, with networks spreading out into the surrounding countryside; nationwide planning came later. In such "newer" and smaller countries as New Zealand, especially where ample resources of water power existed, electrification plans either for the whole country or for each of its main divisions were introduced. Where distances between communities were great, as in the U.S.S.R. and the western United States and Canada, or where indigenous energy resources were lacking, individual small power stations, usually driven by internal-combustion engines, were the rule until the middle of the 20th century.

Following World War II, rapid advances in techniques for the construction of hydroelectric plants and in extra-high-voltage transmission led to the development of major water power resources remote from load centres, as well as to much more extensive networks. The trend, therefore, was toward the replacement of small local stations by network supplies wherever the density of population warranted. These advances were accompanied by equally important developments in the size and efficiency of steam-power generating plants and in nuclear power generation. The main features of electricity supply in the early 1960s, therefore, were the concentration of thermal generation in a relatively small number of large and highly efficient stations burning coal or oil, some to be supplied later from nuclear energy; the development of large water power sources, with extra-high-voltage transmission networks connecting stations or interconnecting separate supply systems; and new forms of small power plants (*e.g.*, wind power plants or small water turbines) for communities in thinly populated regions with poor prospects of early industrial development.

This article is divided into five principal sections: *Historical Background; Generation*, which describes the energy sources, equipment and techniques employed to produce electricity in quantity; *Transmission* and *Distribution*, which describe the methods used to convey electricity to the user; and *Commercial Aspects*.

(E. W. G.)

HISTORICAL BACKGROUND

After the development of the dynamo, now more commonly called the electric generator, and the development of such uses for electrical energy as lighting, the desirability of providing some means of serving electric energy to a number of customers from a common generating system was recognized. A plan to accomplish this was first proposed in 1878 by St. George Lane-Fox in England and by Thomas A. Edison in the United States. The first public power station was placed in operation in London (at Holborn viaduct) Jan. 12, 1882, and on Sept. 4 of that year power supply to the public began in New York city. The systems produced direct current (D.C.) and their service was limited to highly localized areas. Although such systems played a very important part in the early development of electricity supply, it soon became evident that a more suitable means of transmission or some different form of electric current should be employed. The problems arising in long-distance transmission of direct current seemed insurmountable.

In 1881 Lucien Gaulard and John D. Gibbs had obtained English patents for a "series alternating current system of distribution." These were purchased in 1885 by George Westinghouse. William Stanley, working with Westinghouse, perfected the transformer, based on these patents, and developed an alternating-current, constant-potential generator. In 1886 Westinghouse and Stanley demonstrated the practicability of alternating current in Great Barrington, Mass., by transmitting single-phase electric power a distance of 4,000 ft., using one transformer to increase the potential at the source to 3,000 v. and another at the receiving end to reduce it to 500 v. Nikola Tesla had developed the induction motor, which used alternating current, and by 1888 alternating-current (A.C.) supply was becoming common in the United States. It was later widely adopted in Great Britain after the pioneering work of S. Z. de Ferranti. The alternating current generated in de Ferranti's original power station at Deptford was carried to central London, a distance of 28 mi. In 1890, a line was opened to carry 480 kilowatts (kw.) of single-phase power at 4,000 v. and 125 cycles per second 14 mi. from Willamette Falls, Ore., to Portland, Ore.

One of the early results of the strong demand for practicable power transmission was the effort to construct a large-scale hydroelectric development at Niagara falls and to carry the power to Buffalo, N.Y., a distance of 22 mi. In 1887 the city of Buffalo

offered a prize of $100,000 for the best transmission plan. A number of mechanical systems were proposed, most of which were patterned after European installations. None was found acceptable, and no award was made, despite the fact that in 1886 Stanley had already demonstrated successfully the transmission of alternating current. However, the concept was so new that its possibilities were not immediately recognized. In 1890 the International Niagara commission, among whose members was Lord Kelvin, was unable to come to a conclusion, and it was not until 1893, after demonstrations in Portland, Ore., and elsewhere, that the Niagara Power commission selected alternating-current transmission, enabling the development to go forward.

By 1898 alternating-current transmission had developed so well that a 30,000-v., three-phase, 75-mi. line from Santa Ana, Calif., to Los Angeles was placed in service, and in 1899 a 40,000-v., 70-mi. line from Colgate Hydro, Calif., to Sacramento was opened. Development of higher voltage transmission was restricted by the limitations of the pin-type insulator then available, but in 1907 E. M. Hewlett and H. W. Buck developed the first suspension insulators, making practical the consideration of higher voltages. By 1920 voltages up to 132,000 were common, and some lines were being operated at 150,000. In the period 1920–45, voltage levels of 220,000 became common, and in 1934 a high point for a number of years was reached when power generated at Boulder (Hoover) dam on the Colorado river was transmitted to Los Angeles, a distance of 270 mi., at 287,000 v.

Following World War II it became evident in many countries that, with the tremendously increasing demand for electric energy, high-voltage transmission would be needed to transmit larger and larger "blocks" of power. Studies and experiments on demonstration lines with voltages of 500,000 and beyond were undertaken. In Great Britain a 275,000-v. supergrid was constructed to transmit power in large blocks from one part of the system to another, and the original extensive 132,000-v. British grid, construction of which had begun in 1926, was relegated practically to the role of subtransmission. The first of the higher-voltage lines was placed in operation in Sweden in 1954, a 380,000-v. line from Harspranget hydroelectric station, 22 mi. N. of the Arctic circle, to southern Sweden, a total distance of 600 mi., with an intermediate tie about 300 mi. S. of Harspranget to an existing 230,000-v. system. In the United States, Canada and Australia 330,000-v. to 345,000-v. lines came into use, while Germany, France, Sweden, the U.S.S.R. and other countries began using 380,000-v. to 400,000-v. transmission. Studies and experiments were under way in many countries aimed at levels of 750,000 v. and higher.

In most countries two frequencies, one high and one low, have become standard. In the U.S. and Canada, a frequency of either 60 or 25 cycles per second (c.p.s.) is commonly used for general utility purposes, with 60 predominating. For railway power 25 c.p.s. is predominant. On the European continent 50, 16⅔ and 15 c.p.s. are standard; in Great Britain and South America, 50 and 25.

The use of direct current also began receiving renewed attention. In 1954 an experimental direct-current submarine cable transmission line, operated at 100,000 v. and carrying 20,000 kw. of power, was put in service between the mainland of Sweden and the Isle of Gotland, 60 mi. away. Two 30-mi. submarine cables under the English channel from Dungeness in Kent, Eng., to Le Portel, a suburb of Boulogne, France, began operation in 1961, carrying 200,000 v. D.C. with a total capacity of 160,000 kw.; they were designed to link the 275,000-v. British grid and the 225,000-v. French network. (E. R. DE L.)

GENERATION

The industry that provides electrical energy is unique in that it manufactures its product at the very instant the customer wishes to use it. The industry can store energy in the form of coal and impounded water to meet future requirements, but this does not lessen the need for generator capacity to match the customers' demands.

Electric power is expressed as the product of voltage and current. A given value of power can be produced by any one of an infinite number of combinations of voltage and current values. The current may be direct, in which case the electrons progress (usually slowly) always in the same direction through the device receiving power, or it may be alternating, in which case the electrons move back and forth. For many applications (*e.g.*, home lighting) either type of current is suitable; however, A.C. customarily is used because of the greater efficiency with which it can be transmitted from one point to another. A direct-current supply is essential for certain industrial applications—*e.g.*, for the operation of some types of electric motors, for electroplating and for electrometallurgical processes.

Generation From Fuel

In the generation of electric power from fuel, the heat energy of the fuel is first converted into mechanical energy by a prime mover and then into electrical energy by a generator. The principal prime mover in use in the U.S. during the 1960s was the steam turbine, which generated about 60% of the total electric power. Most of the remainder was derived from water power. In other countries the ratio was different, depending upon the available supply of commercially useful hydraulic energy and upon relative fuel costs.

As power systems increased in size, the size and speed of turbine-generator sets also increased. During the years from 1900 to 1920 the 1,200-r.p.m. set predominated, the largest unit being rated 62,500 kilovolt-amperes (kva.). From 1920 to 1935 the 1,800-r.p.m. set predominated, with units up to 200,000 kva. in service. After 1935, 3,600-r.p.m. sets replaced slower-speed sets, and by 1950 nearly all new sets were rated at 3,600 r.p.m., the maximum permissible to produce 60-c.p.s. current. The weight in pounds per kilovolt-ampere rating of turbine generators was reduced from about 15 in 1910 to 5 in 1953. As speeds were increased, the rotating mass per kilowatt of rating was decreased.

In general, the limit to the capacity of generators has not been reached. Economic considerations usually call for a few units of large capacity rather than many small units, but in each case local conditions determine just how large a single unit should be. Every central station requires an individual design tailored to local conditions, such as the amount and character of the land available, the kind of fuel available, the source of condenser cooling water and the kind of electrical load served. In addition to steam generating equipment (*see* BOILER; STEAM: *Steam Generation* and *Steam Power Plants*), stations are structurally divided into (1) generator room, (2) bus system, (3) switchgear and (4) control room.

Generator Room.—The A.C. generator (*see* GENERATOR, ELECTRIC) consists of a hollow cylindrical steel element (stator) with groups of insulated copper conductors held in slots facing its inner surface; within the stator the field structure (rotor) is placed, with groups of insulated copper conductors held in slots on its outer surface. The conductors are so arranged that when direct current is sent through them alternate north and south magnetic poles are established on the rotor surface. A two-pole generator has one north and one south pole, a four-pole generator has two of each kind, and so on. As the rotor is turned by the turbine, the magnetic lines representing a pole are carried across the stator conductors and generate currents in them. The frequency of the alternating current thus produced is determined by the rotor speed and the number of poles. The magnitude of the voltage is fixed by the number of magnetic lines crossed and by the rotor speed. A complete alternation of the voltage is one cycle and results from one revolution of a two-pole rotor, one-half revolution of a four-pole rotor, and so on. Modern stators have three sets of conductors, each called a phase winding, connected to deliver current on three wires. This forms a three-phase armature winding from which voltages of equal magnitude can be obtained between any pair of the three wires.

The turbine-generator sets, completely enclosed in steel casings, are installed side by side with considerable space between them (for the design and operation of the turbine, *see* TURBINE: STEAM). Overhead, a traveling crane is provided so that the casings and rotating parts can be lifted clear for inspection or repair. The turbines are fed by heavily heat-insulated pipes that bring steam

from the boiler room. The generator output is carried in cables through ducts to the bus room. Often there is little equipment on the generator room floor other than the main units, although small separate turbine-generator units may be installed there to supply power for use within the station. Directly beneath the generator room is the condenser well, where steam leaving the turbines is condensed before being returned to the boilers.

The amount of heat resulting from all the losses in a large generator is only about 1% of its output, but in a 60,000-kw. machine this will amount to 600 kw., enough to supply a small factory or many residences. Consequently, large generators are totally enclosed so that a cooling medium, usually hydrogen, can be circulated through them.

Bus System.—The currents from the various generators are pooled on a conducting structure known as the bus system, consisting of a group of heavy copper bars or cables rigidly supported on insulators. The generators are connected to the bus system by individual lines so arranged that in case any section of the bus becomes inoperative the generators feeding that section can be reconnected through alternative paths to the portion of the bus still in service. In some stations the buses are sectionalized, the sections being installed in separate fireproof vaults. A section is disconnected completely in case of trouble on it, and the load is shifted to another section.

Switchgear.—The buses, bus sections and generators are interconnected through circuit breakers, which usually have their contacts immersed in oil to reduce arcing. Solenoids or small electric motors, operated from the control room, open and close the circuit breakers, which are also equipped with tripping relays to open their contacts when excessive currents flow in times of accident.

Switchgear, as well as the associated transformers used to increase voltages for long-distance transmission, is often located in an outdoor yard beside the power station.

Control Room.—The switching of the generators and feeders and the apportioning of the load among them are controlled from a central switchboard. It consists of a number of vertical panels upon which are mounted instruments for indicating and recording the current, the voltage and the power. A desk in front of the board is provided with rows of switches connected through low-voltage circuits to relays on the various circuit breakers. When energized, the relays connect the electric operating mechanisms of the breakers to a power supply circuit and open or close them as desired. Coloured lamps on the board indicate the positions of the breakers.

A signaling system, similar to that used on shipboard between bridge and engine room, is installed between the control room, boiler room and generating room. If the operator wishes a generator started, he presses the proper button, which actuates a signal beside that generator. An attendant starts the machine and, when it is up to speed, the operator gets it in step electrically with the system with the aid of a synchronizer. When the incoming generator is "in phase," it is connected to the desired bus. The operator then by remote control adjusts the turbine governor to cause the generator to deliver the desired amount of load.

The operation of all station equipment is correlated in the control room, so that the station sends out power when and where it is needed. A totalizing instrument on the switchboard indicates the sum of the outgoing power so that the operator can start up machines or stop them as the load demand requires. An indication of the totalized demand is also given in the boiler room to inform the attendants there about steam requirements and aid them in anticipating changes in station output.

Hydroelectric Generation

The extent to which world water power was investigated and developed after 1910 was one of the striking engineering features of the 20th century. Although falling or flowing water formed the earliest of the natural sources of energy to be utilized for providing power, most of the water power in use at the present time was developed after the first decade of the 20th century. Development of the hydraulic turbine made it possible to use any head from 10 to 5,000 ft. with high efficiency and at speeds that permit use of a reasonably cheap electric generator. It became commercially feasible to develop large power stations at sites far remote from any centre of industrial activity and to transmit the energy for long distances to a centre. Developments in electrochemical, electrophysical and electrometallurgical processes, most of which require large amounts of cheap electrical energy, created a demand that can, under favourable circumstances, be satisfied more readily from a water-power installation than from any other source. For a water-power installation to be economically justifiable it must in general be able to generate energy more cheaply than a steam plant. Even in countries with an adequate supply of cheap fuel this may be possible. In countries normally dependent on imported fuel, any available water power is of great value.

Site Selection.—The possibilities of hydroelectric development in any country depend on its physical characteristics. Adequate fall of rain or snow is the first essential.

Water may be collected and stored at a high elevation and led through tunnels or pipelines to a station at a much lower elevation, the difference in elevation being known as the head, or dams may be erected across a river to raise the level of the water behind them. Since the power that can be developed by a given volume of water is directly proportional to the working head, a high-head installation requires a smaller volume of water than a low-head installation to produce an equal amount of power. In addition, the pipelines and the turbines required for the high-head plant are smaller and the installation in general is cheaper. The volume of water that must be stored during wet weather to enable continuous operation throughout the year is also smaller. A small reservoir often will enable sufficient storage to be obtained at a reasonable cost. For these reasons such countries as Norway, France, Switzerland and Italy, which have mountainous regions subject to heavy rainfall in close proximity to industrial regions requiring large amounts of electrical energy, are in favourable circumstances for hydroelectric development.

In a high-head installation the water normally is brought from the supply reservoir through an open canal at a low gradient to a small chamber or forebay on the hillside above the powerhouse. From there a short, steep pipeline conveys it to the turbines. Although high-head sites offer many advantages, they are not essential for the successful development of hydroelectric energy. Most of the world's water-power sites that are reasonably close to industrial communities are in regions in which the gradients are medium. The aggregate power developed from medium- and low-head plants is far greater than that from high-head plants. Low-head plants, utilizing heads up to 40 or 50 ft., usually are located on rivers in which the gradient is small. Sometimes a natural head is available because of the presence of rapids or waterfalls. Some water is diverted and led through a flume or canal to the turbines from which it is discharged into the low-level reach of the river. In other cases a dam is built to make an artificial head. A low-head dam, by raising the natural level of the water, provides a certain amount of storage, but seldom more than is sufficient to store the night flow of the river for use during the day. A river plant of this type can give a continuous output equal only to the dry weather flow of the river. However, it is usually economically sound to install a hydraulic plant of a capacity greater than that corresponding to the minimum dry weather flow and to add a steam plant capable of maintaining the output when the water supply is inadequate.

Various types of low-head development are adopted to best fit the conditions at the power site. Where a dam is built, the powerhouse often is constructed on one flank of the dam with a short headrace or tailrace (whichever is more convenient); the dam itself is used as a spillway over which excess water is discharged in time of flood. Where the river flows in a narrow, steep gorge the powerhouse may be constructed in the dam itself, which then is designed as a hollow, reinforced-concrete structure. Where a river has a long bend it is often possible to cut across the neck of the bend and utilize the head between the two ends of the cut.

Turbines.—Low-head propeller-type turbines, inward-flow pressure turbines and Pelton wheels are used in modern hydroelectric plants; the type chosen depends largely on the available

head. The Pelton wheel is a slower-running machine than the pressure turbine and is therefore better suited for very high heads. In addition, since the water is discharged through its one—or at most two—nozzles, these may be of reasonable size when handling the small volumes of water normally available in high-head plants. On the other hand, the pressure turbine with its full peripheral admission of water is well adapted to utilize the large volumes necessary in low-head plants. Its higher speed of rotation enables the cost of the electrical generators to be reduced. The Pelton wheel is best suited for heads above 700 ft.; the pressure turbine for heads below about 250 ft. in small units and below about 500 ft. in large units; while over the intermediate range of heads the choice depends upon the size of a unit and special circumstances.

Modern pressure turbines, except the propeller type, are of the mixed-flow type with inward radial flow, through guide vanes surrounding the runner, and axial discharge. Pivoted guide vanes are used, speed control being obtained by simultaneous movement of these on their axes. Turbines for heads up to about 40 ft. are usually set in an open forebay. Either vertical—or horizontal—shaft turbines may be used, the vertical being commonest. For higher heads the water must be carried through a pipeline. The turbine is enclosed in a spiral casing designed to distribute the water uniformly around the periphery of the guide vane ring. For heads up to about 120 ft. the casing may be molded of concrete. For higher heads a metal casing is required. This may be of cast-iron, cast-steel or steel plate construction. In order to give rigidity and increased safety in case of pressure surges it is sometimes embedded partially or wholly in the concrete of the substructure. Most modern turbines have a single runner at the base of a vertical shaft on the top of which the generator is mounted. The weight of the shaft, runner and generator is carried by a single thrust bearing. This is a simple and efficient design, and the friction losses are low.

Runners in low-head turbines have a small number of vanes, often not more than four, shaped somewhat like those of a marine propeller. In the Kaplan turbine the vanes can be shifted on their axes to adjust for various rates of water flow. The pressure turbine is built in units capable of developing 100,000 kw., and this size could be increased if necessary. Efficiencies of 93% were obtained in tests of vertical shaft turbines at Niagara falls, and values approximating 90% are common. In a medium-head plant the turbine efficiency may range from 70% at one-quarter load to 90% at full load.

Most Pelton wheels have a horizontal shaft and one or two nozzles. They have been built in sizes up to 40,000 kw. Speed usually is controlled by a deflector that cuts off the jet of water from the wheel, acting in conjunction with a central needle or spear in the nozzle that slowly reduces the size of the jet while the deflector returns to its original position. At constant speed the efficiency of a Pelton wheel falls off comparatively slowly as the load is diminished. The wheel efficiency may range from 76% at one-quarter load to 83% at full load. (*See* also TURBINE: WATER.)

Generators.—Generators to be driven by hydraulic turbines range from the simple open-type machine, often used in small sizes, to constructions similar to those of steam-turbine-driven generators. In large, high-speed generators a forced draft system of ventilation is used. The rotor is designed to withstand the centrifugal forces accompanying the runaway speed that may result during the interval between the sudden dropping of load by the generator and the reduction of water supply to the turbine. Since centrifugal forces vary as the square of the speed, if a runaway speed of twice normal value is reached, centrifugal forces four times normal are produced.

Of the two common types of A.C. generators, the synchronous and the induction types, the former is the only one used on many systems and predominates on all systems. However, the induction type is more rugged, requires fewer controls and no direct current for excitation; the magnetic lines are produced by alternating currents drawn from the synchronous generators on the system. For these reasons the induction generator is sometimes used in isolated hydroelectric plants.

Associated Equipment.—The bus system, switchgear, transformers and control room equipment for a hydroelectric station are generally similar to those found in a steam generating station. Differences occur mainly in the special devices—electrically operated valves, pressure-indicating devices, etc.—needed to control the movement of large volumes of water through the turbines.

Automatic Generating Stations.—The automatic generating station is especially suited to systems where numerous small-power waterfalls are available. In such a case the expense of an operating staff at each station would be prohibitive, but if each station can be made automatic, and all are linked into a common distribution system, the labour cost is reduced to a reasonable value. The first of such stations in the U.S. was put into operation in 1917 on the system of the Iowa Railway and Light Co. and operated in parallel with a steam station situated about two miles distant. The automatic station contained three 500-kw. generators driven by Francis turbines operating under a ten-foot head. Normally the starting and stopping of the turbine-generator sets was accomplished automatically by the response of float switches actuated by the change in the level of water above the dam. Provision also was made for controlling these operations, as well as the gate openings of the individual turbines, from the central powerhouse.

Combined Operation of Hydraulic and Steam Plants.—Except where a large volume of water storage is provided, the variability of river flow makes it impossible to utilize more than a fraction of the total available energy unless generating equipment is installed which will have to be idle the greater portion of the year. It is found that economical operation can be obtained when the capacity of the turbines is such that they can be operated at full load for about six months of the year.

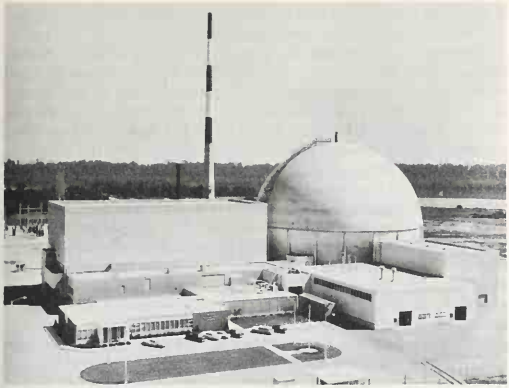
By operating a steam plant in conjunction with a hydraulic plant, it becomes economically feasible to increase the capacity of the hydraulic plant, the shortage of its output at times of less-than-normal flow being supplied by the steam plant. The latter also serves as a stand-by in case of a breakdown of the hydraulic plant. Most large hydroelectric power systems include one or more steam plants. The best method of operation of such a combination depends upon the type of load, the storage capacity, etc., and is determined by consideration of the special circumstances of the individual plant. In general, since for maximum economy a steam plant should operate at or near its rated capacity, at times of low water the steam plant operates continuously, carrying the base of the load, while the peak of the load is carried by the hydraulic plant. On the other hand, in a system having a very long transmission line it is more economical to have the hydraulic plant carry the base of the load and to carry the peak of the load on the steam plant. In this method of operation the cost of the transmission line is less than if the reverse method were adopted, and the saving may more than counterbalance the loss resulting from uneconomical operation of the steam plant.

GENERATION FROM NUCLEAR FUEL

Nuclear reactors release energy in the form of heat from the fission of uranium or plutonium. In a pressurized-water reactor the fuel structure is in the form of plates or rods. The heat of fission is delivered to water that is kept at a pressure of 2,000 lb. per square inch, raising the water temperature to 550° F. This water is circulated through a heat exchanger in which water in a separate piping system is boiled to produce the steam that is fed to the turbines. When steam is not required by the turbines the flow of pressurized water is diverted from the heat exchanger to a special cooling system while the reactor is being shut down.

A period of about eight hours is required to put a nuclear plant back in operation after a normal shutdown. First the instruments and the system components are checked to ascertain that they are operating properly. Then the auxiliary equipment is started, using power from the system to which the plant is to be connected or from a generator driven by a prime mover such as a diesel engine. Next, if water has been added in the reactor system, excess oxygen is removed from it.

To produce heat, the reactor control rods are withdrawn slowly to cause fission to begin. The rate of heat production is adjusted so as to cause the reactor temperature to increase at a rate of about



BY COURTESY OF COMMONWEALTH EDISON CO.

FIG. 1.—DRESDEN NUCLEAR POWER STATION

Located near Chicago, Ill., this station, the first privately financed nuclear power plant in the U.S., was placed in commercial operation Aug. 1, 1960. Dual-cycle boiling-water reactor housed in domed building at right produces steam to drive turbine-generator in high structure at left. The full power output of the plant, 192,000 kw., is sufficient to serve both the plant and a city of more than 200,000 residents

150° per hour. During the warm-up period a 20% increase in the volume of the water occurs, making it necessary to remove some water from the system.

A reactor is designed so that to a considerable degree it is self-regulating; *i.e.*, an increase in the amount of steam required by the turbine causes the reactor to deliver more heat, and vice versa. An automatic control system makes any additional adjustments needed.

During a short interval of zero electric power demand, the turbines are kept in operation to supply the auxiliary equipment. The reactor is kept at the critical point of fission and its output reduced as necessary. If it is known that the electric power demand will be zero for an extended interval, the control rods are inserted completely to make the reactor subcritical. *See* also NUCLEAR ENGINEERING. (G. V. M.)

TRANSMISSION

Arrangements of transmission lines linking power-generating sources and load centres form a transmission grid or network. Such arrangements, together with the associated power generating sources, are referred to as power pools or interconnected electric power systems.

Load centres are points at which the power received in bulk at transmission voltages is reduced to lower voltages for purposes of distribution. The lines used to carry electricity at this lower voltage are known as primary distribution lines. Interposed between the high-voltage lines and the primary distribution lines may be an intermediate voltage level which is designated as subtransmission. Since the voltage at which the primary distribution takes place is still too high for use in home or factory, it must be reduced further by a line or distribution transformer. The lines connecting the secondary distribution lines with the customers' or users' premises are called "services" or service wires. The combination of primary distribution lines, distribution transformers, secondary distribution lines and services is called the distribution system.

Types of Transmission Systems

Although the direct-current system was chronologically the first used, the alternating-current system has almost wholly displaced it. However, increasing interest in direct current for long-distance transmission arose after World War II.

Direct-Current Systems.—The incandescent electric lamp developed by Thomas Edison in the United States and Sir Joseph Wilson Swan in England could be operated at 100–110 v. on a two-wire D.C. system. R. E. B. Crompton, for a network he installed in Vienna in 1883–84, used a five-wire D.C. system having a central neutral wire with +110 v. and +220 v. on one side and −110 v. and −220 v. on the other, the difference in potential between the outermost wires thus being 440 v.

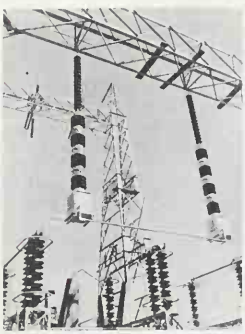
J. Hopkinson and Edison in 1883 devised three-wire D.C. systems that were thereafter commonly adopted. To maintain a balance on such a system, it was necessary to arrange that the loads on each side of the centre neutral line be as nearly equal as possible; furthermore, special generators, batteries or other equipment had to be used to maintain equality of voltage. The system was subject to accidents that could unbalance it, producing hazardous voltages, and it soon was superseded by the more flexible A.C. system. However, the relative freedom of D.C. systems from the effects of capacitance and inductance (*see* below) led to reconsideration of their use for long-distance transmission and for underground and submarine cables. The cable connection constructed in the English channel between England and France and the pioneering Swedish cable constructed from Sweden to the Isle of Gotland are noteworthy examples of underwater D.C. transmission. As for overhead lines, the U.S.S.R. in the late 1950s announced plans to construct an 800,000-v. D.C. transmission system from Volgograd to the Donbass, approximately 300 mi. In all of these undertakings, the power is generated as alternating current, stepped up to transmission voltage, converted to D.C. for transmission, reconverted at the receiving point to A.C. and stepped down to lower voltages for distribution.

Alternating-Current Systems.—Every A.C. circuit has three characteristic properties that may be distinguished at low and moderate frequencies, but become so intermingled at high frequencies that separate identification becomes difficult. These are resistance, inductance and capacitance.

Resistance is the opposition that conductors offer to the flow of electric current.

Inductance is the electrical analogy to inertia. Electric current creates around a conductor a magnetic field that may be represented by lines of magnetic flux encircling the conductor. Whenever the current is changing, the magnetic flux also changes and thereby induces in the conductor a voltage opposing the change of current. Inductance is the ratio of the voltage produced in this way to the rate of change of current. While inductance limits the flow of alternating current in much the same way as resistance, its effect depends upon change in current rather than the flow of current. When direct current is used, there is no change of current and consequently no inductance; however, if there is an interruption in the current, there will be an inductance effect until a steady-state flow is resumed.

When a voltage difference is maintained between two conductors that are separated by an insulator (which may be air or any other nonconducting material), an electric stress is created in the insulator. A positive charge appears on the conductor connected to the positive terminal of the power source and a negative charge appears on the other conductor. If a small isolated electric charge is brought into the vicinity of the charged conductors, it will tend to move toward or away from the conductors; *e.g.*, if the isolated charge is positive, it will be repelled by the conductor having the positive charge and attracted by the conductor having the negative charge. The ratio of the charge on one of the conductors to the voltage between them is the capacitance of the circuit.



BY COURTESY OF WESTINGHOUSE ELECTRIC CORP.

FIG. 2.—COUPLING CAPACITOR POTENTIAL DEVICES

Used to measure the high voltage of the power system and to introduce low-power high-frequency energy into the transmission line

The conductors of an overhead transmission line are relatively small in diameter and separated by large distances, with air acting

as an insulator between them. The capacitance, while relatively small for short lengths, becomes large in the aggregate when the line is long. In underground cables, the conductors are much closer together and the insulating materials between them have a dielectric constant higher than that of air. Hence, the capacitance in underground cables is greater than for lines above ground.

When alternating current is used, the voltages between the transmission conductors are continually changing; the charges on the conductor surfaces consequently flow back and forth, requiring an additional current in the conductors. The over-all losses in the transmission of energy may be substantial. When direct current is used, stable charges remain on the conductor surfaces, and no effect on the transmission of power is noticeable so long as steady conditions are maintained.

The effects of inductance and capacitance make extensive analysis necessary in planning an A.C. transmission system. For methods of analyzing circuits, when the resistance, inductance and capacitance may be considered concentrated and independent of each other, *see* ELECTRICITY. For a transmission line these circuit characteristics are distributed along the line and are so interrelated that they can be considered as concentrated only as an approximation. At frequencies of 50 to 60 c.p.s., the maximum distance suitable to such approximations is roughly 100 mi. for overhead lines and 20 mi. for those underground. For longer lines, or where more accuracy is desired in short lines, special equations are used to determine the voltage and current at any point from a minimum of specified information. By combining these methods with those for circuits having concentrated properties, precise predictions of the behaviour of entire electric systems can be made. (*See* NETWORK THEORY.)

Polyphase Circuits.—A single-phase transmission line is one consisting of two conductors. However, by using more than two conductors several advantages can be obtained. Between each pair of insulated conductors an alternating voltage can be applied, the current going out one conductor and returning in the other. If the alternations of voltage are started at slightly different times in the different pairs of conductors, the currents in each pair of conductors will be reaching the corresponding values in their alternations at different times. Under these conditions, the conductors may be interconnected, with one conductor forming, for example, the outgoing conductor for one circuit and the return for another. Because the currents which meet in a single conductor do not reach corresponding values in their alternations at the same time, the combined current produces a smaller energy loss than that which would occur if the two currents passed through corresponding points in their cycles simultaneously. This difference in timing is called a phase difference, and such combination circuits are called polyphase circuits. The most advantageous arrangement comprises three interconnected circuits, forming a three-phase circuit, as shown in fig. 3 and fig. 4. With either arrangement, the phase difference between successive voltages is one-third of a cycle. Although the power of each of the three circuits is varying as the current and voltage change, the total power for the polyphase circuit is normally constant—that is, as the power in one circuit or phase increases, the total power in the other two decreases in just the right relation so that the total for all three is a constant. A similar arrangement can be established with four circuits or even more, the phase differences between successive voltages being equal to one-fourth of a cycle if there are four circuits, one-fifth of a cycle if there are five circuits and so on. However, the three-phase system is the most efficient and is almost universal.

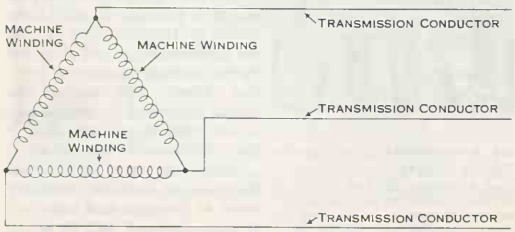


FIG. 3.—DELTA-CONNECTED THREE-PHASE CIRCUIT

Generators and motors are built with several windings having voltages that differ by the correct phase, so that generators and motors are also polyphase units. The constant, even flow of energy makes such motors and generators much smoother and more efficient in operation than they would be if the flow of energy were pulsating as in a single phase or single circuit. Also the magnetic interaction between the phases or windings of a polyphase motor makes it inherently self-starting when it is energized. By contrast, some special means must be provided to make a single-phase motor rotate when it is energized.

In the Y-connected arrangement of fig. 4, the common connection is called the neutral. When the currents in the three circuits have the same value and differ in phase by exactly one-third of a cycle, as intended, the current in the common conductor is always zero, and that conductor can be eliminated entirely. This ideal situation cannot always be maintained, however, and some provision must be made for a small current in the common or neutral conductor. If conditions of operation make it necessary to provide for appreciable currents, a fourth conductor, never larger than one of the other three, is installed. If the current likely to be encountered is small, the neutral may be connected to the earth (grounded), and this current permitted to flow in the earth. Establishing a ground minimizes the possible hazard to apparatus in the event of an insulation failure in the system and is generally a desirable practice. The hazard to persons who might come in contact with live conductors is also more definitely controlled, although in some cases the danger may be somewhat greater.

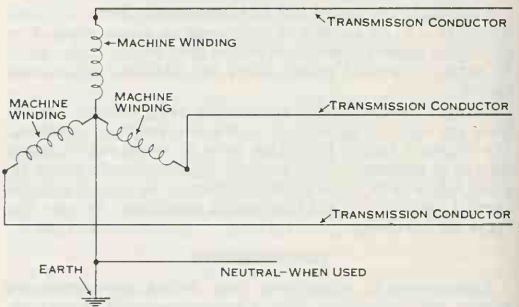


FIG. 4.—Y-CONNECTED THREE-PHASE CIRCUIT

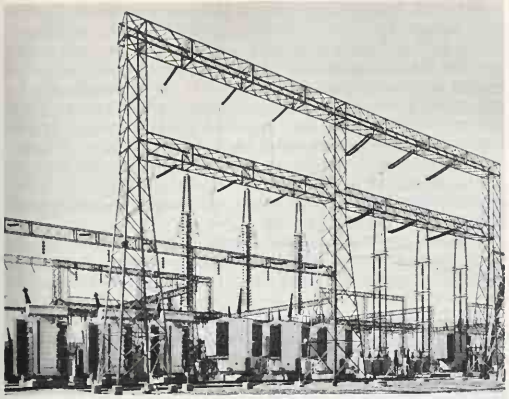
Although most transmission lines are three-phase, two or even more polyphase circuits are often run side by side, so that six or more conductors may be found supported on a single pole or tower.

TRANSMISSION LINE DESIGN AND CONSTRUCTION

After power-plant sites have been selected, surveys are necessary to determine the best routes for the transmission lines. A preliminary air reconnaissance is helpful, followed by a detailed ground survey. A series of studies is needed to determine design features of the line. After acquisition, clearing and grading of the land, the poles or towers are erected, the insulators are mounted and the conductors are strung.

Prior to construction, studies determine the voltage, conductor spacing, number and type of insulators, conductor size and material, pole or tower design and the protective features that will guard the line against damage from lightning or other natural hazards. On large systems it is necessary to check normal and abnormal operating conditions, and extensive studies on miniature systems, called network analyzers, may often be necessary.

The choice of transmission voltage depends primarily on the amount of energy to be transmitted and the transmission distance. For lines connecting hydroelectric plants to load centres, a voltage of about 1,000 v. per mile is often suitable. For new interconnections between systems the voltage is frequently the existing voltage of one of the systems.



BY COURTESY OF DEP'T OF WATER AND POWER, CITY OF LOS ANGELES

FIG. 5.—TERMINATING STRUCTURES

(Left) for a 138,000-v. transmission line; (right) for a 287,000-v. transmission line; lightning arrestors are suspended from the tower with transformers in the background

Corona and Radio Influence.—The major factor affecting the choice of conductor, aside from the current to be carried and the mechanical strength, is the phenomenon known as corona. If the voltage between conductors exceeds a critical value, a hissing sound will be audible and in the dark a glow can be seen around the conductor. The critical voltage depends upon the diameter of the conductor, the roughness of its surface, its distance from other conductors, and atmospheric conditions. Corona is caused by intermittent ionization of the air surrounding the conductor, producing random high-frequency waves and causing loss of energy in the line as well as static or interference in radio and television receiving sets. Practically the only way to minimize corona is to increase the diameter of the conductor. For example, at sea level a minimum conductor diameter of about 1 in. is necessary for a 230,000-v. line and about 1.7 in. for 330,000 v. Sizes must be increased at higher altitudes, where the insulating effect of the atmosphere decreases.

Conductor Materials.—Materials in general use for conductors are copper and aluminum, both of which frequently are wound around a steel core to provide strength. Copper conductors are sometimes made with hollow centres. For very high voltages, a conductor may be composed of two, three, or even four small conductors separated by distances from a few inches to a foot or more; this arrangement is known as a bundle conductor.

Insulation.—High-voltage conductors are suspended from the poles or towers by strings of insulators, each consisting of several disks of porcelain or glass held together by mechanical fittings. For lower voltages, a single insulator of porcelain or glass, called a pin-type insulator, is used. Insulators must be strong enough to support the weight of the conductors and to resist forces caused by wind and ice, the latter being a special hazard.

The formation of ice on conductors can be controlled by discontinuing the usual high-voltage transmission and by transferring through the line a large amount of current at a lower voltage at recurring periods of 30 to 60 min.; the conductor thus is heated sufficiently to inhibit the formation of ice. The scheme is most practicable where a large number of circuits are available so that the wires can be warmed in rotation without discontinuing service on the entire system. However, where conditions warrant, a large circulating current may be used continuously during ice-melting operations without the need of taking lines out of service.

Insulators must withstand not only the normal operating voltages but, wherever possible, abnormal voltages of short duration that may be caused by lightning or by the connection or disconnection of circuits. Insulating properties are reduced by dirt and moisture, but the minimum insulation value must prevent electrical breakdown for normal voltages, even during unfavourable weather. Fortunately, insulator failure nearly always occurs as an electric arc discharge through the air around the insulator, rather than by puncture of the insulator itself. If the line is disconnected from service rapidly by the circuit breakers, the insulators may suffer little permanent damage and the line frequently may be restored to service immediately.

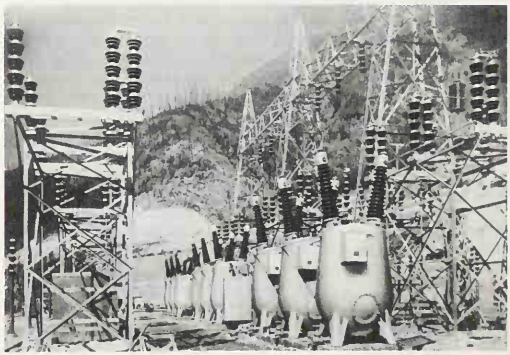
The insulation between conductors on overhead transmission lines is the air space between them and the spacing used must provide sufficient insulation so that breakdown will not occur as a result of the conductors coming in contact, or so close together that a flashover occurs between them, even when deflected by high winds.

Lightning Protection.—Transmission lines are susceptible to lightning, and a direct stroke may produce a serious short circuit unless special lightning protection is provided. A moderately high level of insulation is provided to prevent insulation failure against all except the more severe lightning strokes, and in especially dangerous locations added protective measures are used. However, it is usually uneconomical to insulate transmission conductors at each pole or tower against an electrical breakdown that might result from a severe direct stroke of lightning. (*See* LIGHTNING: *Lightning Protection.*)

Circuit Breakers.—When a transmission line suffers an insulation failure or other accident, it must be disconnected quickly from the unaffected portions of the system. The switches developed for this purpose are called circuit breakers. For high voltage systems circuit breakers are designed to operate in from 3 to 5 cycles (a cycle for a 60-cycle system would be $\frac{1}{60}$ sec.). Circuit breakers in turn are controlled by precise protective relays that detect the location of a failure and cause only the proper breakers to open. The relays can function in as short a time as one-half cycle, and thus a section of transmission line having a failure can be isolated in less than four cycles, or $\frac{1}{15}$ sec.

Two types of circuit breakers are in common use: oil-filled and air- or gas-blast breakers. Since it is neither possible nor desirable to interrupt the current instantaneously, an electric arc is drawn between the breaker contacts as they separate. In the oil-filled breaker a small amount of the surrounding oil is vaporized, producing a high pressure that drives quantities of unvaporized oil into the path of the arc. Moving electrons, which are the main elements of the electric current in the arc, become attached to the particles of vaporized oil and their mobility is so restricted that the arc is extinguished. In the air- or gas-blast breaker the arc is quenched by the force of an air or gas jet.

In addition to interrupting circuits, circuit breakers are also arranged to reclose circuits automatically, because transmission faults are sometimes of very short duration. Reclosing times are usually from 10 to 20 cycles. Should the fault persist after several openings and reclosings, the breaker will finally remain open.



BY COURTESY OF WESTINGHOUSE ELECTRIC CORP.

FIG. 6.—LARGE CAPACITY CIRCUIT BREAKERS FOR CONTROLLING FLOW OF ELECTRIC POWER ON 230,000-V. CIRCUITS

Circuit is interrupted by separating contacts under oil in the large tanks into which circuit is led through porcelain bushings

For low-power lines or where prompt reclosing of a circuit is not necessary, fuses of large size may be used. The usual type is the expulsion fuse, consisting of a long, thin strip of metal inside a tube made of insulating material and filled, in some cases, with a nonconducting, fire-extinguishing chemical in liquid or dry form. When the current exceeds a predetermined point, the metal strip melts, breaking the circuit, and the resulting gases in the container expand suddenly in such a way as to put out the arc formed by the melted fuse.

DISTRIBUTION

Formerly only one step of transformation at each end of a transmission line was used. However, as systems grew, it was soon necessary to use two or more steps of voltage reduction at the receiving end. The high voltage needed for long-distance transmission was too high for lines in city streets, and yet the secondary voltage was too low for supplying the energy to consumers over a wide area. Therefore, at the receiving end of a transmission line a first group or bank of transformers reduces the voltage to an intermediate level in a substation, and the secondaries are connected to bus bars, from which several circuits at the intermediate voltage radiate to different parts of the area, where the voltage is again reduced by other transformers for delivery to individual consumers. The intermediate circuits are known as distribution circuits. In many systems one intermediate level is insufficient and two or more are used.

As systems were interconnected, the transmission circuits of the original system became the distribution circuits of the combined system. Thus, the distinction between distribution and transmission circuits has become vague. It may be said, however, that transmission implies larger blocks of power delivered at higher voltages to a few main substations, while distribution implies smaller blocks of power delivered at lower voltages to many smaller substations or to consumers.

For residential supply it is common practice to provide small transformers on poles along the streets, each to supply a few consumers at 115 or 230 v., usually from one phase of the three-phase system. The circuits supplying these local transformers operate at 2,300, 4,000, 4,600, 6,900, 8,000, 11,500, 13,800 and similar voltages, depending upon local conditions. In most systems it was found necessary to increase the voltage from the initial level to a higher one as the system grew.

Duplicate Circuits.—The need for two or more circuits and duplicate apparatus to provide for emergencies resulted in many circuit arrangements or systems, in an attempt to provide a high degree of service reliability without undue investment. Two or more main transmission circuits are regularly provided from a main generating station to each major substation, each capable of supplying the load, at least for a short time, when the other is out of service. At main substations two or more banks of transformers may be used, so that service may be maintained if one fails. Often the incoming and outgoing circuits are connected to two groups of bus bars, so that one will be available as a stand-by. To provide the great variety of connections needed and to isolate any defective part, several circuit breakers must be provided at each substation, each with its complement of protective relays.

Since severe short circuits occasionally involve both circuits of a pair designed to reinforce each other, resulting in a shutdown, it is evident that duplication of all facilities will not provide complete assurance of continuous service. Therefore it is customary to provide duplication of those elements of greatest importance or most likely to be out of service, and to design the remaining apparatus so that failure is highly improbable and emergency means can be used to restore service quickly.

Duplicate service is seldom provided to residences, but is often provided to hospitals, municipal services and major industrial plants, particularly where material in process may be subject to severe loss during even a short power interruption. Also, substations may be interconnected by several paths, either directly or through other substations en route, resulting in loop systems, multiple radial systems, close-linked systems, loose-linked systems and others too diverse to describe in detail here.

Network Systems.—In metropolitan areas forms of distribution systems known as network systems have been developed. Begun in New York city in 1922, the secondary network system provides that the secondaries of the final transformers, instead of feeding only a few customers, will all be connected directly together over a wide area by large conductors to form a network, and all users will be fed from these secondary mains running along each street. The transformers, instead of being small units, are larger, up to 500 kw. or more, and are frequently fed directly from the generating stations at generator voltage, often 13,800 v., by underground cables. The transformers are located in underground vaults, usually under the streets.

Network protectors—special circuit breakers connecting the transformers to the secondary mains—open whenever power feeds from the network to the transformer, even in small amounts. This may occur when the primary circuits are opened at the generating stations and the transformer idling losses are supplied from the network, or when a defect occurs in a transformer. Otherwise, whenever power can feed from the transformer to the network, the protector closes. Short circuits in the low-voltage network, which rarely occur because of the large conductors, well insulated and underground, are cleared by the melting of the conductors in the region of the short circuit or by the melting of reduced sections known as limiters. Since every main is fed by at least two transformer banks, one at each end, the first melting of the cable does not cause any interruption in service to consumers. Special modifications are used in industrial plants and large office buildings. This system is economical in regions of high load density.

BY COURTESY OF WESTINGHOUSE ELECTRIC CORP.

FIG. 7.—NETWORK TRANSFORMER BEING INSTALLED IN UNDERGROUND VAULT

Used for low-voltage distribution in metropolitan areas, transformers of this type, with network protection mounted directly on the transformer (foreground), give reliable network service for dense load areas

A modification known as the primary network system extends a similar plan to supply the distribution circuits feeding the final distribution transformers in residential areas. These circuits are usually overhead, and fuses or relays and circuit breakers are relied upon to isolate defective circuits. However, each distribution circuit is fed from two sources, so that a high degree of reliability is achieved with relative economy. This plan was not as widely followed as the secondary network system.

(W. A. L.; E. R. DE L.)

COMMERCIAL ASPECTS

Following World War I the need for co-ordination of electricity supply undertakings, private and governmental, was generally appreciated. For economy in widespread electrification, it became necessary for governments to control the concentration of generation, transmission and distribution facilities. National or regional power authorities were established in developing countries, *e.g.*, in Australia, Canada and New Zealand, to be responsible for all aspects of electrification. In Great Britain the Central Electricity board was set up in 1926 to construct the 132-kv. national grid as a primary transmission system—interconnecting the main power stations and giving bulk supplies to distribution undertakings—and to standardize the frequency at 50 c.p.s. In 1943 the North of Scotland Hydro-Electric board was established and in 1948 the British electricity supply industry was nationalized, under the British Electricity authority (later the Central Electricity authority), with 14 area boards.

In the United States, where both publicly and privately owned electric power facilities exist, governmental regulation and co-ordination extend from the federal to the municipal levels. The federal government does not sell electric power direct to consumers, except in the case of a few large industries working on

military contracts. However, it does play an important role in the generation, transmission and sale of electricity. The federal agencies having a direct bearing on the electric utility industry include the Federal Power commission (FPC), the department of the interior, the Atomic Energy commission, the Tennessee Valley authority, and the Rural Electrification administration of the department of agriculture.

For the encouragement of discussion of common problems and the exchange of technical information, several international organizations were initiated. The Conférence Internationale des Grands Réseaux Électriques, the World Power conference and the Union Internationale des Producteurs et Distributeurs d'Énergie Électrique dealt with transmission, generation and consumption problems respectively, while the International Electrotechnical commission was concerned mainly with questions of standardization.

World War II caused such disruption of European electric power systems that there was a serious lack of generating facilities and heavy deficiencies in energy production during the immediate postwar years. A plant deficit of 10,000,000 kw. in 1945 was, however, largely overcome within ten years through the construction of new power stations, while interconnection of these systems improved the reliability and economy of the service. The Electricity committee of the Organization for European Economic Co-operation (O.E.E.C.) prepared the electrical part of the European Recovery program in 1948 and subsequently made several studies of future energy needs and methods of providing for them. In 1951, the Union for the Co-ordination of the Production and Transmission of Electricity (U.C.P.T.E.) was formed to ensure optimum utilization of the power capacity of the interconnected systems of western Europe.

The Economic Commission for Asia planned an expansion of electric power facilities to increase the installed capacity in Asian countries to more than double the 1950 total capacity of some 4,000,000 kw. Postwar reconstruction of electric power systems and development schemes for new generating stations, both thermal and hydroelectric, were assisted financially by the International Bank for Reconstruction and Development.

The high rate of increase in the demand for electrical energy during the first half of the 20th century led to a corresponding rise in the installed capacity of generating plants. At first, an apparent abundance of coal and much undeveloped water power in the industrialized countries allowed production to advance without anxiety about the exhaustion of resources. The discovery of large reserves of petroleum, which was also used for electricity generation in some thermal power stations, had the same effect. But as the rate of use of these resources in relation to the total proved reserves rose very rapidly, the position became disquieting. In the decade following World War II the electricity supply industry was, therefore, much occupied with the introduction of new sources of energy. While many sources were studied intensively and plans were made to develop them, it was soon clear that nuclear energy gave the most promise as a major source for the future.

By the early 1960s, the annual world production of electric power had risen to more than 2,228,000,000,000 kw.hr. The annual per capita consumption in the leading countries (including power produced by industries for their own use) was approximately as follows: Norway, 7,400 kw.hr.; Canada, 5,400 kw.hr.; United States, 4,200 kw.hr.; Sweden, 4,000 kw.hr.; and Luxembourg, 3,800 kw.hr. (E. W. G.)

United States

Privately Owned Power Companies.—In the early 1960s there were about 400 privately owned electric power companies in the U.S., ranging in size from those with assets of only about $100,000 to those with assets of more than $7,000,000,000. About 260 of these companies provided electricity to 98% of the customers served by the privately owned segment of the industry.

The U.S. in this period led all nations in the world both in installed generating capacity and in electric energy production; with a generating capacity of about 183,000,000 kw., it produced approximately 800,000,000,000 kw.hr. annually. The private companies accounted for about 127,000,000 kw. of the total generating capacity and publicly owned utilities, about 38,000,000 kw. Plants capable of producing approximately 18,000,000 kw. were owned and operated by industries, including railways, for their own use.

Private electric utilities are regulated by state commissions or city governments with respect to their intrastate or intracity operations. All of the states had established regulatory bodies by the early 1960s. The various state laws require that power companies be permitted to earn a reasonable profit on that part of their investment which is devoted to public use. The state agencies determine a reasonable profit level by taking into consideration: competitive cost of money; ability to attract money for expansion; financial condition of the company; effects of inflation; physical condition of the property; degree of risk assumed by the investors; and the operations of the company. This does not mean that the regulatory agencies guarantee any particular profit, nor does it mean they will reimburse a company for any losses which might be sustained. However, the agencies usually will order a company to reduce rates if its earnings are considered too high or permit it to raise rates if earnings are too low.

The primary responsibility of the regulatory commissions is to protect utility customers from poor service or unfair charges resulting from a misuse of the monopoly status of the utilities. The commissions must also assure fair treatment for utility investors. Not only are investors a part of the public, but good service depends upon the willingness of investors to supply money to the private utilities.

In addition to state and municipal agencies, the federal government also exerts regulatory influence, particularly upon utilities that operate across state lines, utilities that are licensed by the federal government to develop water power on navigable streams, and utilities that are parts of holding-company systems. The Federal Power commission has authority to regulate the rates and conditions of service where electricity is sold at wholesale in interstate commerce, and it also reviews and approves proposed rates for the sale of electric energy from certain federal and international projects.

The Securities and Exchange commission, under the Public Utility Holding Company act of 1935, has certain powers over holding-company systems, and it also requires electric utilities to provide the information necessary to permit investors to arrive at sound decisions about the desirability of securities which the utilities issue.

Rural Electrification.—Rural electric co-operatives, owned and operated by the people they serve, provided service in every state except Connecticut, Hawaii, Massachusetts and Rhode Island in the early 1960s. More than 96% of U.S. farms were electrified, as compared with 10% when the Rural Electrification administration (REA) was established in 1935, and co-operatives served 51% of these farms. (*See* Rural Electrification.)

Federal Control.—The Federal Power commission issues permits for the planning and licences for the construction and operation of nonfederal hydroelectric projects on waters or lands subject to U.S. jurisdiction. It studies plans for dams to be built by the U.S. army corps of engineers, the department of the interior, or any other federal agency, and it makes recommendations concerning the installation of penstocks and similar hydroelectric facilities at proposed federal projects. The FPC also co-ordinates the efforts of federal agencies toward the most comprehensive development and use of the water and related land resources of the U.S., and issues permits for facilities on the international borders which are to be used to export or import electric energy to or from Canada or Mexico.

The Federal Power act of 1920 limited the licences for hydroelectric power sites to 50 years and provided for recapture by the federal government at the end of that period. Upon the expiration of the licence, the FPC may take one of three possible actions: it may order the project surrendered to the federal government upon payment of a fair price; it may issue a new 50-year licence to the original holder; or, if the original holder does not desire a renewal, it may issue a licence to a new applicant.

The Federal Power act also contains a "preference" clause stipu-

lating that states, municipalities or other political subdivisions be granted preference over private applicants in the licensing of proposed hydroelectric projects. The preference clause is used only when the proposals of both public and private applicants are of equal merit.

In the original Reclamation act of 1902, the department of the interior was authorized to begin construction of water-use facilities to provide for irrigation and municipal water supplies in the arid regions of the west; there was no mention of the production or sale of electric power. In 1906, congress amended the act so that surplus power generated at the dams of irrigation projects could be leased, giving preference to municipal purposes, for periods not exceeding 10 years. By law, revenues from the sale of electricity generated at federal projects must be used to repay the cost of installing the power facilities. This repayment must be made within a prescribed period of time, usually 50 years.

Federal Power Production.—The department of the interior has four subdivisions which deal with the generation and sale of electric energy: the bureau of reclamation; the Bonneville Power administration; the Southwestern Power administration; and the Southeastern Power administration. The total amount of energy marketed annually by these agencies is approximately 50,000,000,000 kw.hr.

The bureau of reclamation is concerned primarily with the irrigation of arid and semiarid lands, and the power it markets is generated at its irrigation projects, at flood control projects of the army corps of engineers, and at certain reservoir projects of other federal or international agencies. In the early 1960s, the bureau operated about 40 power plants with a total capacity of 5,137,550 kw. In addition, it marketed power at four corps of engineers' plants, having a capacity of 745,000 kw., and the 31,500-kw. Falcon dam on the Rio Grande river.

The three remaining department of the interior subdivisions deal primarily in power production and marketing. The Bonneville Power administration was created on Aug. 20, 1937, to market power at the Bonneville dam on the Columbia river in Oregon and Washington; it later was given responsibility for distributing power produced at more than a dozen other dams in the northwestern U.S. The Southwestern Power administration was created in 1943 to market power generated at corps of engineers' projects in Arkansas, Missouri, Oklahoma and Texas, while the Southeastern Power administration was established in 1950 to sell power generated at 10 corps of engineers' projects in the southern and southeastern states.

The objective of all four marketing agencies is to dispose of power from federal projects in such a manner as to encourage the most widespread use of electricity at the lowest possible rates to consumers consistent with sound business principles, giving preference to public bodies and rural co-operatives.

The Tennessee Valley authority (TVA), created by an act of congress on May 18, 1933, developed into the largest single electric power system in the world. It is charged with the development of the Tennessee river and its tributaries in the interests of navigation, flood control and the generation and distribution of electric power. In the early 1960s it was producing annually more than 60,000,000,000 kw.hr., of which about three-quarters was generated by steam-electric plants and the remainder by hydroelectric plants. Its installed generating capacity was more than 11,000,000 kw. *See* TENNESSEE VALLEY AUTHORITY. (A. J. MACD.; X.)

GREAT BRITAIN AND OTHER COUNTRIES

Great Britain.—Under the responsibility of municipal and private supply undertakings, which were governed by various electricity acts from 1900 onward, the position of electricity supply in 1920–21 (excluding the north of Scotland area) could be summarized as follows: units (kw.hr.) sold, 3,100,000,000; number of consumers (estimated), 1,092,600; average price per unit, 2.482*d*.

World War I showed the need for planning on a national basis, and electricity commissioners were appointed in 1919 to promote and regulate regional schemes and to reorganize the industry by voluntary agreements. In 1925 there were 572 electricity undertakings and 438 generating stations. The Electricity act of 1926 established the Central Electricity board to be responsible for main transmission and interconnection of selected generating stations giving bulk supplies to distributing authorities. The construction of the 132-kv. national grid was completed, with standardization of the supply frequency at 50 c.p.s., by 1938.

The outbreak of World War II caused the postponement of further measures for the reorganization of distribution and reduction of the number of undertakings. In 1940–41 the position was: units sold, 23,818,000,000; number of consumers, 10,358,800; average price per unit, 1.038*d*.

Progress in construction and in the spread of rural electrification during the war was dictated by economic and military necessity. At its close the serious shortage of generating plant and the need for reorganization became obvious. For several years, "load shedding" and the discouragement of energy consumption were resorted to and it was not until 1955, after nationalization of the industry from April 1, 1948, that such steps became unnecessary. The British (later Central) Electricity authority had 14 area boards covering all of the country with the exception of the north of Scotland, which was the responsibility of the North of Scotland Hydro-Electric board, established in 1943. In 1955 two boards were combined to form the South of Scotland Electricity board.

The number of generating stations was reduced, and new ones were built, under a succession of annual construction programs which increased the total installed capacity from 11,680 megawatts (Mw.) in 1947 to 24,235 Mw. at the end of 1957. In 1955 a national ten-year program for the construction of 12 nuclear power stations, of total capacity 1,500 Mw. to 2,000 Mw. (increased in 1956 to 6,000 Mw.), was formulated and four of these stations were begun in 1957 and 1958. Construction of a 275 kv. supergrid started in 1954.

Hydroelectric plans were started in Wales and in Scotland. In Northern Ireland responsibility for electricity supply was vested in a separate supply board, together with Belfast corporation, and there were small independent undertakings in the Isle of Man and in the Channel Islands.

Commonwealth of Nations.—*Australia.*—Public electricity supply was developed mainly under state electricity commissions with their own networks. Coal, brown coal and lignite were used for generation, but water power, particularly in the eastern section of the Australian Alps and in Tasmania, was being developed. The estimated total hydroelectric power resources were 3,400,000 kw. on the mainland and 1,250,000 kw. in Tasmania. The installed capacity of 4,860 Mw. included 15.7% of hydroelectric power, more than half of which was in Tasmania.

Canada.—Development of electricity supply, based on the abundant water power, was undertaken through provincial power commissions and private organizations. Some of the commissions were responsible for generation as well as transmission and distribution, while others bought energy in bulk and distributed it. By the early 1960s, annual production was more than 103,000,000,000 kw.hr., about three-quarters of which was produced by hydroelectric plants.

Ceylon.—The Ceylon government sponsored the development of electricity supply and supervised electric power schemes by local authorities. After 1952 such developments were aided by the Colombo plan. Between 1950 and 1957 the production of electrical energy for public use rose from 82,000,000 to 210,000,000 units. By 1960 the installed capacity of generating plant was about 65 Mw., of which 34 Mw. was thermal and 31 Mw. hydroelectric. The water power resources have been estimated as between 500 and 1,000 Mw. and these were being developed to reduce the thermal capacity needed.

India.—Although electricity supply started in 1903 under the Indian Electricity act, progress was slow until nationalization in the 1940s. Both the central and provincial governments undertook developments and the annual production of electrical energy for public supply rose from about 2,400,000,000 units in 1939 to more than 10,900,000,000 units by 1960, though this was still only 28 units per capita. Installed capacity was about 3,550 Mw., of which 2,650 Mw. was thermal. Of the estimated hydroelectric potential (40,000 Mw.), less than 3% had been exploited.

New Zealand.—Regional power boards were responsible for development and in 1945 a Rural Electrification Reticulation council was established to co-ordinate rural supplies. With abundant water power, progress in electricity production was rapid and almost 95% of the total electrical energy for public supply (5,275,000,000 units) was generated from this source. Installed capacity was 1,272 Mw., all but 72 Mw. being hydroelectric. In the North Island little hydroelectric power remained for development and geothermal power was used. The first 69-Mw. geothermal plant was in operation in 1958.

Pakistan.—In 1951 electricity production was only 226,000,000 units (3 units per capita) but the 75 Mw. of this installed capacity was increased to 227 Mw. by 1957, when the annual production reached 1,220,000,000 units or 14 units per capita. Hydroelectric development of 380 Mw. by 1965 was planned for West Pakistan although the water resources were distant from load centres.

Rhodesia and Nyasaland.—In Southern Rhodesia electricity supply was the responsibility of an Electricity Supply commission established in 1936. Heavy power demands by the mines led to rapid development, and by the 1960s the total production of electricity for public supply exceeded 2,700,000,000 units, of which more than 90% was generated thermally. Installed capacity was about 700 Mw., all but 40 Mw. being thermal. But construction of the Kariba gorge scheme, on the Zambesi river, with an eventual capacity of 1,000 Mw., progressed.

Europe.—Electricity supply followed different patterns according to the degree of industrialization and the availability of fuels or water power. Private enterprise was responsible for the earlier developments but, especially after World War II, nationalization spread until most countries had government-sponsored electricity authorities. The great damage done to power systems during that war led to a total European deficiency in generating plant of some 10,000,000 kw. in 1945. Reorganization was started under the European Recovery program, assisted by plans made by the Electricity committee of the Organization for European Economic Co-operation. In the following years the Electric Power section of the United Nations Economic Commission for Europe studied the question with particular reference to international interconnection of supply systems. In 1951 the Union for the Co-ordination of the Production and Transmission of Electricity (U.C.P.T.E.) was set up and this encouraged international exchanges of energy between its members—France, Belgium, Netherlands, Luxembourg, Federal Republic of Germany, Austria, Switzerland and Italy. The total energy exchanged within this group was about 3,000,000,000 kw.hr. in 1950 and about 6,000,000,000 kw.hr. in 1955. By the middle 1950s the deficiencies had been largely made up and interconnected transmission networks operating at 220 kv., with sections at 380 kv.—which was standardized for future transmission—were functioning satisfactorily. This allowed advantage to be taken of the differences in the annual regimes of thermal and hydroelectric systems and so reduced the total installed capacity, on an international basis, needed to meet power demands. Hydroelectric projects were developed rapidly, particularly in the Scandinavian countries, France, Italy, Switzerland, Austria and the U.S.S.R., but in some countries the end of economically exploitable plans was in sight. The potentialities of nuclear power generation, geothermal energy, tidal power and wind power were studied for the future.

Another important interconnection was that between Denmark —a country lacking in energy resources—and Sweden. For about 40 years preceding 1956, exchange of energy had taken place, largely from Sweden to Denmark. The capacity of the connecting link was 150 Mw. and up to 300,000,000 kw.hr. annually was imported by Denmark.

Norway, with an estimated year-round hydroelectric potential of 12,500,000 kw., corresponding to an annual energy production of about 120,000,000,000 units, had the highest per capita consumption in the world. About 26% of the Norwegian hydroelectric stations were state owned. Of the total annual consumption, 40% was taken by electrochemical industries and 36% by farms and residential and business houses.

Iceland had technically exploitable water power of 4,500,000 kw. to produce 38,000,000,000 units per annum, as well as geothermal resources and peat deposits. There was, however, difficulty in finding outlets for the energy which could be produced.

Central and South America.—Except in the large cities and in some exceptional small areas, electricity consumption remained low. If Mexico is included with the countries of Central America, the average annual per capita consumption was only about 240 kw.hr., while for South America it was about 300 kw.hr. Development of electricity supply was mainly through private undertakings or by municipal authorities but, in the 1950s, national authorities took an increasing share. There were very small resources of coal, but petroleum was produced on a significant scale in ten countries, particularly in Venezuela. These countries also had resources of natural gas. Hydroelectric power potential (estimated as some 60,000 Mw., based on ordinary minimum flow) was high and had been little developed by the middle of the century.

Indian and Pacific Oceans.—The territories enclosed in an area 5,000 miles square in the southern Pacific ocean include Australia, New Zealand and Ceylon. Excluding these, electricity supply was little developed and generation was mainly from thermal sources.

The Hawaiian Islands, Indonesia and the Philippines were the main components with a total population of about 49,000,000. Each had an annual production of 1,500,000,000 to 1,900,000,000 units but, while the consumption per capita in Hawaii was about 2,400, that in Indonesia and the Philippines was only about 20 and 85, respectively. Papua and New Guinea had undeveloped deposits of coal and lignite and an estimated hydroelectric potential of about 15,000 Mw.

Africa.—The position of electricity supply in so large an area was, of course, very varied. Proved reserves of coal and lignites were about 36,000,000,000 tons, with a probable total about ten times as large. Relative to the rest of the world, hydroelectric resources were much larger at 187,000 Mw. (at ordinary minimum flow) though only about 1,500 Mw. of this had been developed. Discoveries of oil and natural gas in the middle 1950s improved the position, particularly in Algeria. The average annual consumption of electrical energy per capita for the whole continent was then 134 kw.hr. but, without South Africa, this figure reduced to 55 kw.hr.

Republic of South Africa.—The Electricity Supply commission, set up in 1922, was mainly responsible for electrical development. In 1948 it purchased, for £14,500,000, the Victoria Falls and Transvaal Power company, one of the largest electricity supply undertakings in the commonwealth. The commission built large thermal power stations, and energy, generated from local coal, was cheap. By the early 1960s the annual production, of which 56% went to the mining industry, was more than 22,000,000,000 units.

Asia.—This area (excluding U.S.S.R.), which contained almost exactly half of the world population, was largely underdeveloped in electricity supply except for a few countries, notably Japan, Israel and North Korea. In the middle east, proved reserves of petroleum were more than half of the world total of 35,706,000,000 tons; coal and lignite proved reserves were only 12,000,000,000 tons and potential water power (at ordinary minimum flow) was 67,000 Mw., of which about 21% had been developed, mainly in Japan. Under national authorities responsible for electrical development, progress was fairly rapid.

See also references under "Electric Power" in the Index volume.

(E. W. G.)

BIBLIOGRAPHY.—National Electrical Manufacturers' Association, *Chronological History of Electrical Development From 600 B.C.* (1946); L. K. Kirchmayer, *Economic Operation of Power Systems* (1958); A. T. Starr, *Generation, Transmission and Utilization of Electric Power* (1957); Organization for European Economic Cooperation, *Interim Report on the European Recovery Programme* (1948), *The Electricity Supply Industry in Europe* (1954), *Production, Transmission and Distribution of Electricity in Europe* (1957); *Proceedings* of the Geneva Conference on the Peaceful Uses of Atomic Energy, Aug. 1955, vol. i (1956); *Transactions* of the Fifth World Power Conference, Vienna, 1956 (1957); *Proceedings* of the Tenth British Electrical Power Convention, 1958 (1958); American Society of Mechanical Engineers, *Nuclear Reactor Plant Data*, vol. 1, *Power Reactors* (1959); J. Guthrie Brown, ed., *Hydro-Electric Engineering Practice* (1958); H. K. Barrows, *Water Power Engineering*, 3rd ed. (1943).

ELECTRIC TRACTION. The term electric traction is applied to those means of transport in which the vehicles are powered by electric motors supplied with electricity from distant generating stations. Such systems include (1) the streetcar or, in British terminology, the tram, operating typically in single, although sometimes in multiple, units, primarily on track laid in the streets; (2) the electric trolley bus, operated on the streets on rubber tires; (3) the interurban railway, with single-unit or multiple cars operating between cities, the track laid primarily on private right of way; (4) rapid-transit lines, operating multiple-unit electric trains on private rights of way, often underground or elevated in metropolitan areas; and (5) electrified main-line railroads, including suburban service, and trunk-line operations powered by electric locomotives. Electricity is also used in the propulsion of vehicles in other ways, but these forms are not usually included in the category of electric traction. Examples include the battery-powered electric automobiles of the early decades of the 20th century; battery-propelled vans for city delivery or warehouse use; and the modern diesel-electric locomotive, whose wheels are driven by electric motors powered by diesel engines.

Historical Development.—The basic principles of electricity were known for nearly a century before successful electric-traction motors were developed, and thus throughout the period in which use of the steam locomotive was expanding rapidly. As early as 1834, a Brandon, Vt., blacksmith, Thomas Davenport, built a small battery-powered electric motor and used it to operate a small car on a short section of track. Four years later, Robert Davidson of Aberdeen, Scot., built a five-ton electric locomotive which propelled itself on an experimental run on the Edinburgh-Glasgow line, but the project ended when a group of locomotive drivers and stokers destroyed it at Perth to end the threat to their jobs. In 1847, Moses Farmer, in an experiment at Dover, N.H., operated a car which could carry two people. But the most important of these early undertakings was that of Charles G. Page of the U.S. patent office, who in 1851, with financial aid provided by congress, built a small electric car and operated it from Washington, D.C., to Bladensburg, Md., at speeds up to 19 m.p.h. But the experimental run suffered many mishaps, and congress would provide no more funds.

The great obstacle in the way of success of these experiments was the need to rely on storage batteries as a means of power. Since these were expensive and inefficient, they seriously limited the potentialities of electric traction. Not until the early 1870s was the principle of generating electricity by a dynamo, that is, rotation of an electric motor by mechanical means, developed by Z. T. Gramme and Hippolyte Fontaine. This invention paved the way for elimination of the use of batteries and the development of commercial electric traction. The first successful non-battery electric locomotive was developed by the German engineer Werner von Siemens, and exhibited at the Berlin industrial exhibition in 1879. Power was supplied through a third rail. Two years later Siemens built the first commercial electric line, a mile and a half in length, at Lichterfelde, near Berlin. Power was fed through the two rails (which were insulated from each other), but trouble was encountered because of short circuits in wet weather and unfortunate shocks to horses and pedestrians crossing the tracks, even though only 100-v. power was used. A similar type of vehicle was developed by Thomas Edison and operated on an experimental basis at Menlo Park, N.J., in 1880. One primary problem remained to be solved before successful commercial operation was possible, namely, that of improved methods of transmitting power. Use of the running rails, or a third rail, was not only dangerous but conducive to shorting and loss of power. A number of persons sought to solve the problem during the 1880s, with consequent lawsuits over patent infringements and differences of opinion as to appropriate credit for the developments.

One approach was modeled on cable-car operation. The third rail was suspended in a slot between the rails, and contact provided by a "plow." Such a system was built in 1884 in East Cleveland, O., by Edward M. Bentley and Walter H. Knight, and operated for nearly a year, until constant mechanical difficulties led to its replacement by horses. The conduit principle was later used in New York city and Washington, D.C., which banned overhead wires in the downtown areas, but higher cost and water hazards caused companies elsewhere to avoid it. Most of the other inventors concentrated on the use of overhead wires. In 1884 a Kansan, John C. Henry, operated a two-wire overhead system on an experimental basis. He sought to solve another basic problem, that of controlling speed, by use of a complicated gear box similar to that employed on automobiles. Prior to this time, the cars, just as cable cars, had to run with full power if they were to move at all. Meanwhile, Leo Daft, a former photographer who operated a small electric power company in New Jersey, was experimenting with various types of electric traction. In 1885, he placed in operation a three-mile commercial line in Baltimore, Md., using an electric locomotive and third-rail power. This line did not succeed and was replaced by horses; and finally Daft shifted to the two-overhead-wire system, power being gathered by a small carriage or "troller" running on the wires. In the late 1880s a number of smaller transit systems were electrified with the Daft system. From the little troller of the Daft system came the word trolley, widely used in subsequent years in conjunction with streetcars.

In this same period, a Belgian *émigré* to the United States, Charles J. van Depoele, developed a somewhat similar system using only one trolley pole and wire, the rails being used as the return circuit. He demonstrated the system at the Toronto agricultural fair in the summer of 1885, hauling large numbers of passengers. On the basis of the success of this operation, a line, one of the first intercity routes, was installed from St. Catharines to Thorold in Ontario, and the Montgomery, Ala., system was electrified by Van Depoele in 1885–86. The original Van Depoele system involved a two-wheeled carriage which ran along on the wire and was connected to the car by a flexible cable. But this system proved unsatisfactory, particularly when cars met; and he replaced it with a trolley pole and contact bearing against the bottom of the wire, a practice which became standard.

The last of the group of major inventors was Frank J. Sprague, a serious student of electricity, rather than an amateur. He became interested in a project for electrifying the London underground, and then returned to the United States to work with Edison. But Edison was suspicious of the academically trained Sprague, who soon left to form his own company to produce and install electric-traction facilities. In 1886 he received the contract to electrify the Richmond, Va., system, which he succeeded in doing despite many technical obstacles. An underrunning trolley wheel replaced the contact of the Daft system, with a rope attached to the pole for adjustments. For the first time the motors were mounted directly on the axles instead of on the platforms, thus eliminating the chain drive of the Van Depoele cars. Sprague also utilized a rheostat to control the flow of power and thus the speed, and developed the system of connecting the two motors in series when the car started up. Operation of the Richmond system began in Feb. 1888. While this was by no means the first streetcar system—being preceded not only by the one in Montgomery but also by installations in Appleton, Wis.; Lima and Mansfield, O.; Scranton, Pa.; Detroit and Port Huron, Mich.; and Windsor, Ont.—it is generally regarded as the first successful large-city operation.

For the next several years a bitter competitive struggle in the supplying and installation of electrical equipment was waged between the Sprague enterprise, which was soon absorbed by Edison General Electric, and the Van Depoele undertaking, which was merged into the Thomson-Houston company. Eventually General Electric also absorbed Thomson-Houston, but was soon faced by strong new competition from Westinghouse. For decades these two companies dominated the field. Streetcars themselves, however, were built by a number of companies. The general principle of streetcar operation remained the same over succeeding decades as that introduced by Sprague, although various technical improvements, such as the replacement of metal by carbon brushes in the motors, were made. One of the last major inventions, also by Sprague, was the multiple-unit (MU) control system, which allowed a number of cars run in multiple to be controlled by the train motorman. First introduced on the South Side Elevated line

in Chicago in 1895, MU revolutionized urban and suburban rapid-transit systems.

Predecessors of the Streetcar.—Once the first successful installations were made, the use of the streetcar spread very rapidly, and in a decade it came to dominate the urban transport picture. Before considering this growth, however, some reference to the earlier history of urban transport is necessary. Paris is generally considered to be the site of the first public transport, in 1819, and in 1831 the first such vehicle operated in New York city. These were horse-drawn omnibuses, running on the streets without rails. But following the example of the railroad, the practice of laying rails in the streets for the cars, to permit a smoother ride and more efficient use of power, developed. On Nov. 14, 1832, the New York and Harlem Railroad company established street railway service with horsepower along the Bowery between Prince street and 14th street, the first of its character in the world. This mode of transport spread very slowly until the late 1850s, when it was introduced into Brooklyn (1853), Boston and Cambridge (1856), Philadelphia (1858) and Chicago, Cincinnati and Baltimore (1859). In the next two decades the horsecar spread more rapidly, until all large cities and many small towns, some with only a few thousand people, had their lines. In 1890 there were 6,600 mi. of horsecar track in the United States, with 28,000 cars and 105,000 horses and mules for power.

In Great Britain, an unsuccessful attempt was made to gain legal permission for a tramway in London in 1857; but in 1860, an American, G. F. Train, opened a line in Birkenhead and three lines in London. Lines were established at Salford in 1862 and Liverpool in 1865. Granting of powers to establish tramlines was transferred from parliament to the local authorities in 1870. The Tramways act effecting the transfer placed detailed obligations on the companies and essentially gave the local authorities complete control over the enterprises. In the next two decades, as in the U.S., horse-tram systems spread rapidly in Great Britain, as they did also in Europe.

The first competitor of the horsecar was the cable car, which was introduced on Sacramento and Clay streets in San Francisco, in 1873, the invention of Andrew Halladie. Cars were drawn by an endless cable running in a slot between the rails and passing over a steam-driven shaft in the powerhouse. The system was particularly adapted for use on steep hills, and reached its most extensive use in San Francisco and Seattle. There were also installations in Washington, D.C., New York, Chicago and a few other cities. But the heavy cost of installation and the rise of the electric car checked its expansion. While the cars operated much more smoothly than the early electric cars, they suffered from the need for running at a constant speed; and breaking or jamming of the cable tied up all cars on a particular line. At the peak, around 1890, about 488 mi. of line were in operation. Most of the trackage was quickly replaced by electric cars, but the Seattle lines lasted until the 1930s, and a portion of the San Francisco system was still in operation in the 1960s, preserved in later years largely as a tourist attraction. Some lines also were installed in Europe.

The Spread of the Electric Car in the United States.—Commencing with the electrification of the extensive West End Street railway in Boston in 1889, following the success of the Richmond system, the electric car replaced the horsecar and cable car at a rapid pace; and many new lines were installed. In 1890, there were 1,260 mi. of electric line and in 1902, 22,000 mi. Only about 250 mi. of horse and cable lines then remained. In larger cities, large companies were formed from the small independent lines which had gained the original franchises; and one company soon gained control over all operations. These were often affiliated with electric power companies. New York city, which had a number of companies for many years; Washington, D.C., with two companies; and a few other cities were exceptions to the unified operations rule. The enterprises operated under municipal franchises, which typically required the companies to pave the area between and adjacent to the tracks, and limited the fare, frequently to five cents. In a number of cities the companies engaged in a long battle with the water systems operators over electrolysis damage to water mains, but this was finally eliminated by careful bonding of rails so that the current would not wander off into the ground. The Cincinnati system employed two trolley wires to solve the problem.

From a technical standpoint, motors and cars were gradually improved. The tiny single-truck (four-wheeled) cars gave way to heavy two-truck (eight-wheeled) cars with much greater carrying capacity; the monitor roofs were changed to arch roof design; and wood bodies were replaced by steel. However, many companies were hard-pressed to meet demands of traffic in rush hours, and many of the obsolete cars remained in service for years. The steam-generated power was often replaced by power obtained from electric companies. The period from 1902 to 1917 was the one of most rapid expansion, and by the end of the period 80,000 cars were in service on 45,000 mi. of track. Chicago had the most extensive system, with about 1,000 mi. of line. Possession of a streetcar line was a "must" for a growing town; typically, any town of 10,000 or more, and many as small as 5,000, boasted systems. The larger city systems extended their lines farther out into suburban areas, in some cases engaging in land speculation schemes in conjunction with new car lines. It was a very common practice for a transit company to own an amusement park, often on a suburban line, to increase evening and week-end travel.

Expansion of the Tramways in Great Britain and Europe.—The spread of the tramway in Great Britain and the continent of Europe was equally rapid. During the 1890s and in the first two

G KRAMBLES FROM B. J. MISEK

FIG. 1.—CABLE CAR WITH TRAILER IN SAN FRANCISCO, THE FIRST AND LAST CITY IN THE UNITED STATES TO USE CABLE CARS

G. KRAMBLES FROM B. J. MISEK

FIG. 2.—STREETCAR TYPICAL OF THE EARLY PERIOD OF ELECTRIC STREET-RAILWAY DEVELOPMENT

decades of the 20th century, horsecar lines were replaced and extensive electric line systems built. By 1908 about 2,500 mi. of line were in operation in Great Britain, a figure which then remained stable for two decades. The development in Europe was equally rapid and continued over a longer period. The Siemens works became major suppliers of streetcar equipment not only for Germany but for many other countries as well; and other equipment firms developed in Switzerland, France, Italy and elsewhere. Many of the German cities came to boast of highly efficient streetcar systems, and throughout Europe the electric car became the chief means of urban transport. In the larger cities of Asia, Africa and South America, tramway systems were built, largely with English capital. Smaller cities in these continents were less well supplied, however.

The term tramway was generally used in Great Britain, and its equivalent in many European countries and in South America, whereas the word was virtually unknown in this sense in the United States and Canada. The term in its broader context refers to any line operating wheeled vehicles over track for the transportation of goods or persons; and thus the early railroads, especially before the use of steam locomotives, were often called tramlines. But with the coming of electric power, the term was narrowed to refer to the electric car operating on streets and rural roads.

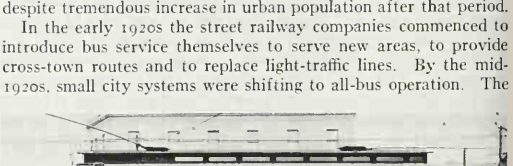
British tram operation was distinguished from American largely by the use of double-deck cars, which never passed beyond the experimental stage in North America. Pay-as-you-enter (PAYE) cars were rare in Britain, despite their widespread acceptance in the United States, and two-man operation was almost universal. This was likewise true in Europe, although in later years PAYE cars became more common. European cars were usually single-deckers and typically much smaller, comparable to the pioneer American cars and the Birneys of later decades. On the other hand, operation of tram-trains with two or three cars, even in small cities, was common. Zone fare systems were much more common in both Britain and Europe than in the United States. Cars very frequently used bow or pantograph power collectors, in contrast to universal use of the single trolley pole in the United States. Lines in South America and Asia typically used English, German or Swiss equipment and operating methods.

The local control of tramways in Great Britain gave rise to difficulties in areas in which two or more municipalities were contiguous, by checking development of through service. As a consequence, the Light Railways act of 1896 set up a board of commissioners to control intercity tram and light railway operations, and cities lost their power to veto construction of such routes. As a consequence, in some areas, new companies were formed which brought together the tramways in adjacent cities and provided more unified service.

The Decline of the Streetcar and Tramway.—The typical streetcar enterprise first began to encounter financial difficulties in the World War I period, particularly in the United States. As wage and materials costs rose, the companies were squeezed by the fixed fares, set almost universally by the franchises; and city governments were most reluctant to permit fares in excess of five cents. The situation became so bad that in 1919 the federal government appointed a commission to conduct an extensive inquiry into the problems of the industry. Gradually, by action of city governments or the states, fares were raised above five cents, and control over rates was frequently shifted to the state public utility commissions. The typical fare of the late 1920s was 10 cents, with discounts for tokens, often three for 25 cents; by the 1950s, 20 cents was a common figure, with 25 cents in a number of cities.

The other, and basically more serious, problem was the rise of the motor vehicle. The automobile was almost as old as the electric car but it developed much more slowly, and not until after 1912 was it regarded as a source of difficulty for urban transit. The first inkling of serious trouble came in the fall of 1914, when in Los Angeles a number of persons commenced to operate automobiles in transit service on major routes, carrying passengers for five cents. The jitney—so called for the five-cent or "jit" fare—spread rapidly across the country and seriously affected street railway business, as well as prejudicing transit officials against the motor vehicle in any form. The jitneys, plus competitive bus operations which developed in some cities from them, were for the most part gradually driven out of business by regulation. In some cities the local governments were forced to choose between the jitney and the streetcar and took the latter. Rarely was independent bus competition with streetcar lines permitted after 1925. However, as automobile ownership spread, use of public transport fell; the peak year in transit travel was 1923, despite tremendous increase in urban population after that period.

In the early 1920s the street railway companies commenced to introduce bus service themselves to serve new areas, to provide cross-town routes and to replace light-traffic lines. By the mid-1920s, small city systems were shifting to all-bus operation. The



G. KRAMBLES FROM B. J. MISEK

FIG. 3.—A TWO-TRUCK, MONITOR-ROOF CAR DESIGNED FOR TWO-MAN OPERATION, THE KIND USED MOST WIDELY IN THE U.S. DURING THE PERIOD 1900–20

direct operating expenses of the bus, per mile, were greater than those of the streetcar in this period, but the heavy expenses incidental to the trackage were avoided. Gradually the policy of bus substitution increased, although up through World War II the superiority of the streetcar over the bus for heavy-density routes in large cities was still generally accepted. However, after the war, substitution increased at an accelerated pace, so that by 1960 most cities (including, for example, Chicago, Cleveland, New York, Minneapolis-St. Paul, Buffalo, Houston, Seattle and Portland) had completely eliminated streetcar lines. In only a few cities, including Toronto and Pittsburgh, was the streetcar still the mainstay of public transit, although some major lines were still operated in other cities. However, many of the existing lines were scheduled for imminent replacement.

The substitution of buses for streetcars was in large measure a result of lower over-all cost, the relative costs of streetcar service being artificially increased somewhat by paving requirements. But the substitution was speeded up by restrictions on one-man operations of streetcars in some cities, a rule not applied to buses; by the desire of cities to have the rails removed from the streets; by the separation of transit and power enterprises under the terms of federal legislation of the 1930s; and by the absorption of control of a large number of city transit systems (about 80) by National City Lines, an enterprise financed by bus, tire and gasoline producers and committed to a policy of bus operation.

The inadequate profit situation from the World War I period

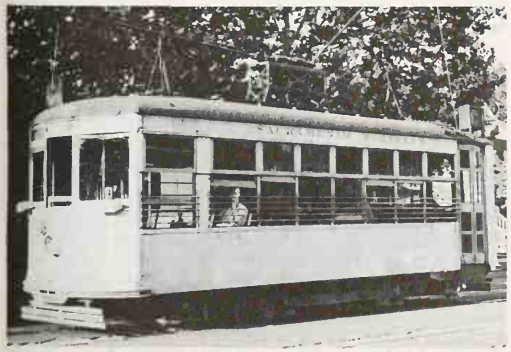
G. KRAMBLES FROM B. J. MISEK

FIG. 4.—A TYPICAL STEEL-BODIED STREETCAR OF THE 1920S

onward and the desire of many companies to retain rail service led to several important innovations in streetcar service itself, which slowed down bus substitution but in the end did not curb the trend. One was the introduction of the one-man car, which became almost universal by the late 1920s except in the larger cities. Another was the Birney car, a single-truck, lightweight, one-man car for light-traffic lines. Other lightweight cars were introduced during the 1920s. In the 1930s, after extensive research, came the so-called PCC car, named for the Electric Railway Presidents' Conference committee. Radical departures in design were embodied, including magnetic brakes, rubber cushioning of wheels (with a vulcanized "sandwich" of rubber and steel between the wheel and hub), improved and smoother acceleration, and streamlining. Nearly 4,000 cars of this type were placed in service in the following decade, with large fleets in Washington, Toronto, St. Louis, Minneapolis-St. Paul, Chicago, Pittsburgh and other cities; and some in relatively small systems such as those of Johnstown, Pa., San Diego, Calif., and Atlantic City, N.J. Even these improved cars, however, could not check the decline of the street railway.

Increased financial problems of streetcar systems and popular dislike of private monopoly in this field led some cities to undertake direct municipal operation. San Francisco was the first major city to do so, partly to permit construction of tunnels for improved service to outlying areas. Detroit, Boston, Cleveland, Seattle, New York, Chicago and some smaller cities eventually took over the systems, either directly or through transit authorities; in 1958 the Los Angeles system was acquired. Municipal management for the most part followed much the same policies in substitution of buses for streetcars as did private enterprise.

In Great Britain the tramway was replaced by the bus even more rapidly than in the U.S. The tramway systems (by this time mostly in municipal hands) at first used buses as feeders to the tramlines, but eventually became subject to substantial competition of private bus lines operating from suburban areas. Some systems proceeded to replace all trams at an early date, but others modernized their



G. KRAMBLES FROM B. J. MISEK

FIG. 5.—A SINGLE-TRUCK, ONE-MAN BIRNEY CAR, WIDELY USED ON LIGHT-TRAFFIC LINES IN THE 1920S

equipment and sought elimination of bus competition. In this period, serious problems were encountered in licensing the intercity bus operators by local authority, and in 1930 the Road Traffic act transferred control to regional traffic commissioners. Under the new control, direct competition of bus lines with tramways was generally prohibited; in some cities new tramlines were built as late as the 1920s. In the 1930s, with improvements in buses and introduction of the large-capacity double-deck bus, substitution continued at a rapid pace, especially as tram equipment and track wore out. Increased track costs consequent to World War II speeded tram abandonment, and by the early 1950s it was evident that with one or two exceptions all tramlines would be gone in a few years. London saw the end of its trams in the early 1950s. The last major tram system was that of Glasgow, which employed relatively modern double-deck cars. Traffic congestion on the streets and sharp increases in rail costs were additional factors leading to replacement, as well as general community redevelopment plans. From the standpoint of fuel availability, the replacement of electric by gasoline power was regarded in some circles with misgivings.

In European countries, the streetcar maintained its position much more securely. Some substitution took place; Paris saw the last of its streetcars in the 1930s, for example; and in other parts of France and in Italy many cities shifted to bus operation. But most of the German systems, even in relatively small places, remained largely intact; and extensive use of streetcars was still to be found in Scandinavia, the Low Countries, Switzerland, Austria and the countries of eastern Europe, where the lines were subject to much less auto competition. Most of the systems were municipal (Lisbon being a primary exception), and private bus competition was not permitted. In South America, by contrast, extensive replacement occurred; for example, the Caracas, Venez., tramlines were replaced as a part of a general municipal redevelopment plan. Some of the larger cities, however, such as Buenos Aires, Arg., maintained extensive tram operations.

The Trolley Bus or Trackless Trolley.—From the earliest days of streetcars, experiments were made with electric-powered vehicles supplied with energy from wires. Of necessity, two trolley wires were employed because of the lack of rails for a return circuit. In 1910 a vehicle of this type was placed in service in Laurel canyon in Hollywood, Calif., to replace buses on steep grades. In the next decade several similar installations were made in the U.S. Not until 1928, however, was the manufacture of trolley buses, designed as such, undertaken; and the first major installation was made in Salt Lake City, Utah. These buses offered the advantages of electric propulsion—more quiet operation, avoidance of fumes, and faster acceleration—plus curb loading, but were less flexible than the bus and necessitated use of overhead wires. Relative costs of the trolley bus and motor bus depended on the particular local circumstances, including traffic volume and cheapness of electric power. After 1950 relatively few new installations were made, and a few cities, including Salt Lake City, replaced their lines with motor buses. However, a number of cities, including San Francisco, Seattle, and Fort Wayne, Ind., relied heavily upon the trolley bus; and planned to continue operations with it. A number of installations were made in Great Britain, especially in London, but there was a stronger tendency than in the U.S. to replace them. Belfast, N. Ire., however, still relied almost entirely on this type of vehicle. In Europe, Italy was the stronghold of the trolley bus.

The Interurban.—The interurban railway was a peculiarly American institution, although its counterpart was found in some European countries. This type of operation was distinguished from streetcar service by the use of heavier, higher speed cars, and by the extensive use of private right of way outside of the cities. By contrast to the electrified main-line railway, it emphasized passenger rather than freight service in its heyday; used cars individually supplied with power rather than electric locomotives for passenger service; and had much more lightly built track. The typical interurban operated between cities, rather than merely into the suburbs of a metropolitan area.

The line usually regarded as the first interurban was completed from Newark to Granville, O., in 1889; the first major high speed

G. KRAMBLES FROM B. J. MISEK

FIG. 6.—THE PCC CAR, REPRESENTING THE LAST STAGE IN STREETCAR DEVELOPMENT IN THE U.S. INTRODUCED IN THE MID-1930S, IT WAS USED EXTENSIVELY BY LARGE-CITY SYSTEMS

road was built from Cleveland to Akron, O., in 1895. In the next decade interurban lines spread rapidly, particularly in the area between Ohio and Illinois, and a dense network of lines grew up, serving all major cities in the area. Limiteds were operated between larger cities, such as from Toledo to Cleveland and Columbus to Cincinnati, but a large portion of the business was local in character, from farm or village to nearby city. It eventually became possible to go from Freeport, in western Illinois, or Elkhart Lake, near Sheboygan, Wis., to Little Falls, east of Utica, N.Y., by interurban (with many changes of cars); and the completion of a few additional lines would have permitted travel from St. Louis, Mo., to New York or to points in central Maine. The greatest concentration of lines was in the states of Ohio, Indiana, Illinois and Michigan, although California had substantial mileage; and interurbans were to be found in the valleys of Utah and the fruit country of the Pacific northwest, in the piedmont area of the Carolinas, in the woods of northern Maine and the coal regions of Pennsylvania; and one line climbed over a 10,000-ft. pass in the Rocky mountains to get from Victor to Cripple Creek, Colo.

The earliest lines used streetcar type equipment, but this gave way to long, high-speed wooden cars, those designed and built by the Niles Car company being in appearance among the most distinguished pieces of transportation equipment ever used in the United States. Toward 1920, heavy steel cars came into use, and after 1920 a number of lightweight one-man cars, designed to speed up operations and reduce costs. The companies originally concentrated on passenger service, but after 1910 commenced to build up package-freight business. A number of lines, especially in the west, turned to carload interchange traffic with the main-line railroads.

The peak of interurban service was reached in the period shortly before World War I, with about 16,000 mi. of line in operation, plus more than 2,000 mi. of rural trolley lines in New England and Pennsylvania which operated at slow speeds with streetcar type equipment. The industry had never been too profitable, and the coming of the motor vehicle very quickly sealed its doom. The short-haul traffic on which it depended was particularly vulnerable to automobile use; and in some instances bus competition developed as well, although in other cases state regulatory policies prevented it. Abandonments commenced on a small scale as early as 1915, continued through the 1920s, speeded up rapidly after 1925. By 1929 even the large systems had begun to disintegrate, and the depression struck the final blow. By the end of the 1930s, the typical interurban was gone, except for scattered remnants. Some roads which had developed extensive carload freight business were able to survive as freight carriers; and some of these continued passenger service as late as the early 1950s, but by this time they had become essentially short-line railroads. The others which survived for a time were those which carried a heavy density of commuter traffic on high speed routes out of large cities. Chicago being the centre for three lines of this type. But these lines were also hard-pressed financially despite the stability of their traffic volume. The inability to increase volume of business per employee necessitated rate increases greater than the rise in the general price level, if profit were to be maintained; and these, if carried through, would have caused serious loss in business.

By the end of the 1950s, the last of interurban passenger service, strictly speaking, was gone and only four of the high-speed suburban operations remained. The last typical interurban passenger run, that from Thorold to Port Colborne, Ont., was discontinued in March 1959.

Great Britain, with its dense network of main-line railroads, never developed a system of interurbans, although there were intercity tramlines. However, a number of lines were built in Europe. The Vicinaux system in Belgium was the most complete, serving virtually every city and town in most of the country. Germany had a number of high speed interurban routes, and some lines were to be found in Switzerland (distinct from the electrified railroad system), the Netherlands, Austria, Italy, France and other countries. Interurbans were never common in the Scandinavian countries. As with the streetcars, a number of the European systems continued in operation through the 1950s, because of the lesser use of automobiles, but some abandonment occurred.

Minor Forms of Local Transit.—Apart from the types of local transit vehicles mentioned, experiments have been made with other forms. Steam-powered streetcars were used for a time, and steam power was employed extensively on the first elevated systems. As storage batteries were improved, renewed attention was given to this source of power for cars on light-traffic lines to eliminate the need for trolley wires, and some battery cars were used for a time in New York city, Florida and elsewhere. As late as the 1950s, Switzerland had one system of this type, the car batteries being recharged at each end of the line. For the most part, however, the coming of the motor bus ended the storage-battery car.

For many years, experiments were made with monorail cars; and in the early 1960s proponents of this system were proclaiming it as the solution to large-city transit problems, because of its cheaper construction compared with subways. While a number of lines were built in earlier decades, only one line continued to operate successfully, namely, that at Wuppertal, Ger. The objections to the monorail are the unsightly overhead construction required, and the difficulty of switching cars. It actually offers few if any advantages over an elevated railroad.

Rapid Transit.—The inadequacy of surface vehicles for carrying the traffic of large cities became evident in the middle of the 19th century, and the first underground railway was built in London in 1863, the first elevated in New York city in 1871. Steam propulsion was used, but had many disadvantages; and the lines were electrified, commencing with the Metropolitan West Side line in Chicago in 1895. In the succeeding years an extensive network of elevated and subway lines was built in New York city, the elevateds in Manhattan ultimately being eliminated because of their unsightly features. Boston and Philadelphia built subways after

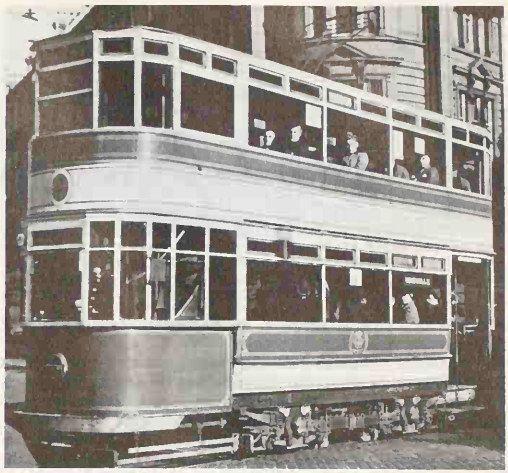
1895; while Chicago, which had developed an extensive elevated system, completed its first underground passenger route in 1943; and Toronto in 1954, the first in Canada. In the United States in 1956, subway and elevated lines carried 1,900,000,000 passengers, out of a total of 8,800,000,000 carried by all urban transit. By comparison, in 1940, 2,300,000,000 were carried out of a total of 10,500,000,000. For a number of years the total track mileage was about 1,250, with about 9,200 elevated and subway cars in use. In London an extensive underground system was developed, originally using steam, and then electricity. Paris was blanketed by the rapid transit lines of the Métro (for Métropolitaine). Berlin, Hamburg, Stockholm, Madrid, Vienna, Warsaw, Moscow and some other cities had one or more subway or elevated lines. Many cities in Europe and North America contemplated subway systems, but the heavy cost of construction was a serious deterrent; given cost levels and potential fares and traffic volume, such an enterprise cannot pay its own way completely under usual conditions, and tax subsidy is necessary.

Electric Traction for Suburban and Main-Line Railroads.—As use of electricity for local transit and interurbans developed, the main-line railroads turned their attention to the possible use of electricity. Initially it was employed to eliminate the smoke menace in long tunnels, the first installation being that of the Baltimore and Ohio railroad at Baltimore, Md., in 1895. The New York Central electrified the approaches to its Grand Central terminal in 1906. Next came substantial electrification of suburban commuter lines because of the advantages of greater flexibility, cleanliness and speed of acceleration. Major installations were those of the Illinois Central in Chicago; and the Long Island, the Reading, the Pennsylvania, the New York Central and the New Haven lines in the Philadelphia and New York metropolitan areas. The heavy investment and the limited profitability of commuter traffic put an end to further extension. In Great Britain, the area south of London was served by an intensive electrified system of the former Southern railway. In Europe, there were electrified suburban services out of Copenhagen, Paris, Brussels and other cities, and in the Republic of South Africa, out of Johannesburg and Cape Town.

Some electrification of longer distance trunk-line railroads was undertaken in the United States, either where extremely heavy traffic volume made such service advantageous, or where cheap electric power and heavy grades favoured electric operation. The principal example of the former is the Pennsylvania railroad's New York-Washington and Philadelphia-Harrisburg lines, electrification of which was not completed until the 1930s. Examples of the latter situation include the pioneer installation of the Milwaukee Road in the Rocky mountains and the Cascades. The Virginian and the Norfolk and Western railroads also had considerable electrified trackage. Further expansion of electrification was brought to an abrupt end in the 1930s by the development of the diesel locomotive. A limited amount of electrification was undertaken in Great Britain, and very extensive amounts in Europe. The Netherlands and Swiss railway systems were almost entirely electric, and substantial main-line electrification was to be found

G. KRAMBLES FROM B. J. MISEK

FIG. 7.—INTERURBAN CAR, TYPICAL OF THE HEYDAY OF THIS FORM OF INTERCITY TRANSPORT. SOME WERE REPLACED BY STEEL EQUIPMENT; OTHERS WERE USED UNTIL THE 1930S



G. KRAMBLES FROM B. J. MISEK

FIG. 8.—THE DOUBLE-DECK CAR CHARACTERISTIC OF BRITISH TRAMWAYS. THIS TYPE OF CAR WAS USED ONLY EXPERIMENTALLY IN THE U.S.

in Sweden, Norway, France and other countries. The denser traffic per mile and the shortage of coal and oil in a number of these countries favoured the use of electric power.

Voltage.—In the early days of electric traction, many experiments were made with low-voltage operation in an effort to minimize shocks. But with the perfection of the trolley system, 600 v. was found to be most satisfactory, and became more or less standard. Direct current was employed, largely because it provided superior starting ability and used lighter motors. Thus converters, and later rectifiers, were required in the substations where the alternating current of the commercial power systems was obtained. As interurban and suburban operation spread, 600 v. D.C. operation was not entirely satisfactory, because of the limited distance the power could be transmitted without excessive voltage drop. About 1900 a number of lines adopted single phase alternating current, typically at about 6,000 v., a system promoted by Westinghouse, but found this more costly in the end because of the much greater weight of the cars and more serious maintenance problems. The last interurbans built typically used 1,200 or 1,500 v. D.C. power. On suburban and main-line operations different systems were used, including 600 v. D.C. and third rail; 1,200 v. D.C. and overhead trolley; and low-voltage A.C. motors with the voltage stepped down on the cars. The final development, as used on the New Haven railroad, was the installation of rectifiers on the cars, which allowed the use of A.C. in the trolley lines, but more efficient D.C. motors. In Europe greater use was made of high-voltage alternating current at the same frequency as commercial lines (50 or 60 cycles per second). *See* also SUBWAY (UNDERGROUND RAILWAY); RAILWAY.

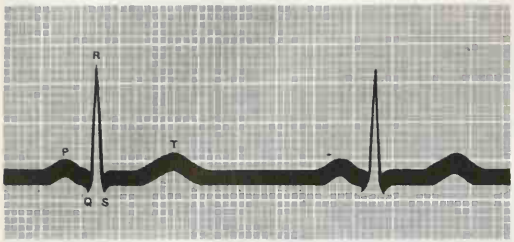
BIBLIOGRAPHY.—F. Rowsome, Jr., *Trolley Car Treasury*, tech. ed., S. D. Maguire (1956); J. A. Miller, *Fares, Please!* (1941); W. Owen, *The Metropolitan Transportation Problem* (1956); Federal Electric Railway Commission, *Proceedings* (1920); Henry W. Blake and Walter Jackson, *Electric Railway Transportation* (1924); American Transit Association, *Transit Fact Book* (yearly); past volumes of *Street Railway Journal* and its successor, *Electric Railway Journal*, and *Mass Transportation*; H. W. Linecar, *British Electric Trains*, 2nd ed. (1949); G. W. Hilton and J. F. Due, *The Electric Interurban Railways in America* (1960); Charles Klapper, *The Golden Age of Tramways* (1961).

(J. F. D.)

ELECTRIC WAVES: *see* ELECTROMAGNETIC WAVES.

ELECTRIFICATION, RURAL: *see* RURAL ELECTRIFICATION.

ELECTROCARDIOGRAPHY, the recording and analyzing of the electric currents produced by the heart when it contracts. An instrument called the electrocardiograph receives the electrical



TWO NORMAL HEART BEATS RECORDED IN A STANDARD LEAD II ELECTROCARDIOGRAM. HEAVY HORIZONTAL LINES ARE 5 MM. APART; HEAVY VERTICAL LINES ARE 0.2 SEC. APART; P WAVE RECORDS THE BEGINNING OF AURICULAR ACTIVITY; QRS-T WAVES REPRESENT VENTRICULAR CONTRACTION

impulses and records them on an electrocardiogram (EKG or ECG).

The discovery that heart muscle sets up an electric current when it contracts was made in the second half of the 19th century. Efforts to record this voltage were too crude to be clinically useful until Willem Einthoven (*q.v.*), a Dutch physiologist and physician, invented the string galvanometer in 1903. Subsequently, electronic devices, with amplification of the current, were added to the instrumentation but did not wholly replace the string galvanometer.

Electrocardiograms are made by applying electrodes to various parts of the body to lead off the tiny heart current to a suitable recording instrument. The four extremities and the chest wall became the standard sites for applying the electrodes. After the electrodes are in place, a millivolt from an extraneous source is introduced so that the instrument can be calibrated; standardizing the electrocardiograms makes it possible to compare them from person to person and from time to time in the same person.

The electrocardiogram is best interpreted when the reader (generally a physician) has clinical information about the patient. Under these conditions it can be a valuable adjunct to the practice of medicine. In general, it is useful in telling about the state of the heart muscle and in diagnosing disturbed heart rhythms. Sometimes it can only indicate nonspecific heart muscle damage and sometimes it can diagnose the cause of damage, as in coronary thrombosis.

See also CIRCULATION OF BLOOD: *The Electrocardiogram;* DIAGNOSIS: *Electrocardiograph;* HEART, COMPARATIVE PHYSIOLOGY OF: *Electrocardiograms.* (E. B. BY.)

ELECTROCHEMISTRY, that branch of chemistry which is concerned with the relation between electricity and chemical change. Under ordinary conditions, the occurrence of a chemical reaction is accompanied by the liberation or absorption of heat and not of any other form of energy, but there are many spontaneously occurring chemical reactions which, when allowed to proceed under special circumstances, liberate electrical energy, an electric current being generated. Conversely, the energy of an electric current can be utilized to bring about many chemical reactions which do not occur spontaneously. A process of the former type involves the direct conversion of the chemical energy which is liberated in a reaction to electrical energy, and an apparatus whereby such a process is brought about constitutes an electrical cell. On the other hand, in processes of the latter kind, electrical energy is directly converted into chemical energy, which is stored up in the products of the reaction: such a process is one of electrolysis; *i.e.*, an electrolytic process. By virtue of their chemical energy, the products of an electrolytic process have a tendency to react spontaneously with one another, reproducing the substances that were consumed during the electrolysis. If this (reverse) reaction is allowed to occur under proper conditions a large proportion of the electrical energy used in the electrolysis may be regenerated. This possibility is made use of in accumulators or storage cells, sets of which are known as storage batteries. Storage cells are often referred to as secondary cells to distinguish them from primary cells. A primary cell is not designed and constructed for recharging after its original supply of energy has been consumed. The charging of an accumulator is a process of electrolysis; a definite chemical change is produced by the electric current passing through it. In the discharge of the cell, the reverse chemical change occurs spontaneously; the accumulator is acting now as a cell that produces an electric current. The "storage" of electrical energy in a secondary cell thus involves its conversion into chemical energy which can be reconverted into electrical energy when desired.

In addition to these purely electrochemical processes, it is customary to include under the heading of electrochemistry those processes in which the energy of an electric current is first converted into heat, which then serves to bring about a chemical reaction which would not occur spontaneously at ordinary temperatures. These electrothermal processes represent the indirect conversion of electrical energy into chemical energy through the medium of heat. An example of an apparatus in which they can occur is the electric furnace. Still more indirect processes, however, are not regarded as electrochemical. For example in the generation of electricity on a large scale, the electrical energy is actually derived from the chemical energy of fuels such as coal, coke or oil, but the operation is not considered to be an electrochemical one, since the chemical energy is first converted into heat by combustion of the fuel, heat is then converted into mechanical energy by some form of heat engine, and finally mechanical energy is transformed into electrical energy by means of the dynamo. Although in this very indirect process only a fraction of the available chemical energy of the fuel can actually be obtained as electrical energy—mainly because of the low efficiency of the heat engine—yet no known type of primary cell can be expected to compete economically with it, since in the operation of such cells relatively expensive material, usually zinc, is consumed continuously. Attempts to devise an efficient primary cell, capable of producing electrical energy directly from the chemical energy of ordinary fuels, have frequently been made, but although many difficulties have been overcome, none of the fuel cells yet designed has been successful on a commercial scale.

Finally, the passage of electricity through gases generally causes chemical changes, and this subject forms a separate branch of electrochemistry. When the current passes in the form of the so-called "silent" discharge, the chemical effects must be attributed to the direct action of the electrical energy supplied, but such a process is not an electrolysis, in the usual sense of the term. In cases where the discharge takes the form of an electric arc, much heat is liberated at high temperatures and the chemical changes produced are the result of combined electrical and thermal conditions. Such processes are partly electrothermal in character.

General Principles.—Substances which are reasonably good conductors of electricity may be divided into two groups, the metallic or electronic conductors and the electrolytic conductors. The metals and a few substances such as graphite, manganese dioxide and lead sulfide exhibit metallic conductivity; the passage of an electric current through them produces heating and magnetic effects but no chemical changes. Electrolytic conductors or electrolytes comprise most acids, bases and salts, either in the molten condition or in solution in water or other solvents. Plates or rods composed of a suitable metallic conductor dipping into the fluid electrolyte are employed to conduct the current into and out of the liquid; *i.e.*, to act as electrodes. When a current is passed through an electrolyte between suitable electrodes, not only are heating and magnetic effects produced but also definite chemical changes occur at or in the neighbourhood of the electrodes, the process being one of electrolysis. At the negative electrode or cathode, the chemical change may be either the deposition of a metal, or the liberation of hydrogen and formation of a basic substance, or some other chemical reduction process; whereas at the positive electrode or anode it may be either the dissolution of the anode itself, the liberation of a nonmetal, the production of oxygen and an acidic substance, or some other chemical oxidation process. In some cases, these primary products of electrolysis then react with the electrolyte or with the material of which the electrodes are composed, yielding secondary products. It was shown by Michael Faraday that the weight in grams of any of the primary products of an electrolysis liberated by Q coulombs of

electricity is equal to $Qz/96{,}500$, where z is the amount of that substance known as the chemical equivalent of the substance: hence to produce a chemical equivalent of any of the products of an electrolysis requires 96,500 coulombs of electricity.

An electrolyte, prepared either by the melting of a suitable substance or by the dissolving of it in water or other liquid, owes its characteristic properties to the presence in it of electrically charged atoms or groups of atoms produced by the spontaneous splitting up or dissociation of the substance. In solutions of the so-called strong electrolytes, most of the original substance, or in some solutions perhaps all of it, has undergone this process of electrolytic dissociation into charged particles or ions (*see* SOLUTIONS). When a potential difference is established between electrodes dipping into an electrolyte, positively charged ions move toward the negative electrode and ions bearing negative charges move toward the positive electrode. The electric current is carried through the electrolyte by this migration of the ions (*see* ELECTRICITY, CONDUCTION OF: *Conduction in Liquids*). When an ion actually reaches the electrode of opposite polarity, its charge may be neutralized. The ion is thereby converted into an ordinary neutral atom or group of atoms. It is this discharge of ions which gives rise to one of the types of chemical change which occur at electrodes.

Copper sulfate, $CuSO_4$, when dissolved in water is an electrolyte, since it dissociates under these conditions into positively charged copper ions, Cu^{2+}, and negatively charged sulfate ions, SO_4^{2-}. When an electric current is passed through the solution by way of platinum electrodes the copper ions move toward the negative electrode (the cathode) and as they reach it are discharged, giving metallic copper which is deposited on the cathode. This process—occurring at only one of the two electrodes—is called a half-reaction. It may be represented by the simple chemical equation:

$$Cu^{2+} + 2\,\mathcal{E}^- \rightarrow Cu \qquad (1)$$

$\mathcal{E}^-$ represents one unit of electricity; *i.e.,* one electron when Cu^{2+} and Cu are regarded as representing an individual ion and atom respectively or one faraday of electricity when Cu^{2+} and Cu are regarded as representing one gram formula weight (63.54 g.) of the respective species. This equation does not tell quite the whole story and can be deceptive unless it is accompanied by a short explanation. Some competent investigators have presented evidence, regarded by many scholars as good, that not all of the copper sulfate in aqueous solutions is dissociated into free cupric ions and sulfate ions. The remainder is present as undissociated copper sulfate (or as "ion pairs"). The formula, $CuSO_4$, is used below to represent the undissociated electrolyte. If the undissociated copper sulfate contains a large fraction of the total copper in the solution the following equation is a more accurate representation of the cathode half-reaction than equation (1):

$$CuSO_4 + 2\,\mathcal{E}^- \rightarrow Cu + SO_4^{2-} \qquad (2)$$

When only the general principles of electrolysis are being discussed the difference between the two equations may be ignored. The truth lies between the two, but either may be used alone to represent the half-reaction if one is reminded that the Cu^{2+} may be partially associated with the SO_4^{2-} and that either the formula Cu^{2+} or the formula $CuSO_4$ is used to represent whatever equilibrium composition actually exists in the solution. A complete description of the electrolysis would require a detailed knowledge and a detailed statement of the concentrations which exist as a result of the following equilibrium:

$$CuSO_4 \rightleftarrows Cu^{2+} + SO_4^{2-} \qquad (3)$$

When a small fraction of the copper in the solution has been deposited one of the following complications may arise: (1) if the source of electrical power is controlled so that the potential drop across the cell cannot rise or can rise only a small amount the electrical current decreases nearly to zero and electrolysis slows nearly to a halt; (2) if the voltage supplied by the power source is allowed to rise indefinitely other electrode reactions begin; *e.g.,* water may be electrolyzed; *i.e.,* gaseous hydrogen may be released by electrolysis of some of the water in the solution. Much is known about such phenomena but they will not be discussed here. The simpler situation—the limited potential difference—is sufficient basis for the discussion of some interesting and important principles of electrolysis.

If the voltage drop across the cell is limited, *i.e.,* if it is allowed to increase only a little, equation (1) or (2) represents practically all that occurs at the negative electrode. When the electrolysis begins, a layer of copper appears quickly; then the current decreases rapidly as the rate of electrolysis slows almost to a halt. The explanation is simple. Let us examine the situation when a small amount of electricity has passed through the cell. Let us say that 0.001 faraday has been allowed to pass. The amount of copper deposited is 0.001 equivalent; in other words it is $(0.001)\frac{63.54}{2}$ g. Most of that copper was derived from electrolyte which was initially close to the electrode. The concentration of copper salt in that region has thereby been decreased.

It might seem, at first glance, that the copper moving into the region as a carrier of electricity would replace all of the copper consumed and that the concentration would be maintained at nearly the original level. Only a fraction of the current, however, is carried by the copper ion. That fraction is known as the transference number of the copper ion. It is a complicated function of temperature and solution composition. It is about 0.25 in this illustrative process. After 0.001 faraday has passed, only 0.00025 equivalent of copper has been brought into the region from which 0.001 equivalent has been removed by electrolysis. The electricity not carried by the copper ion, 0.00075 faraday, was carried by the sulfate ion (carrying negative charges) moving in the other direction toward the positive electrode.

When 0.03177 g. of copper, *i.e.,* 0.001-equivalent, has been deposited the solution around the electrode has suffered a net loss of 0.00075 equivalent of Cu^{2+} and a net loss of 0.00075 equivalent of SO_4^{2-}. In other words, the solution has lost 0.00075 equivalent of the electrolyte, copper sulfate. One of the results of the decreased concentration is that the specific resistance of the solution near the electrode becomes high. Another is that a higher electrolyzing voltage is needed for the deposition of Cu^{2+} ion from a dilute solution than from a more concentrated one. Either of these factors is sufficient to decrease the rate of electrolysis almost to zero if the electrolyzing voltage is prevented from rising freely.

It is quite feasible to use the electrolysis of solutions of copper salts for precise determinations of the concentration of copper in those solutions. In the analysis the copper may be removed quantitatively from the solution within an hour or less. It is also practicable in various industrial processes to deposit great quantities of copper in a short time. Both operations depend upon the stirring of the solutions being electrolyzed. The stirring moves new material into the regions about the electrodes. In a well-stirred solution the migration of the ions in the electrical field is important only for short distances since the stirring moves ions into virtual contact with the electrodes. If the solution is stirred even gently, an electrolysis which would otherwise require many years for completion may be made to occur in a few hours.

As Cu^{2+} ions move toward the cathode, there to be reduced to metallic copper, sulfate ions move to the anode and in doing so carry part of the electricity. Under ordinary conditions the sulfate ions are not oxidized. The substance oxidized is water. The electrolysis of the water produces elementary oxygen which is evolved as bubbles of gas. The valence of the hydrogen of the water is not altered; the hydrogen remains in solution, some of it as hydrogen ion and some of it as a component of bisulfate ions which are also produced. If the solution is not stirred the region around the anode therefore becomes acidic. (The acidity may be simply demonstrated in a colourless solution by the colour change of an acid-base indicator.) Either of the following equations is a good approximation of the anode reaction for the passage of two faradays through the cell:

$$SO_4^{2-} + H_2O \rightarrow \tfrac{1}{2}O_2(\text{gas}) + H^+ + HSO_4^- + 2\,\mathcal{E}^- \qquad (4)$$

$$2SO_4^{2-} + H_2O \rightarrow \tfrac{1}{2}O_2(\text{gas}) + 2HSO_4^- + 2\,\mathcal{E}^- \qquad (5)$$

When the electrolysis begins, the concentration of the sulfate

ions (or of copper sulfate) is high. Equation (5) adequately represents the changes which occur if it is explained, as above, that SO_4^{2-} stands temporarily both for sulfate ions and for undissociated copper sulfate which may dissociate to liberate the ions. When the electrolysis is nearly complete the accumulated anode changes are more nearly represented by equation (5) than by equation (4). The two are equivalent, of course, if it is understood that HSO_4^- is intended to represent the substances produced whatever they may be, in the proportions in which they are actually formed.

For some purposes a single equation may be wanted for the overall reaction which occurs when two faradays of electricity are passed through a stirred solution of copper sulfate. The simplest equation is:

$$CuSO_4 + H_2O + \text{electrical energy} \rightarrow Cu + H_2SO_4 + \tfrac{1}{2}O_2(\text{gas}) \qquad (6)$$

Under ordinary conditions no appreciable amount of undissociated H_2SO_4 molecules is formed (*see* SULFURIC ACID). The formula merely stands here for the mixture of H^+, HSO_4^- and SO_4^{2-} ions which actually exists in a sulfuric acid solution of moderate concentration. Similarly $CuSO_4$ in equation (6) stands for both the associated and dissociated copper sulfate, because we are not concerned here with the equilibrium between the free and associated ions.

A simpler case is that of an aqueous solution of copper chloride, $CuCl_2$, in which the ions are Cu^{2+} and Cl^-, when conditions are such that the Cl^- ions yield only chlorine gas when they reach the anode. The equation for the complete electrolysis is

$$CuCl_2 + \text{electrical energy} \rightarrow Cu + Cl_2. \qquad (7)$$

The quantity of electricity required to produce, for example, 1 g. of copper by the electrolysis of either of the above solutions is $96{,}500/31.77 = 3{,}037$ coulombs, since 31.77 g. of copper is the chemical equivalent in copper salts of this type. But the quantity of electrical energy needed for this purpose is given by the product of the quantity of electricity and the voltage which was applied across the electrodes during the electrolysis. If the current through a given electrolyte is to be I ampere, then the applied voltage, E_I, is given by:

$$E_I = E_0 + IR + \pi, \qquad (8)$$

where IR is the voltage used in overcoming the resistance, R ohms, of the column of electrolyte between the electrodes; and π denotes the overvoltage of the cell, a quantity which depends on the nature and concentration of the electrolyte, the nature of the electrodes and the current density; *i.e.*, the ratio of I to the surface area of the electrodes in contact with the electrolyte. If the current density is large, π may be considerable but as I is decreased π may approach zero. The minimum value of E_I which is approached as I is diminished is E_0, the decomposition voltage of the given electrolyte with the given electrodes. This represents the voltage which must be exceeded if a particular electrolysis is to occur at all. For a 0.5-molar aqueous solution of copper chloride $E_0 = 1.065$ v. The electrical energy needed to liberate 1 g. of copper from this solution under any given conditions is $\frac{96{,}500}{31.77} E_I = 3.037\ E_I$ joules. The minimum value of this quantity under any condition is $3.037 \times E_0 = 3.037 \times 1.065 = 3.234$ joules. This value represents also the decrease of chemical energy when 1 g. of copper and the equivalent amount of chlorine (1.116 g.) are converted into copper chloride in aqueous solution. In practice some more electrical energy is always needed, since E_I is necessarily somewhat greater than E_0. The additional energy is used in overcoming resistance and polarization effects and is thereby converted into heat.

In virtue of their chemical energy, the products of an electrolysis tend to react with one another if allowed to come into contact. Thus, copper and chlorine interact readily, forming copper chloride and liberating a considerable quantity of heat—the heat of reaction. In a primary cell, however, in which a copper electrode and a chlorine electrode dip into a solution of copper chloride, this same reaction can be caused to occur so as to yield electrical energy. For this purpose the copper electrode may consist of a plate of the metal, but, in order that chlorine may function as an electrode, a conductor of some material that is not attacked by chlorine must be used to act as an intermediary between the gas and the solution. A plate of platinized platinum or a rod of porous carbon saturated with the gas can act as a chlorine electrode.

When the two electrodes are connected by means of a wire outside the cell, a current passes through the circuit formed by the electrolyte, the electrodes and the external wire. Whenever the cell is allowed to deliver a current through an external circuit, chemical changes occur at the electrodes, viz., copper dissolves from one electrode forming Cu^{2+} ions and chlorine goes into solution as Cl^- ions at the other, so that, although the copper and chlorine do not actually come into contact, the total chemical reaction is

$$Cu + Cl_2 \rightarrow CuCl_2 + \text{electrical energy}, \qquad (9)$$

which is exactly the reverse of that occurring in the electrolysis of copper chloride solution.

For every gram of copper which dissolves during the action of the cell, 3.037 coulombs of electricity is obtained. The amount of electrical energy obtained, however, is $3.037 \times E'_I$, where E'_I is the actual voltage or potential difference between the electrodes of the cell, when it is furnishing a current of I amperes. For the cell $E'_I = E_0 - IR + \pi$ where these symbols have the same significance as above, π being the overvoltage of the cell under the given conditions. If I is made small by using an external circuit of high resistance, π approaches zero, E'_I approaches E_0, and the electrical energy obtainable is approximately $3.037 \times E_0 = 3.037 \times 1.065 = 3.234$ joules per gram of copper dissolved. This quantity represents the maximum amount of electrical energy obtainable from the given chemical change and is a measure of the chemical energy which is causing the reaction to proceed. It should be noticed that the heat evolved when a reaction occurs under ordinary conditions is not generally equal to the chemical energy. Hence, in general, some heat is evolved or absorbed even when a primary cell is operated in such a manner that it is furnishing the maximum amount of electrical energy corresponding to the chemical change. For example, an ordinary lead accumulator absorbs heat from its surroundings when it is being discharged; *i.e.*, when it is actually liberating electrical energy.

Let us return for a moment to the question of electrolysis. It must be pointed out that, if a copper chloride solution is electrolyzed between two copper electrodes, copper is deposited on the cathode, but under these conditions copper is also dissolved from the anode at an equal rate, so that no net chemical change occurs in the cell as a whole. Copper is merely transferred from the anode to the cathode and the decomposition voltage is zero. Processes of this kind are employed in refining metals. Using anodes of the crude metal, conditions are arranged so that impurities in these anodes do not deposit with the metal on the cathode during electrolysis. The voltage required if the solution is well stirred is merely that needed to overcome resistance and polarization effects.

Apart from electrolytic conductivity, which has been considered above, and electronic conductivity, which is ascribed to the movements of free electrons through a metallic conductor, the conductivity of gases should be mentioned. Under ordinary conditions, gases are practically insulators, the small electrical conductivity which they do possess being attributed to the accidental presence of a few gaseous ions (charged molecules). When a high potential difference is established between electrodes in a gas, these ions move toward the electrodes of opposite polarity with considerable velocities, so that often when they collide with neutral molecules of the gas, they are able to split up those molecules into ions. With a sufficiently high voltage, this ionization by collision becomes important, and, as more and more ions are formed, the conductivity of the gas increases. But ions carrying charges of opposite sign may also collide with one another and form neutral molecules again. With a given voltage the current through the gas attains a steady value when the rate of formation of ions is equal to their rate of recombination. When a current passes through oxygen gas, some oxygen molecules, O_2, may be ionized into O^+ and O^-, whereas others may yield O_2^+ and a free electron.

If now the ions O^- and O_2^+ collide, a molecule of ozone, O_3, may be formed. Chemical changes due to the passage of electricity through other gases arise in a similar way.

The heating effects, which are produced when large electric currents are driven by means of high voltages through metallic conductors or gases, are utilized in electrothermal chemical processes. Alternating currents can generally be used for this purpose.

Industrial Applications.—Electrolytic processes are employed commercially in the extraction of metals such as aluminum, magnesium, sodium, zinc and nickel from their ores; in refining metals such as gold, silver, copper, lead and tin; in electroplating with various metals and alloys; and in producing substances such as caustic soda, hypochlorites, chlorates, hydrogen and oxygen, the so-called "per-salts," etc. Electrothermal chemical processes have been applied to the smelting of ores of metals such as iron and zinc; to the production of ferroalloys; to the refining of steel; and to manufacture of substances such as calcium carbide, carborundum and phosphorus. The preparation of ozone and the production of nitric oxide from air (*see* NITROGEN, FIXATION OF) are processes depending on the passage of electricity through gases.

See BATTERY; ELECTRICITY, CONDUCTION OF; ELECTROMETALLURGY; ELECTROPLATING; ISOTOPE; OZONE; ZINC; *see* also references under "Electrochemistry" in the Index volume.

BIBLIOGRAPHY.—A. J. Allmand and H. J. T. Ellingham, *Principles of Applied Electrochemistry*, 2nd rev. ed. (1924); I. N. Pring, *The Electric Furnace* (1921); H. S. Harned and B. B. Owen, *The Physical Chemistry of Electrolytic Solutions*, 3rd ed. (1958); Malcolm Dole, *Principles of Experimental and Theoretical Electrochemistry* (1935); D. A. MacInnes, *The Principles of Electrochemistry* (1939); R. A. Robinson and R. H. Stokes, *Electrolyte Solutions* (1955); *Principles and Applications of Electrochemistry*, vol. i, *Principles* by H. J. Creighten, 4th ed. (1943), vol. ii, *Applications* by W. A. Koehler, 2nd ed. (1944); P. Delahay, *New Instrumental Methods in Electrochemistry* (1954); J. O'M. Bockris and B. E. Conway (eds.), *Modern Aspects of Electrochemistry* (1954); G. Kortüm and J. O'M. Bockris, *Textbook of Electrochemistry*, 2 vol. (1951). (H. J. T. E.; X.)

ELECTROCUTION, the popular name, invented in the United States, for the infliction of the death penalty on criminals by passing through the body of the condemned a current of electricity sufficient to cause death. The method was first adopted by the state of New York, a law making this method obligatory having been passed and approved by the governor on June 4, 1888. The first criminal to be executed by this method was William Kemmler, on Aug. 6, 1890, at Auburn prison. The validity of the New York law had previously been attacked in regard to this case, as providing "cruel and unusual punishment" in violation of the constitution; but it was sustained in the state courts and finally in the federal courts (*Re Kemmler*, 1889, 136 U.S. 436).

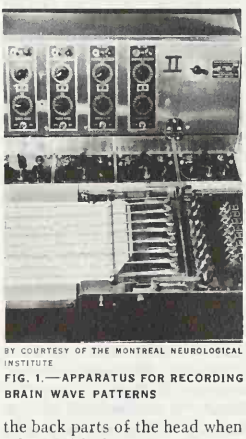
Typical execution equipment consists of a 2,300-v., single-phase, 60-cycle A.C., serviced induction voltage regulator and autotransformer with necessary panel-mounted switches and meters. This unit furnishes up to 2,000 v. current to the electric chair. The condemned person is brought in, seated in the chair and fastened securely with straps around chest, groins, arms and ankles. The moistened sponge-lined, properly molded electrodes are strapped to the head and the calf of one leg. The wires are attached to the copper terminals of the electrodes, the face mask is fastened with a strap around the headrest. A single shock of current with a starting potential of 2,000 v. is applied, and immediately reduced to approximately 500 v., then raised and lowered at 30-second intervals for a total application period of two minutes, during which period the current varies from four to eight amperes. The switch is then opened and the official physician makes an examination before pronouncing the condemned person legally dead. When properly performed the operation is believed to be painless, and instantaneous death occurs. There can be little doubt that circulation and respiration cease with the first contact, that consciousness is blotted out instantly and that the prolonged application of the current ensures permanent derangement of the vital functions beyond recovery. To give complete assurance, the law provides for an autopsy immediately after the execution.

Post-mortem examinations of electrocuted criminals revealed a number of interesting phenomena. The temperature of the body rose promptly after death to a very high point. At the site of the leg electrode a temperature of more than 128° F. was registered within 15 minutes in many cases. After removal of the brain, the temperature recorded in the spinal canal was often more than 120° F. The heart, at first flaccid when exposed soon after death, gradually contracted and assumed a tetanized condition. The blood was profoundly altered biochemically; it was of a very dark colour and rarely coagulated. *See* CAPITAL PUNISHMENT; CRIMINAL LAW. (W. L. D.)

ELECTRODE, an electrical conductor, most commonly metallic, that usually serves as either of the two terminals of an electrically conducting medium and thus conducts current into and out of the medium. The conducting medium can be an electrolytic solution (as found, for example, in a storage battery), a solid, a molten mass, an ionized gas or a vacuum. The electrode from which electrons emerge is called the cathode and is designated negative or −; the electrode which receives electrons is called the anode and is designated positive or +. In an electron tube (*q.v.*) the anode is called the plate, and conducting elements that regulate the electron flow inside the tube are also called electrodes. *See* ELECTRICITY, CONDUCTION OF; ELECTROCHEMISTRY; ELECTROMETALLURGY.

ELECTROENCEPHALOGRAPHY is the technique of recording and interpreting the electrical activity of the brain. It was introduced in 1929 by Hans Berger of Jena, Ger. Records of the electrical activity of the brain, commonly called "brain waves," are called electroencephalograms or electroencephalographs. The common abbreviation for both the technique and the records is EEG.

Instrumentation.—The living brain generates very small electrical currents during its activity, even during sleep and in deep coma. Some of these minute currents can be recorded from the brain surface by means of small wires (electrodes) attached to the scalp, where they appear as electrical potentials of about 100 microvolts (μv) in normal persons (1 μv equals $\frac{1}{1,000,000}$ of a volt). The wires are connected to a vacuum tube amplifier where they are magnified more than 1,000,000 times. They can then be made to move an electromagnetic pen that writes the EEG on continuously moving paper.



BY COURTESY OF THE MONTREAL NEUROLOGICAL INSTITUTE

FIG. 1.—APPARATUS FOR RECORDING BRAIN WAVE PATTERNS

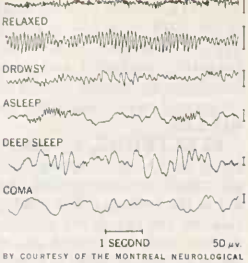
The electrical waves are not broadcast from the head as radio waves but must be conducted to the EEG machine through wires attached to the scalp or placed directly on the brain surface during neurosurgical operations for the removal of diseased parts of the brain, such as scars or tumours. Records taken directly from the brain surface are called electrocorticograms, or cortical electrograms.

Types of Pattern.—Most of the EEG record in normal adult persons is made up of rhythmic oscillating waves, regularly repeated at about 10 cycles per second (8 to 12). They are called alpha waves, or the alpha rhythm. They are best obtained from the back parts of the head when a person is resting quietly, but not asleep, with the eyes closed. The waves disappear, or are blocked, when the eyes are opened (and attention is concentrated) and when a person is excited or startled. Then the alpha rays are replaced by low-voltage, rapid, irregular waves. This is explained by the fact that thousands of brain cells must "get into step," or beat in synchrony, before their electrical waves can be detected in the EEG. When the cells are activated while the person is looking at something, or is emotionally excited, the cells become desynchronized. Hence a decrease in amplitude of the EEG may actually indicate increased brain activity, but this is not always the case.

A more rapid rhythm is characteristic of the normal EEG from the central and front parts of the head. It consists of smaller

waves, called beta waves, that are repeated at 18–25 cycles per second. These waves are more closely related to the function of the sensory-motor parts of the brain; they are blocked by voluntary movements in a manner similar to the blocking of the alpha rhythm by opening the eyes.

During sleep, even more brain cells get into step and generate even higher voltage electrical waves, but they are also slowed down to 2 or 3 waves per second, sometimes with short spindles of waves at 14 per second known as "sleep spindles." Very slow waves also appear during coma. The irregular slow waves (1 to 3 per second) that arise from the vicinity of a localized area of brain damage (sometimes due to a brain tumour) have been called "delta waves." Rhythmic slow waves at frequencies between 4 and 7 per second have been called "theta waves." Slow waves are abnormal in the adult waking subject but are normal in infants and young children. The alpha rhythm appears as a slow irregular rhythm (about 2 to 4 per second) in infants three to six months of age. EEG patterns change gradually in the growing child until the age of 8 to 12 years, when adult patterns appear. They are stabilized in the late teens and remain fairly constant until old age.



BY COURTESY OF THE MONTREAL NEUROLOGICAL INSTITUTE

FIG. 2.—TYPICAL BRAIN WAVE PATTERNS FROM A STATE OF EXCITEMENT TO ONE OF DEEP COMA (*See* TEXT)

The EEG is not the same in each individual. Every person has a different characteristic pattern that may be related to his personality type, but such relationships have not been well established. Identical twins have almost identical characteristic EEG patterns, so that a strong hereditary determinant has been well established. Patients with certain hereditary brain diseases also have similar EEG patterns.

In addition to these "resting rhythms" of the brain, there are transient electrical potentials that can be recorded from certain areas in response to the arrival of sensory nerve impulses. These are called evoked potentials. They are most useful in tracing sensory pathways and sensory receiving areas.

Application.—The EEG has proved to be a useful method for studying how the brain works, as well as for tracing connections between one part of the central nervous system and another. It is only a very gross method, however, since it records only a small sample of the synchronous activity from the brain surface. Most of the more complicated activities of the brain—those that underlie conscious sensation, thought and behaviour—in man cannot be related closely to the EEG. Consequently the EEG has not been found particularly useful in the study of many mental disorders or diseases, except when these are caused by gross damage to the brain, or when they are related to epilepsy.

The most dramatic changes in the EEG occur in patients with epileptic seizures. These are caused by excessive synchronous electrical discharge by large masses of brain cells so that high-voltage discharges appear in the EEG. Characteristic abnormal wave forms appear in different kinds of epilepsy. For example, rapid sporadic waves or "spikes" may point to a localized lesion of the brain that causes seizures of focal onset. A sudden burst of high voltage spike and wave complexes that appear simultaneously from both sides of the head in a regular rhythm at 3 per second is the electrical sign of *petit mal* epilepsy. (*See* EPILEPSY: *Causes and Diagnosis.*)

The EEG also has been found useful for locating brain tumours or local areas of injury, and in diagnosing other general diseases affecting the brain, such as encephalitis, or systemic diseases such as toxemia and diseases of the liver or kidneys that indirectly affect brain function.

See also ATTENTION: *Physiological Studies;* BRAIN: *The Brain Stem;* NERVOUS SYSTEM, SURGERY OF. (H. H. JA.)

ELECTROLYSIS. The passage of an electric current through a substance, accompanied by definite chemical changes which are independent of the heating effects of the current, is known as electrolysis (formed from Gr. *lysis,* "to loosen"), and the substance is called an electrolyte. *See* ELECTRICITY, CONDUCTION OF; ELECTROCHEMISTRY; *see* also references under "Electrolysis" in the Index volume.

ELECTROLYTE, in chemistry and physics, a substance which conducts electric current as a result of a dissociation into positive and negative ions, which migrate toward and frequently are discharged at the negative and positive electrodes, respectively. In those instances in which an ion is not discharged at a given electrode, some other substance present in the solution or forming part of the electrode is instead always oxidized at the positive electrode or reduced at the negative electrode. The most familiar electrolytes are acids, bases and salts, which ionize in solution in such solvents as water, alcohol, etc. Many salts, such as sodium chloride, behave as electrolytes when melted in the absence of any solvent; and some, such as silver iodide, are electrolytes even in the solid state.

See ELECTRICITY, CONDUCTION OF: *Conduction in Liquids;* ELECTROCHEMISTRY; *see* also references under "Electrolyte" in the Index volume.

ELECTROMAGNET, a magnet which employs electric current to generate its magnetic field. In 1820 H. C. Oersted discovered that an electric current could turn a magnetized needle. In the same year both D. F. J. Arago and Sir Humphry Davy found that an electrified wire could magnetize steel needles and attract iron filings. These basic experiments demonstrated that magnetism (*q.v.*) was related to electric currents. In one of Davy's experiments steel needles were fastened to a silver wire in different directions, some being parallel and others transverse to the wire. When the ends of the wire were connected to a battery the parallel needles attracted iron filings in much the same way as did the wire itself, and they lost this power when the connection with the battery was broken. Each of the transverse needles, however, acquired two magnetic poles and retained its magnetism after the circuit was broken. Pieces of iron wire placed transversely also became magnetized, but they became demagnetized when the connection was broken. At about the same time, Arago and A. M. Ampère (*q.v.*) magnetized a steel needle which was wrapped in paper and placed within a helical coil of wire carrying a current.

These preliminary discoveries led to the production of an electromagnet by William Sturgeon, a native of Whittingham, in Lancashire, Eng. He described (1825) an electromagnet which consisted of a bent bar of soft iron, about a foot long and half an inch in diameter, coated with varnish, on which was wound a coil of 18 turns of bare copper wire. When a current from a single cell was passed through the wire the 7-oz. bar could support a 9-lb. iron mass. When the connection was broken, the weight immediately fell. Sturgeon's experiments soon attracted the attention of other investigators, among them Joseph Henry in the U.S., who in 1831 described an important improvement which he had effected. Instead of insulating the iron bar he covered the wire with silk and wound a great number of turns around the bar in the same direction. A horseshoe-shaped electromagnet which he made in this way for Princeton university lifted 750 lb. when excited by the current from a small battery, and another one, for Yale college, supported 2,086 lb. Henry also found that electromagnets could operate at a considerable distance from the battery and developed an early version of the electromagnetic telegraph.

Magnetic fields are the basis for electric motors and generators. Electromagnets are used in electric bells, telephone receivers, radio and television speakers, tape recorders and many other devices. The electromagnet has also played a primary role in research and is an indispensable tool in the quest to understand the universe.

Electromagnetic Design.—An electromagnet, most typically, is designed to produce a field of a given volume, shape and intensity. The first design step is to calculate an electrical current density configuration to give the desired magnetomotive force and magnetic field. Whenever possible a magnetic core structure,

usually iron, is used. A magnetic core acts as a channel or conductor for the magnetomotive force and usually serves to apply it most efficiently and effectively to the working volume. On the other hand, in cases where the field is too strong (above 30,000 gausses), or when the current-carrying conductors must be placed as close as possible to the work volume, a core does not operate well.

Once the desired current density configuration is established, the proper number of wires or turns is determined by the input voltage, the current required and the resistance of the wire.

Because heat generated by coil resistance reduces the intensity of the generated field and may even destroy the magnet, provision should be made for cooling. A common design employs water circulating through pipes embedded in the winding; or, hollow wires that conduct both electricity and water may be used. Cooling with liquid helium, E. Rosner *et al.* (1965) reported a 132,000-gauss field powered only by six automobile batteries.

Francis Bitter (1936) achieved a field of 100,000 gausses by employing 1,700,000 w. of power and an elaborate cooling system. To eliminate the cooling problem, P. Kapitza (1924) developed methods of pulsing electromagnets for very short times. In a coil 1 mm. in diameter he produced a field of 500,000 gausses for .003 sec., and in a 2-cc. coil he obtained 320,000 gausses for .01 sec. In 1956 H. P. Furth, M. A. Levine and R. W. Waniek showed that there was a limit to this procedure that never permitted more energy density in the magnetic field than could be absorbed in boiling away the metal in the coil. They produced fields of 1,600,000 gausses by allowing a part of the coil to vaporize. However, the mechanical stress in these coils had become a more severe problem than the heating. The magnetic pressure in this coil was almost 7,500 tons per square inch.

Electromagnetic Mechanisms.—In a large number of devices there is a piece of iron forming part of the magnetic circuit of an electromagnet which moves when current flows in the exciting coils. The movable iron usually forms the armature of the electromagnet and is attracted when the current flows in either direction. In many of these devices the electric circuit includes a contact breaker so arranged that the movement of the armature toward the electromagnet interrupts the circuit. The core then becomes demagnetized and the circuit is re-established as the armature is drawn back by a spring. In this way vibratory movement of the armature is maintained as the current becomes intermittent. This combination of electromagnet and contact breaker is used in such devices as electric bells, electromagnetic hammers and electric automobile horns. In all such mechanisms the movement produced by the magnetic action is such as to diminish the reluctance of the magnetic circuit. The magnitude of the attraction per unit area of the surface of the armature, where the normal flux density is B, is represented approximately by $B^2/8\pi$.

Another type of mechanism, used especially in the loud-speaker and oscillograph, is one in which the received and amplified alternating current, instead of flowing through coils fixed on the cores of a magnet, flows through a movable coil in the annular gap of an electromagnet of suitable form. A description of this moving coil system will be found in the article LOUD-SPEAKER.

The magnetic recorder uses tiny electromagnets in its erasing, recording and reproducing heads. In this method of recording, sound vibrations produce varying degrees of magnetization along the length of a permanently magnetizable record medium (*see* TAPE RECORDING, MAGNETIC).

The Electromagnet in Research.—In addition to its engineering applications the electromagnet has played a primary role in many scientific investigations. The numerous uses of electromagnets as research tools basically involve the deflection of moving charges, the deflection of moving magnetic dipoles in a nonuniform field and the alignment of magnetic dipoles.

A particle with a charge of e coulombs moving at a velocity of v cm. per second across a magnetic field of B gausses will undergo a force $f = evB\sin\theta$ dynes where θ is the angle between the magnetic field and the direction of motion. This force will result in a radius of curvature r for the particle and is given (in centimetres) by $r = mv/B$ where m is the mass (in grams) of the particle. The effect of the magnetic field in bending the paths of moving charged particles was used by Sir Joseph Thomson in determining the ratio of charge to mass for an electron. Later investigators have been able to use magnetic fields to study the characteristics of other particles.

Most atoms and nuclei, because of their circulating charges, act like electromagnets or, more technically, as magnetic dipoles. When placed in a magnetic field, they tend to orient themselves in the direction of the field. Perhaps the earliest application of this effect was by M. Faraday who discovered the rotation of the plane of polarization of light in a material which had been placed in a magnetic field. Another example of the alignment of the dipoles in a material is the rotation of the plane of polarization of light reflected from polished pole pieces of a magnet.

One of the most interesting applications of this effect is in its use in the production of temperatures in the neighbourhood of absolute zero. This is accomplished by the adiabatic demagnetization of certain paramagnetic salts. The method was suggested independently by P. Debye (1926) and W. F. Giauque (1927). A sample to be cooled is embedded in a paramagnetic salt (a salt whose magnetic dipole moment is very large), and the entire assembly is cooled to as low a temperature as possible in a bath of liquid helium. A strong magnetic field is then applied which serves to orient all the dipoles in the same direction, releasing heat from the salt into the helium. The magnetic field is then turned off and the dipoles again become random. Because of the earlier release of heat, the temperature of the paramagnetic salt drops. Temperatures as low as 10^{-3}° K. have been reached in this way.

One of the most striking effects of the magnetic field on the dipole is the Zeeman effect, in which the spectral lines of a source of light shift and split into many components in a magnetic field. This effect has been used to enumerate the electron orbits in atoms. It is a primary tool in the investigation and understanding of atomic structure. Indeed, the Zeeman effect provided one of the earliest tests of the quantum theory. The tendency of a magnetic field to make charged particles move in circles has an important application in the containment, or "bottling," of hot gases. When a gas is heated to a temperature of over 50,000° C., an appreciable portion of the gas is ionized, releasing electrons. As the temperature is raised, the number of ionized particles increases, and eventually the gas becomes completely dissociated into a mixture of ions and electrons. These charged particles in a strong magnetic field will describe helices along the magnetic field lines; the motion across the magnetic field lines is reduced to a nominally low level, permitting containment of the hot gas. One of the most important applications of this technique is to control thermonuclear reactions.

The ability of magnetic fields to move charged particles in circular orbits permitted the construction of such high energy accelerators as the cyclotron. It is possible to apply the same accelerating magnetic field to a given particle many times in succession, and extremely high energies thus can be imparted to the particles without the necessity for generating and maintaining a high accelerating voltage.

A large electromagnet is also used in the betatron. This machine is used to accelerate electrons to velocities approaching that of light, whereas the cyclotron accelerates the much heavier nuclear particles at much smaller velocities.

Since the maximum energy imparted to the particles in an accelerator depends on the size of the machine, enormous electromagnets are used in some accelerators. For example the electromagnet in the 184-in. cyclotron at Berkeley, Calif., weighed more than 4,000 tons, and there were 300 tons of copper in the exciting coils. The yoke of the magnet was 59 ft. long, 30 ft. high and 184 in. wide.

Later developments in circular particle accelerators resulted in shaping and programing the magnetic field to keep the particle orbits in as thin a tube as possible. This greatly reduced the amount of magnetic core material required for such machines but, on the other hand, it made the design and material requirements far more difficult. Machines that produce particles with more than 1,000,000,000 ev of energy may be as much as 1,000 ft. in diameter

or about a mile in circumference, with tolerances in the placement of the magnetic core material of less than .001 in. These accelerators depend on a strong focusing principle to reduce the size of the orbit tube so that only a ring of electromagnets with a magnetic volume about 1×4 in. in cross section is used.

See also references under "Electromagnet" in the Index volume.

(E. T. J.; M. A. Le.)

ELECTROMAGNETIC WAVES are waves of "electric" and "magnetic" force and for this reason have often been called electric waves. In this article, both electric waves and electromagnetic waves are used interchangeably. After an introductory definition of terms and the relation of electromagnetic waves to all waves, this article is divided into three main parts: early history; Maxwell's field theory; and generation and propagation of electric waves.

Among the various material particles in the universe, some are charged positively and some negatively. The particles of similar signs repel each other and those of opposite signs attract. Stationary electric particles are said to be surrounded by an electric field of force and the moving particles by electric and magnetic fields. This is an accepted way of saying that if an electric particle is placed at any point of space surrounding one or more electric particles, it will be acted upon by a force by virtue of its electrification. That part of the total force which depends only on the charge of the particle and not on its velocity is called the electric force. The remaining part, depending on the velocity of the particle as well as on its charge, is called the magnetic force. The ratio of the electric force to the charge of the particle is called the electric intensity or the electric field strength at the point occupied by the particle. The ratio of the magnetic force to the product of the velocity of the particle and the charge is taken as the measure of magnetic induction at the point occupied by the particle. To detect the existence of electric and magnetic force and, hence, of electric and magnetic waves, a supply of electric particles is needed as well as a means of detecting their motion in response to electric and magnetic forces. Metallic wires are ideally suited for this purpose since they contain numerous negatively charged particles, the electrons, which are easily detachable from the atoms and easily movable. When used for detection of electric waves the wires, and more elaborate structures, are called antennas. An antenna is connected to some device, such as a radio set or a television set, which translates the variations in the electronic current in the antenna into sound or the visual picture.

When an electric charge is suddenly displaced, a ripple of electric and magnetic force is generated in much the same manner in which a water ripple is produced by dropping a pebble, or by sharp movement of a stick. The water ripple or wave is an up and down movement of water which is propagated on the surface in all directions with a certain velocity; it takes time for the wave to reach a point at some distance from origin of the wave and after a while the wave is gone and the water is quiescent again. The water moves up and down, but the disturbance travels along the surface. The movements are the essential features common to all waves. Electromagnetic waves do not require a material medium for their existence and propagation for they are merely waves of force. Lightning is an example of a sudden movement of electric particles; it starts an electromagnetic wave which, on reaching a radio receiver, is translated into a loud noise. (S. A. S.)

Starting with the initial development of a system for wireless telegraphy by Sir Oliver Lodge in England in 1894, the use of electric waves as a medium for the communication of intelligence through the atmosphere has expanded at an ever-increasing rate. The middle of the 20th century found them in use in wireless telegraphy and telephony, in radio broadcast and telecast (television) and in radar systems which have been developed for both military and civilian purposes. In the realm of power transmission and distribution, the application of the principles of electric wave propagation to the study of surges set up in systems by lightning and switching resulted in great reductions in the numbers of interruptions to service and hazards to life and equipment. It has been shown that the operation of all electrical devices involving energy changes can be explained in terms of the transfer of electrical energy in the form of electric waves. Familiar examples of this are the transmission of both industrial electric energy and telephone messages over long lines, the transfer of energy from the primary to the secondary of a transformer and from the stator to the rotor of an induction motor. Furthermore, the entire gamut of radiations, starting with the longest-known radio waves and progressing through the radio spectrum to microwaves, on through radiant heat and infrared light, visible and ultraviolet light, the broad X-ray spectrum and, finally, cosmic rays, are known to be electric waves of increasing frequency and decreasing wave length.

EARLY HISTORY

Introduction.—The general idea that electrical effects are propagated in the form of waves was slow in developing. Starting with the observation of Thales of Miletus in 600 B.C. that a piece of amber could be made to acquire the property of attracting small particles by rubbing it with a piece of cloth, and of Lucretius of Magnesia (95–52 B.C.) of the power of lodestone to attract iron, it was not until the 1830s that Karl Friedrich Gauss of Göttingen made the first attempt to deduce the fundamental law of electromagnetic action in terms of an electric field propagated at a finite velocity. In that long interval, extensive experimental studies carried out in all the countries of the civilized world had placed both electrostatics and magnetostatics upon a firm scientific basis and had also brought them together in electromagnetism. In order to understand the significance of the various discoveries and deductions which finally resulted in the modern conceptions regarding electric waves, typical views of some of the better-known scientists in the development period will be reviewed. (*See* also Electricity and the references listed in the bibliography at the end of this article for additional details.)

Action at a Distance.—It is not known that either Thales or Lucretius made any attempt to explain the electric and magnetic effects noted by them, and so it can be concluded that they were accepted without question as examples of action at a distance. The principle that a force could not be communicated between bodies other than by some such tangible means as pressure or impact had been maintained even by many of the ancient Greek philosophers, including Aristotle (384–322 B.C.). That others were to hold the same view is evidenced by the famous book *De magnete* which was published in 1600 by William Gilbert, physician to Queen Elizabeth I. The opinion is expressed there that electrical phenomena are due to something material which is liberated from bodies when electrified by friction, without being accompanied by any change in form or loss in weight. René Descartes (1596–1650), famous French mathematician and philosopher, likewise did not accept the possibility of action at a distance and believed that all space was occupied by an ether consisting of moving particles which did not affect the senses but yet could transmit a force by contact and collision. Descartes also believed, as did Leonhard Euler (1707–83), a Swiss mathematician, that magnetism consisted of vortexes in an omnipresent ether, an idea to be expressed again by James Clerk Maxwell in 1861.

Velocity of Electricity.—Although it was originally believed that electric and magnetic effects materialized instantly, this view, like that of action at a distance, was to be challenged. Sir William Watson, who was a believer in the ether concept (*see* Ether [in Physics]), attempted to measure the velocity of an electric current in 1747–48. Although a circuit four miles long was used for this purpose, his instruments showed no delay. In 1834 the famous telegraph engineer Sir Charles Wheatstone used a revolving mirror and obtained a value equal to three-halves that of light. Since the measurements were made on a large coil of wire, it is certain in view of later knowledge that the true velocity along the conductor was not obtained. More reasonable results were obtained in 1850 by A. H. L. Fizeau and E. Gounelle, who made measurements on telegraph lines in France, when they obtained values of one-third and two-thirds that of light for iron and copper wires, respectively. In each of the above cases, the investigator was actually measuring the velocity of the propagation of the electric current rather than the velocity of electrical charges in a con-

ductor, since it is known that the latter has a maximum possible value in the order of millimetres per second. Many determinations were made of the velocity of electric waves, including visible light and other radiations, by a wide variety of methods during the 19th and 20th centuries, one of the most recent values being $c = 2.99773 \times 10^8$ m. per second, as given by N. E. Dorsey of the U.S. bureau of standards in 1944.

Origin of Electric Wave Concept.—In 1752 Euler expressed further opinions which might properly be considered as the forerunners of a long series of developments which finally resulted in a formulation of modern electric wave theory. He envisioned all electrical, light and gravitational phenomena as derangements in the ether which were propagated at the same velocity. Michael Faraday of the Royal institution in London also did not believe in either action at a distance or instantaneous propagation; he considered all space to be in a state of stress when electrified or magnetized, and that the fields of stress could be portrayed by lines of force, an idea that had been expressed previously in *Epistola Petri Peregrini de Maricourt de Magneta* in 1269 and by Niccola Cabeo in *Philosophia Magnetica* in 1629. Like Euler, Faraday believed that a single ether should account for both electrical and optical effects, rather than a separate one for each, as had been considered previously by many other scientists. According to Faraday, electric lines of force originated at positive charges and terminated on negative charges of electricity, magnetic lines of force formed closed curves, and all lines of force were both in a state of tension and mutually repellent. He further postulated that light consisted of electric and magnetic lines of force in motion, but later suggested that both light and radiant heat might well consist of transverse vibrations of the stresses portrayed by lines of force. Faraday did not succeed in measuring the velocity of propagation of electric and magnetic effects, but he did discover a possible connection between light and magnetism by observing in 1845 that when a strong magnetic field was set up parallel to a beam of polarized light passing through glass it would produce a rotation of the plane of polarization, a phenomenon since known as the Faraday effect. A similar result was noted by E. F. Wartmann in 1846 in connection with the passage of radiant heat through rock salt. In 1848 Gustav Robert Kirchhoff in Germany noted as a coincidence that the ratio between the electromagnetic and electrostatic units of charge was equal to the velocity of light; in 1858 Georg F. B. Riemann presented a paper at the Göttingen academy in which he assumed a finite velocity of propagation of electric effects and deduced that it must be equal to the same ratio of units.

James Clerk Maxwell, director of the Cavendish laboratory at Cambridge university, was inspired both by Faraday's portrayal of electric and magnetic forces by means of lines of force and by Lord Kelvin's (William Thomson) analogy between the flow of heat and the conduction of current over submarine telegraph cables, the Dover-Calais cable having been laid in 1851. Thus he too visualized all electric and magnetic phenomena as states of stress in an ether which had mechanical properties and believed that they were propagated in the form of waves of stress. He started with the fundamental laws of electricity and magnetism which had been definitely established as such and applied the mathematical procedures which had been developed by J. L. Lagrange (1736–1813) in his *Mécanique analytique* to the immediate vicinity of a point where the fields of electric and magnetic force were assumed to be present; thus he developed the famous Maxwell field equations. These constituted the first general mathematical formulation of the problem which tied the various effects together in their correct mutual relationships. Maxwell's first paper on this subject, "On Faraday's Lines of Force," appeared in 1855. In this Faraday's conceptions of the nature of fields were put into a form suitable for mathematical treatment by introducing the idea of a tube of force. This was assumed to be a tubular surface whose elements were the lines of force passing through a small closed curve in the field. Thus, these tubes would originate on the surface of positively charged bodies and terminate on corresponding negative ones, the magnitudes of the two charges included within a single tube being the same. In a series of three articles in 1861 he also worked out a special theory for magnetic fields, by looking upon a tube of magnetic force as a vortex in the ether whose axis of rotation coincided with the direction of the force. Although Maxwell made little reference to this vortex theory in later papers, it appears to have served as a basis for the final form of his theory as stated in 1864 in the paper "On a Dynamical Theory of the Electromagnetic Field." One of his greatest contributions to electromagnetic theory appears in this paper in which he introduced the idea of a displacement current which could set up a magnetic field in addition to that set up by charges moving through wires or electrolytes. Such a current can be of appreciable magnitude in both free space and insulating materials. The inclusion of this displacement current removed one of the greatest difficulties which had been encountered in the development of the theory.

Recognition at this point should be given also to the work of Ludwig V. Lorenz (1829–91) in Copenhagen, who extended the well-known potential theory of F. E. Neumann to include the delay which would result from the propagation of electrical effects at a finite velocity. Independently from Maxwell he visualized visible light as an electromagnetic phenomenon rather than a stream of particles as had been considered by Sir Isaac Newton (1642–1727) and others. In 1867 Lorenz published a theory of electromagnetism which involved the use of a displacement current and further associated the effects directly with distributions of charges and current elements producing them. (C. S. Rs.)

MAXWELL'S FIELD THEORY

Basic Field Concepts.—Maxwell's field theory is based on the following concepts. The first one, the electric intensity $\mathbf{E}$, was defined earlier in this article as the force per unit charge held fixed at a given point. It is characterized by a certain magnitude and direction and hence it is a vector. Some material media, metals in particular, contain free electrons. Under the influence of $\mathbf{E}$ these electrons move and electric current is thus produced. The amount of electric charge crossing a given area S per unit time is the measure of the current $\mathbf{I}$ through S. If the area S is plane and perpendicular to $\mathbf{E}$, the ratio $\mathbf{I}/S$ is called the average current density over the area S. The limit of the average current density as S approaches zero is the electric current density $\mathbf{J}$ at the point approached by S. This current density is proportional to the electric intensity, $\mathbf{J} = g\mathbf{E}$, where g is called the conductivity of the medium. The conductivities of metals are very large.

In some other media, called dielectric media, most electrons are bound to the atoms and molecules. Under the influence of $\mathbf{E}$ the bound electrons will be displaced with reference to the positive particles. Furthermore, if a thin metal plate is introduced in the dielectric, or empty space for that matter, then under the influence of $\mathbf{E}$ some electrons in the plate will be displaced from one face to the other. For some particular orientation of the plate the displacement of electrons will be maximum. The ratio of the displaced charge to the area of the plate is called the displacement density $\mathbf{D}$ at the particular place in the field. In noncrystalline media the displacement density is proportional to the electric intensity, $\mathbf{D} = \epsilon\mathbf{E}$, and the coefficient of proportionality is called the dielectric constant of the medium. If $\mathbf{D}$ is varying with time, the time rate of change of $\mathbf{D}$, that is $\partial\mathbf{D}/\partial t$, is the density of the famous Maxwellian displacement current. Maxwell's hypothesis was that displacement current produces a magnetic field just as conduction current, considered in the preceding paragraph, does. It was already stated in the introduction that the magnetic induction $\mathbf{B}$ is defined as the force on a moving charged particle divided by the product of its speed and charge.

Maxwell's Equations in Homogeneous Media.—Consider a surface such as a face of a circular disk, or a rectangular plate, or a spherical shell. Suppose that this surface is divided into infinitesimal elements of area dS. If $\mathbf{J}_n$ is the component of electric current density $\mathbf{J}$ perpendicular to dS, the current through the elementary area is $\mathbf{J}_n dS$, and the total current through the area S is the sum of the currents through all elements of S, that is, $\int\int \mathbf{J}_n dS$. Similarly the total displacement current is $\frac{\partial}{\partial t}\int\int \mathbf{D}_n dS$.

Let $\mathbf{B}_s$ be the component of the magnetic induction $\mathbf{B}$ tangential to the edge of the surface S. If ds is the length of a typical element of the edge, then the sum of elementary products $\mathbf{B}_s ds$ around the edge, that is, the integral $\int \mathbf{B}_s ds$, is found to be proportional to the total electric current through S. Hence,

$$\int \mathbf{B}_s ds = \mu\Big[\int\int \mathbf{J}_n dS + \frac{\partial}{\partial t}\int\int \mathbf{D}_n dS\Big]$$

where the coefficient of proportionality μ is called the permeability of the medium. This is the Ampère-Maxwell field equation. In this equation it is understood that if an observer sums the elementary products $\mathbf{B}_s ds$ in the clockwise direction around the edge of the surface S, the electric current through S is calculated in that direction in which a right-handed corkscrew would advance if its handle were turned in the clockwise direction. If the magnetic field strength $\mathbf{H}$ is defined as $\mathbf{B}/\mu$, the Ampère-Maxwell equation becomes

$$\int \mathbf{H}_s ds = \int\int \mathbf{J}_n dS + \frac{\partial}{\partial t}\int\int \mathbf{D}_n dS$$

Similarly it is found that

$$\int \mathbf{E}_s ds = -\frac{\partial}{\partial t}\int\int \mathbf{B}_n dS$$

This is the Faraday-Maxwell field equation. The negative sign indicates that the flux of $\mathbf{B}$ through S is calculated in the direction opposite to that of current in the Ampère-Maxwell equation.

The above two equations are the famous Maxwell's equations which govern electromagnetic field phenomena. From them it can be shown that electric disturbances are propagated as waves and that in empty space the velocity of these waves is approximately equal to 3×10^8 m. per second. Waves are characterized by their frequency and wave length.

The frequency of a wave is the number of complete alternations per second in the condition of the medium. In the electromagnetic case, the frequency is the number of times per second the following cycle is repeated at any given point: the electric force increases from zero to a maximum, decreases back to zero, reverses its direction and again increases to a maximum, and finally decreases to zero once more. The wave length is the distance the wave travels in one cycle; it is also the distance between the successive crests or troughs in the wave. The velocity of the wave is equal to the frequency times the wave length. Since in air the velocity is constant, low-frequency waves are long waves and high-frequency waves are short waves. Electromagnetic waves used for national broadcasting are medium-frequency waves (medium by the electric standards); these frequencies range from about 500,000 to 1,500,000 cycles per second, and the wave lengths range correspondingly from 600 m. down to 200 m. Short waves are used for international broadcasting and for various communications purposes. The wave length of electromagnetic waves is large even for frequencies which by normal standards would be regarded as very high. This is because the velocity of electromagnetic waves, 300,000 km. per second, is exceedingly high. The velocity of sound waves, for instance, is only about 330 m. per second; the wave length for the lowest audible frequency, about 16 cycles per second, is about 21 m.; while for the highest audible frequency, about 15,000 cycles per second, the wave length is about 2½ cm.

The waves generated by an electric charge oscillating in space with a small amplitude are spherical in the sense that the ripples in electric and magnetic field strength are spherical. However, at great distances from the centre of the wave, the intensity of the wave is zero on the axis of oscillation and maximum in the direction perpendicular to the axis. The waves from two or more centres of disturbance may reinforce each other in some directions and cancel in others. It is easy to see how this may happen. Consider two sources producing waves of the same wave length and in phase; that is, at their origin the crests of the waves occur at the same time. If a point P is equidistant from both sources, the crests arrive simultaneously and reinforce each other. Similarly the troughs arrive simultaneously, and become deeper. The same situation occurs if the distances to point P are unequal but differ by one or more full wave lengths. If, however, the distances differ by half a wave length, or by an odd number of half-wave lengths, then the crests of one wave will coincide with the troughs of the other and the intensity of the resultant wave is decreased. When the two waves are of equal intensity, they will cancel each other completely. Intermediate situations arise in those directions in which the distances traveled by the two waves differ by some other fraction of a wave length, the waves tending either to reinforce or to cancel each other. This phenomenon is called the interference of waves. It is utilized for focusing electromagnetic waves and producing highly concentrated beams, as in radio beacons. The antenna from which the waves are sent out is made large compared with the wave length and is so designed that in one particular direction the waves going from all parts (or almost all) of the antenna reinforce each other. In other directions there will be automatic cancellation. The necessity of making the antennas large compared with wave length puts a premium on short waves.

Like all waves, electromagnetic waves are reflected from obstacles. If an obstacle is small compared with the wave length, then, under the influence of the incident wave, it becomes the source of another wave which spreads from it almost equally in all directions. When the obstacle is large, then each point of the obstacle acts as the source of a new wave and the total wave is the result of interference of all these elementary waves. Thus the waves reflected by large obstacles become highly directed. This phenomenon of reflection of waves is utilized in detection of objects by radar (*q.v.*) and by sonar.

Electromagnetic waves are used in radar, and sound waves in sonar. In each case a strong wave is sent toward the object, and subsequently the reflected wave is detected. The lapse of time, between the direct wave and the reflected, gives the distance to the object.

Reflection is also useful in long-distance radio communication. All objects which are large compared with wave length cast shadows; that is, the waves do not bend easily around corners. The earth is an extremely large object, and it casts a very deep shadow. Thus radio communication to points far below the horizon would be impossible, were it not for the existence of an ionized reflecting layer about 50 km. above the ground. The radio beam directed at an angle to this layer is reflected from it as shown in fig. 1 and then reflected from the earth. Unfortunately, there are several reflecting layers at different heights.

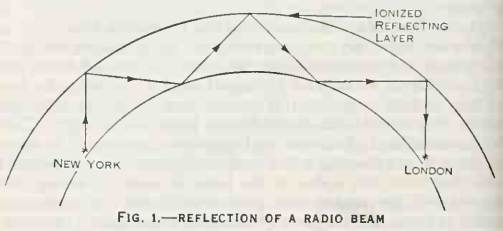


FIG. 1.—REFLECTION OF A RADIO BEAM

Sometimes the waves travel by different paths and arrive at the receiving station at slightly different times. Then the signals carried by the waves become somewhat mixed and impair the quality of long-distance broadcasts. Variations in the heights of these ionized layers with weather conditions add to the troubles of the broadcasters.

GENERATION AND PROPAGATION OF ELECTRIC WAVES

Early History.—Maxwell's electromagnetic theory was published in 1864, but electric waves were not identified until 1879, when David E. Hughes received a signal ½ mi. away from a spark transmitter. Many scientists were of the opinion, however, that this was a case of ordinary induction. Heinrich Rudolf Hertz (1857–94) of Germany is given credit for the first direct experimental verification of Maxwell's theory. In 1886, at the Technical

college at Karlsruhe, in addition to generating waves by means of the oscillatory discharge of a condenser through a loop provided with a spark gap and then detecting them with a similar type of circuit, he carried out a long series of experiments dealing with the reflection, refraction, polarization, interference and velocity of electric waves. In his later experiments, an induction-coil type of transmitter was used which gave intense waves at about 100 mc. (3 m.). Horns and reflectors were also introduced as aids in focusing and receiving the waves, and prisms and lenses made of paraffin, pitch and other dielectric materials were used for the purpose of verifying the optical properties of the waves predetermined by Maxwell.

Additional studies of the properties of Hertzian waves, as they were commonly called for several decades, were carried out by G. F. FitzGerald (1851–1901) and F. T. Trouton (1868–1922); Sir J. Ambrose Fleming (1849–1945) of the University of London; Sir Joseph John Thomson (1856–1940), director of the Cavendish laboratory; and others. A typical arrangement of equipment used in their experiments is shown in fig. 2. This apparatus, aside from demonstrating the optical properties of Hertzian waves, is of special interest since it shows the use of horns, one to focus the waves from the transmitter into a beam, the other to increase the sensitivity of the receiver.

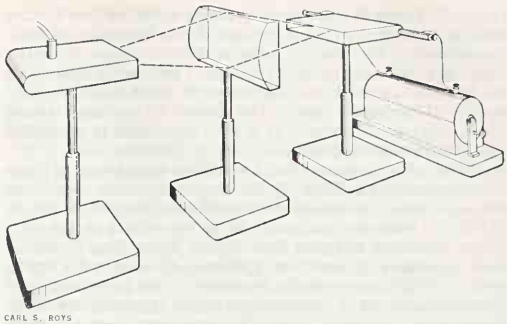
The work of Hertz was followed by some demonstrations of high-voltage, high-frequency discharges by Nikola Tesla in the United States in 1892, utilizing several new types of oscillation transformers which were subsequently used for wireless work. One of these, since known as a Tesla coil, is still encountered whenever a spectacular display of high-voltage, high-frequency discharges is desired.

The work of D. E. Hughes, Hertz and Tesla inspired Sir William Crookes, publisher and independent experimenter, to write a paper in 1892 for the *Fortnightly Review*, in which he proposed the use of electric waves as a means of transmitting telegraph signals through space. Some of the suggested needs for ensuring the success of such a project were: simple methods for generating signals of any desired wave length; more sensitive receivers which would respond to certain frequencies and not to others; and a means for directing the waves from the transmitter to the receiver.

Sir Oliver Lodge, English physicist, who was inspired in turn by Crookes's paper, is to be given the credit for developing the first system for wireless communication when, in 1894, he demonstrated successful reception over a distance of 100 yd. with the aid of a tuned circuit and the coherer developed by Edouard Branly of Paris in the receiver. Subsequent developments permitted him to extend this range to $\frac{1}{2}$ mi.

During the years 1894–96, Guglielmo Marconi, Italian electrical engineer and inventor, attained an operating range of several miles with improved equipment. This consisted of a Hertz oscillator as a transmitter and a tunable receiver provided with a sensitive detector. He stated that transmission over a distance of 100 mi. should be possible. This was challenged by the scientists since they believed that the waves traveled in straight lines and would go by the horizon and out into space at even much shorter distances from the transmitter. As the result of extensive experimentation, Marconi discovered, by grounding one of the terminals of the transmitter and connecting the other to a long wire which was supported a considerable distance above and parallel to the ground, that the waves appeared to act as though guided around the surface of the earth, and he was thus able to carry on successful transmission over almost unbelievable distances. It is interesting to note that a high, vertical antenna and improved coherer receiver was used by A. S. Popov of Kronstadt in 1895–96 for studying atmospheric electricity and detecting the approach of distant storms. Popov also suggested the use of his apparatus for "the transmission of signals to a distance by the aid of quick electric vibrations as soon as a means for producing such vibrations possessing sufficient energy has been found."

The work of Marconi in this period reached a climax in 1901 with the reception of the letter "S" in Morse code at St. John's, Nfd. This was sent out from a ten-kilowatt station located at Poldhu in Cornwall, Eng. Marconi continued his experiments



CARL S. ROYS

FIG. 2.—FOCUSING OF ELECTRIC WAVES WITH PARAFFIN LENS (c. 1890)

with a receiver mounted on the steamship "Philadelphia" and succeeded in 1902 in receiving the Poldhu station at a distance of 2,000 mi. during the dark hours, but could achieve only one-third that distance during daylight. Further experiments showed that such a great difference between the practical limits of night and day transmission decreased with increased wave length, which resulted in a long-wave-length era in the history of electric waves. During this period, wave lengths as great as 10,000 m. (30 kc.) were of common occurrence for long-range transmission. Continuous waves in this range were often obtained from high-frequency generators which had been developed by Tesla in the United States in 1890; by Sir David Salomons and L. Pyke in England in 1891; by W. Duddell in England in 1905; and by R. A. Fessenden in 1907 and E. F. W. Alexanderson in 1908 in the United States. These were of the inductor type, often driven by steam turbines, and could be built to deliver several kilowatts at frequencies up to 100,000 cycles per second. In 1907 Rudolf Goldschmidt in Germany developed an induction type of alternator whose function was to multiply the frequency produced by another alternator. The construction of the Goldschmidt alternator was similar to that of a single-phase, wound-rotor induction motor, the input winding being excited by a primary alternator. When driven at synchronous speed, frequency multiplication was accomplished by the effects of the rotating components of the pulsating fields set up by both stator and rotor windings, which were tuned to resonate at the harmonic frequencies generated during the multiplication process. Although high-frequency alternators have continued to have practical limitations upon output and frequency, they are of more than historic interest because, in addition to being in use for long-wave-length radio, they are sometimes constructed for high-frequency, inductive heating installations.

In 1903 Valdemar Poulsen, a Danish electrical engineer, obtained a patent in England covering an adaptation of Duddell's singing arc for radio purposes. In this the arc was formed between a carbon anode and copper cathode in an atmosphere of a hydrocarbon gas and a transverse magnetic field. Principally as a result of the further efforts of Poulsen and P. O. Pederson in Denmark, H. T. Simon in Germany and Fleming and W. L. Upson in England, long-wave stations with an output as great as 1,000 kw. were made possible by 1920. One typical installation was at the Annapolis Naval station, where there were two 350-kw. Poulsen arc generators; another was the 1,000-kw. Lafayette station in France. These operated in the range 20,000 to 25,000 m. (12 to 15 kc.) and permitted a telegraph speed as great as 50 words per minute. Aside from the Poulsen arc generator, many smaller types of arc and spark gap equipment were developed for long-wave use, and after 1930 many arc-type oscillators were built for the heating and melting of metals by induction.

The shortest radio waves reported by the mid-1920s were produced by means of an arc-type generator. In 1923 A. Glagoleva-Arkadieva of the University of Moscow succeeded in generating waves over the band 0.0729 to 50 mm. (410,000 to 600 mc.) the shorter wave lengths lying within the infrared region.

As a result of the discovery of the three-element vacuum tube

by Lee De Forest in 1906, and of feedback action and the vacuum tube oscillator by both De Forest and E. H. Armstrong in 1912, transatlantic and transpacific communication by means of electric waves had been made possible by 1915. These new developments led John R. Carson of the Bell system to make a mathematical analysis of modulated waves. This showed the necessity of band width, which was to prove to be of great importance in connection with radio broadcast, initiated in the U.S. in 1921.

In 1920 radio amateurs started a series of experiments in transatlantic telegraphy at 200 m. (1,500 kc.) but attained success during the night only. Subsequent experiments by Marconi at 100 m. (3,000 kc.) were also successful, for he was able to obtain satisfactory nighttime reception from Poldhu to the Cape Verde Islands, a distance of more than 2,500 mi. Continuing his experiments at shorter wave lengths, he made in 1924 his second great discovery in the use of electric waves for the transmission of intelligence when he received 32-m. (9.4-mc.) waves over a distance of 2,400 mi. It is now known that he was receiving waves reflected from one of the ionized layers of the ionosphere. This new success of Marconi's led Oliver Heaviside in England and A. E. Kennelly at Harvard university to suggest that it was caused by reflections from an ionized layer located at a high altitude above the earth's surface. In England W. H. Eccles in 1912 had already analyzed wave propagation through a charged medium, and Sir Joseph Larmor in 1924 applied the same principles to a region containing freely moving electrons. Since then it has been established that there are, in effect, several reflecting layers in that portion of the earth's atmosphere now known as the ionosphere. This discovery of Marconi's initiated an era of radio transmission by means of ionospheric reflection. It was estimated that by mid-20th century about 95% of the world's long-distance radio communication was carried on at wave lengths of less than 25 m. by means of reflections from the ionosphere.

The Microwave Era.—Microwaves are usually defined as those electric waves occurring in the 0.1–10-cm. (300,000–3,000 mc.) region. When point-to-point communication is desired rather than general broadcast, the necessary output rating of the transmitter can be greatly reduced if the waves are focused in a narrow beam by means of a reflector or horn, or one of these in combination with a metal-plate lens. For example, if the output of one watt were concentrated in a conical beam whose vertical angle is $\frac{1}{2}°$, then it would require an output of 210 kw. to produce the same field intensity uniformly distributed over a complete spherical surface. This corresponds to a directive gain, or simply gain, as it is often called, of 53 decibels. An additional gain can be obtained by providing the receiver with a similar arrangement so that it is able to collect the energy over an appreciable area of the beam; *e.g.*, to increase its capture area. Although pronounced focusing effects can be obtained at all wave lengths, highly directive radiators must have dimensions of the order of magnitude of several wave lengths, and it is only at short-wave lengths, particularly in the microwave range, where such effective devices as horns, metallic reflectors and lenses are practical. However, directive arrays at all wave lengths have been in use since the earliest days of commercial wireless telegraphy and telephony.

Other advantages resulting from the use of microwaves are the large number of wide bands available for television and other purposes at such high frequencies, and the possible use of special types of equipment, including wave guides, cavity resonators, etc., many of which have exceedingly low losses and are small in size. Since the third discovery of Marconi, the minima for the lengths of electric waves being studied in the laboratory and for commercial use have been placed by the limitations upon the maximum frequency and power output of the available generators.

Microwave Generators.—The development of the three-element, space-charge vacuum tube for use as a generator of sustained oscillations, modulator, amplifier or detector open up new vistas for the use of electric waves. At first the possibilities seemed unlimited, but it was found that there were restrictions upon its general use at high radio frequencies. The first limiting factor encountered was attributable to the effect of the capacitances between the leads and electrodes of the tube. These capacitances were usually in the order of the magnitude of a few micromicrofarads. Going to higher frequencies, the transit time of an electron while traveling from cathode to anode of a tube became comparable with the period of the signals so as to cause a high grid input conductance. Among those who made analytical and experimental studies of this effect during the pioneer period of 1930–40 were B. J. Thompson, F. B. Llewellyn, W. R. Ferris and D. O. North in the United States and R. W. Sloane, E. G. James and W. E. Benham in England. It was found that this input conductance is caused by a grid conduction current which is induced by the electrons as they approach and pass by the grid. It is given by

$$G_g = Kg_m f^2\tau^2 \text{micromhos.}$$

In this, K = a constant for a given tube and depends upon the cathode-grid and grid-anode transit times; g_m = the transconductance of the tube in micromhos; f = frequency of grid signal; τ = transit time of an electron to the vicinity of the grid.

Special tubes have been developed in which these effects are minimized by using small and closely spaced electrodes with special connections and then by operating them with a high-anode voltage. Some of the more familiar are known as the acorn, doorknob and the later planar type in which the cathode, grids and anode are located in closely spaced parallel planes. The lighthouse tube first developed by E. D. McArthur of Schenectady, N.Y., is a well-known planar example and has become common usage at frequencies up to 3,000 mc. Other planar types have been constructed for direct connection to either coaxial cables or wave guides, while still others are provided with built-in cavity resonators. The interelectrode spacings in some of these vary from a few mils down to a fraction of a mil, with grid wire diameters and spacings being even smaller.

New principles of operation for a high-frequency oscillator tube were announced by A. Arsenjeva-Heil and O. Heil in Germany in 1935 and for both oscillators and amplifiers by R. H. Varian and S. F. Varian in California and by W. C. Hahn and G. F. Metcalf in Schenectady in 1939. Although these various tubes differed in detail, the operation of each depended upon the well-recognized fact that an electron acquires energy when accelerated by an electric field and delivers energy to the field when it is decelerated. Tubes of this type are known as velocity-variation or velocity-modulated tubes. In these, a stream of electrons is first accelerated to a high velocity by falling through a potential difference of at least several hundred volts. They next enter a narrow gap forming part of a cavity resonator system where they are further acted upon by a radio-frequency field, resulting in velocity modulation corresponding to the signal. As they continue through a comparatively long drift space, the faster ones catch up with the slower so that there is a bunching action. In this state the stream of electrons is now amplitude-modulated and induces a strong signal as it passes through the gap of a second resonator, finally to reach the anode, or collector, and to return to the cathode by way of an external connection. The commonest tube of this type is known as the klystron, a trade name introduced by the Sperry Gyroscope company, and gains as great as ten decibels have been obtained at 4,000 mc. over a band width of ±5 mc. The noise figure, which is defined as the ratio of the available noise at the output to the noise input of the tube, is about ten decibels higher than that for space charge amplifiers, so that the klystron could not be used to advantage as an amplifier of weak signals.

The velocity-modulated tube can be used as an oscillator by providing appropriate coupling, as by means of a section of coaxial cable between the output and input terminals. In the case of an oscillator, the first resonator assembly can be omitted if a negative electrode is substituted for the collector. This causes the electrons to oscillate back and forth within the gap of the resonator so as to result in a bunching effect. Such a tube is known as a reflex klystron and it has been used successfully for the production of oscillations up to 50,000 mc. (0.6 cm.), the usual output being from about $\frac{1}{2}$ w. up to a few watts, depending upon frequency.

In 1921 A. W. Hull of Schenectady described a two-element

vacuum tube, which he named the magnetron, in which the plate current was controlled by means of a magnetic field parallel to the axis of the tube. Such a field causes the electrons to travel in curved paths, resulting in a sharp decrease in plate current with increased magnetic field as cutoff is approached, this being the condition for all the electrons to return to the cathode. Because of this negative control characteristic, it was found that such a tube would function as a low-frequency oscillator by introducing coupling between the plate and field coil circuits. In 1924 E. Habann in Germany found by splitting the anode into two equal semicylinders that the tube would oscillate at a high frequency without any change in magnetic field; in 1933 E. S. McGaw in England was able to operate a tube of this type at 1,000 mc. (30 cm.); in 1936 G. R. Kilgore in Pittsburgh, Pa., obtained 100 w. at 600 mc. (50 cm.) with an efficiency of 25%; and in 1936 C. E. Cleeton and N. H. Williams operated one at 47,000 mc. (0.64 cm.).

A. L. Samuel of the Bell Telephone laboratories applied for a patent in 1934 on a multiple-cavity anode for a magnetron, which was further improved in 1940 by N. F. Alekseev and D. D. Malairov in the U.S.S.R. and by scientists at the University of Birmingham, Eng. In this a traveling wave is set up in the gap between cathode and anode with which the bunched electrons move in synchronism so as to transfer energy from the electrons to the fields in the cavities. Energy is then extracted from the field by means of a coupling element in one of the cavities and delivered to the load. One of the principal uses of the magnetron is to generate pulses of high-frequency energy for radar, it being quite common to produce a thousand 3,000-mc. pulses per second, each lasting one microsecond and having a peak power as great as two megawatts. This corresponds to an average power of two kilowatts. Magnetrons are also used for continuous wave generation when an exceptionally high frequency is desired, as in high-frequency cooking, for example.

The traveling-wave amplifier tube consists of a beam of electrons surrounded by a coaxial helix, the traveling wave being initiated at the end of the helix adjacent to the accelerating anode by means of a radio-frequency signal. By adjusting the velocity of the electrons in the beam, they can be caused to remain in step with the traveling wave and to deliver energy to it, this action being similar to that occurring in a magnetron. The initial development work on the traveling-wave tube was done by R. Kompfner at Oxford university in 1946. This was followed by additional study and consequential improvements by J. R. Pierce and L. M. Field of the Bell Telephone laboratories. The tube passed through the experimental stage and in the late 1950s found practical applications as a broad-band amplifier of microwaves, a typical tube having a gain of 30 decibels over a band of 3,600 ± 800 mc. and an output of several watts.

Microwave Transmission.—It is not practical to transmit microwaves over appreciable distances by means of open wires, principally because of the loss of energy by radiation. A coaxial cable is usually selected when a two-conduction line is desired since it is self-shielded. However, it is expensive and its losses increase rapidly with frequency. These consist of the Joulean loss in the outer and inner conductors, and any possible dielectric, leakage and reflection losses introduced by the supporting structure of the inner conductor.

A detailed study of the losses in a coaxial cable reveals that the loss in the outer conductor is usually small in comparison with the others. The possibility of transmitting waves within a hollow tube and thereby eliminating the cost, weight and losses in the centre conductor and its supports was considered at an early date in the application of electric waves. In 1893 Heaviside expressed the opinion that it was not possible, but J. J. Thomson suggested it as a possibility in his book *Recent Researches in Electricity and Magnetism* in connection with a mathematical study of the results of suddenly releasing a charge in a closed cylinder. During the following year Larmor analyzed the special case of waves generated by a spark-gap transmitter mounted within a hollow cylinder, to be demonstrated later by Sir Oliver Lodge. In 1896 Viktor von Lang set up waves in a long, hollow tube. Additional experimental work was carried out by P. K. L. Drude of optical fame in 1898, by R. H. Weber in 1902 and by H. Zahn and O. Schriever in the years 1910–20. To return to analysis, a rigorous mathematical study of the propagation of electric waves through circular and rectangular pipes filled with a dielectric was made by Lord Rayleigh (John W. Strutt) in 1897; through dielectric conductors by D. Hondros and P. J. W. Debye and through hollow tubes by Louis Silberstein in 1910.

It was not generally appreciated, however, that metallic pipes could be used as a practical means of transmitting electric waves until 1933, when several patent applications were filed by G. C. Southworth. These were followed by a series of papers based upon the work of Southworth and some of his colleagues at the Bell Telephone laboratories, notably S. A. Schelkunoff, Sallie P. Mead, J. R. Carson, A. E. Bowen and A. P. King. Contemporary work was done also by W. L. Barrow, Julian Schwinger, L. J. Chu and others in the United States; and by G. G. MacFarlane, D. G. Allen and others in England. After 1940 articles and texts covering a multitude of aspects of the problem were prepared by mathematicians, scientists and engineers, and hollow wave guides of standardized sizes were put in use as coupling elements in both transmitters and receivers as well as in radar equipment.

The principle of operation of a hollow wave guide is similar to that of an acoustical speaking tube in that the electric waves are confined within the tube by reflection from the walls. Reflections of electric waves can take place from smooth surfaces of conductors, such as silver, copper or gold, with very little loss as compared with the losses resulting from the presence of the inner conductor in a coaxial cable. Greater amounts of power can also be transmitted for the same maximum potential gradient, for wave guides have more favourable field patterns.

The applications of wave guides must be ordinarily limited to the microwave range, since at least one of the dimensions of any guide must be greater than $\frac{1}{2}$ wave length. This precludes their use for power transmission, for example, since the wave length corresponding to 60 cycles per second is about 3,100 mi. If the frequency of the input signal to a guide is decreased, it will be found that the attenuation will suddenly rise with further decrease. This can be explained in a qualitative sort of way by considering the hypothetical case of a transmission line consisting of two channel bus bars, as shown in fig. 3. It is obvious that this line can be operated at any frequency, including zero, provided that there is not a breakdown of the insulation separating them. If the two conductors are brought together so as to form a hollow tube, because of the potential difference between the top and bottom halves large transverse currents will flow at low frequencies and the input power will be rapidly attenuated. If the frequency can be increased without limit, a point will be reached where the potential difference between the opposite sides is opposed by a sufficiently high-induced voltage to prevent the flow of such a current. Thus, a hollow wave guide has a high-pass characteristic. Sections of wave guides can be used as matching and tuning elements, thus taking the place of the transformers, coils and condensers at lower frequencies. A section with an open or a closed end is known as a cavity resonator.

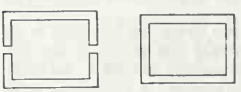


FIG. 3.—EVOLUTION OF RECTANGULAR WAVE GUIDE

Some additional types of microwave tuning elements are tuning screws and posts, irises or diaphragms and dielectric windows. In general, the values for Q, the quality factor, are high as compared with the maximum values obtainable for lower-frequency equipment, because of the relatively low losses as compared with the energy content of the electric and magnetic fields present. *See* ELECTRON TUBE; RADIO; TELEVISION; *see* also references under "Electromagnetic Waves" in the Index volume.

BIBLIOGRAPHY.—J. A. Fleming, *The Principles of Electric Wave Telegraphy and Telephony*, 4th ed. (1919); H. H. Skilling, *Exploring Electricity* (1948); Sir Oliver Heaviside, *Electromagnetic Theory* (reprint) (1950); Sir Edmund Whittaker, *A History of the Theories of Aether and Electricity*, 2 vol. (1951); M. Born and E. Wolf, *Principles of Optics* (1959); J. C. Maxwell, *Scientific Papers*, ed. by W. D. Niven, 2 vol. (1952). (C. S. Rs.; S. A. S.)

ELECTROMETALLURGY, the science and technology of metallurgical operations conducted by electrochemical means; specifically, the use of electrolytic processes in the extraction and refining of metals and in electroplating; the term also encompasses the use of electric arc, resistance and induction furnaces (*see* FURNACE, ELECTRIC; METALLURGY).

Sir Humphry Davy (1778–1829) first isolated the alkali metals by electrolysis of fused salts. Davy's brilliant student, Michael Faraday (1791–1867), succeeded in setting down the theoretical relations between equivalents of electricity and equivalents of mass in electrolytic cells. Other pioneers in the field included Johann W. Hittorf, Svante Arrhenius and Charles M. Hall (*qq.v.*). The early studies of electrolytes were aimed at the structure and degree of dissociation of ionic solutions, types of electrical conductivity and the separation of elements from aqueous and fused salt electrolytes. Later studies of thermodynamics established the energy relations involved in electrochemical reactions.

Electrometallurgy is concerned with mechanisms of electrical conduction, rates of ion migration, galvanic and electrolytic cells, thermodynamics of cells and polarization.

Electrical conductivity may be classified broadly into two main categories—electronic and ionic. Electronic conduction occurs by the passage of electrons while ionic conduction occurs by the migration of ions. In a closed electronic circuit no net mass transfer accompanies the flow of electrons along the current path; however, a mass flow does take place within electrolytic circuits, although no net mass flow occurs in the external circuit. This observation can be accounted for by the migration of relatively heavy ions of different masses in the electrolyte, in comparison with the motion of light mobile electrons in the external circuit. Faraday's laws govern the amounts of the electrode reactions in the following ways:

1. The amount of chemical oxidation or reduction produced by an electric current is proportional to the quantity of electricity passing between the electrode and electrolyte.

2. The amounts of different substances liberated by the same quantity of electricity are proportional to the chemical equivalent weights of the oxidized or reduced substances.

These laws represent a statement of 100% current efficiency which is the ultimate goal in any electrolytic process. In actual practice this is never achieved. (*See* also ELECTRICITY: *Faraday's Law of Electrolysis.*)

In galvanic cells or batteries, if two electrodes immersed in an electrolyte are connected by an external wire or circuit, the current flows through the wire from the positive cathode to the negative anode (*see* BATTERY). In direct contrast, in electrolytic cells, in which an external emf (electromotive force) is impressed across the electrodes, the cathode is negative and the anode is positive. This is expressed by the positive and negative sign convention, which is based on the defining relations for positive and negative ions. In electrolytic solutions positive ions, for example, Cu^{2+}, Ag^{-} or Fe^{2+}, are defined as cations; negative ions like Cl^{-}, $SO_4{}^{2-}$ or $(CN)^{-}$ are anions. The electrode toward which cations move when a suitable emf is impressed across the cell terminals is called the cathode; that toward which anions move is the anode. For all cells, galvanic and electrolytic, oxidation processes occur at anodes and reduction processes at cathodes. (*See* also ELECTRICITY, CONDUCTION OF: *Conduction in Liquids.*)

METALLURGICAL PRACTICE

Electrometallurgical operations can be classified into three important categories: electrowinning or extraction, electrorefining and electroplating. Electrowinning is a procedure in which (1) a raw ore is leached with an aqueous solution and the resulting electrolyte, rich in the metal to be recovered, is electrolyzed in order to produce a relatively pure metal deposit at the cathode, or (2) a raw ore is fused (or converted into an easily fusible salt) and the resulting fused electrolyte is electrolyzed, producing a relatively pure metallic cathode deposit. Electrorefining produces metals of extremely high purity by the electrolysis of solutions in which the anode is the impure metal to be refined, and the cathode is the pure metal deposit. The recovery of precious metals from the impure anodes is an important secondary refining operation in simple electrorefining. Electroplating consists of covering a metallic or nonmetallic object with a metal coating by making the object to be plated a cathode in an electrolytic cell which contains an electrolyte rich in the metal being deposited (*see* ELECTROPLATING).

Names of Electrolytic Processes

Metal	Fused salt electrowinning	Aqueous electrowinning	Electrorefining	Electroplating
Aluminum	Hall	—	Hoopes	—
Beryllium	—	—	—	—
Bismuth	—	—	Betts*	—
Cadmium	—	Midvale	—	Cadalyte
Calcium	Seward; Brace	—	—	—
Chromium	—	—	—	Fink
Cobalt	—	—	—	—
Copper	—	—	—	Acid; Cyanide
Gold	—	—	Wohlwill	—
Lead	—	—	Betts	—
Lithium	Osborg	—	—	—
Magnesium	Dow	—	—	—
Manganese	—	—	—	—
Nickel	—	—	—	Watts
Rhodium	—	—	—	—
Silver	—	—	Moebius; Thum	—
Sodium	Downs; Castner	—	—	—
Tin	—	—	Fluosilicic	—
Zinc	—	—	—	—

*As by-product of Betts lead process.

Common metals obtained commercially by electrowinning nonaqueous melts include aluminum, beryllium, calcium, lithium, magnesium and sodium. Metals commonly produced by electrowinning aqueous solutions are copper, cadmium, chromium, cobalt, manganese, nickel and zinc. Metals that are refined by the electrolysis of aqueous solutions are silver, bismuth, copper, gold, lead and tin. Metals and alloys which are commercially deposited by electroplating include copper and brass, cadmium, chromium, gold, nickel, rhodium, silver, tin and zinc. Many of the processes for electrowinning, refining and plating are referred to by proper names, indeed, sometimes vaguely by the names of the originators of the electrolytic cells or operations used since the 1850s. The accompanying table presents a fairly complete listing of the important names referred to in electrolytic processes. If the process has no proper name, a common name is indicated.

Electrowinning.—*Aluminum.*—A typical electrowinning operation can be described by the production of aluminum. The raw materials charged into a Hall electrolytic cell are purified alumina, Al_2O_3, cryolite, Na_3AlF_6, and small amounts of aluminum fluoride, calcium fluoride or sodium chloride. These latter additions serve to lower the liquidus temperature of the electrolyte and also to lower the density of the electrolyte. The Hall cell consists of a steel shell about 8 ft. long, 5 ft. wide and 2 ft. high, lined with a carbon cathode. Several cylindrical carbon anodes are suspended from an overhead bus bar into the electrolyte. Under the influence of the electric current oxygen ions collect at the carbon anodes and carbon monoxide and carbon dioxide are formed; molten aluminum is deposited on the cathode (the carbon lining) and collects at the bottom of the cell in a pool several inches deep.

Cadmium.—This metal is obtained electrolytically as part of the electrolytic zinc recovery process. Cadmium, which is found as an impurity in zinc sulfate electrolytes, is precipitated with zinc dust. The resulting zinc-cadmium sludge is filtered and the purified liquor is sent to zinc recovery cells. The sludge is redissolved in sulfuric acid, resulting in a cadmium-rich cadmium and zinc sulfate solution. The solution is electrolyzed, as in zinc recovery, with aluminum cathodes and lead anodes. The tendency for the cadmium to plate out as "trees" is minimized by adding glue to the electrolyte. At an emf of 2.6 to 2.7 v. and a current density of 4 amp. per square foot, cadmium plates out preferentially to zinc. Cathodes are stripped of cadmium every 12–24 hours, and the metal thus obtained is melted in iron pots, then cast into ingots of 99.95% purity or better. The over-all efficiency of cadmium recovery starting from the raw zinc-cadmium ore is 85% to 95%.

Magnesium.—In the United States magnesium is produced by the electrolysis of fused magnesium chloride. Most of the pure chloride is produced from sea water by alkaline precipitation, chlorination and evaporation. The electrolytic cells are iron or

steel shells or pots containing suspended carbon anodes and iron or steel cathodes. Before electrolysis the cells are externally heated to 675°–725° C., but once the current is turned on, resistance heating between anode and cathode maintains the temperature. The magnesium floats to the top of the molten chloride where it is tapped. Chlorine is the anode by-product. Anhydrous magnesium chloride is added to the bath from time to time to keep the process continuous. Cells are arranged in series of 100, operating at about 6 v. per cell.

Sodium.—Sodium is produced electrolytically by fused salt electrolysis. The older of the two major processes, using the Castner cell, is no longer employed in the United States. The Castner cell consists of an iron pot in which is suspended an iron or nickel cathode. Sodium hydroxide is the fused electrolyte, the reaction being carried out first above the melting point of the salt. In actual practice, current efficiency of only 40% is achieved; the voltage is about 5 v. Hydrogen gas is liberated at the cathode and oxygen at the anode, presenting a great danger of explosion in the process.

The Downs cell employs two of the cheapest chemical raw materials, sodium and calcium chlorides. It consists of a steel box lined with refractory brick. A graphite anode extends upward through the bottom, the cathode is an iron ring extending through the sides of the cell. A conical bell over the anode collects the only side product, chlorine. The metal, lighter than the bath, rises to the surface and flows into an external sump. The electrolyte is a 67–33 mixture of calcium chloride and sodium chloride. Current efficiency is greater than 85% at 7 v.

Zinc.—In the electrolytic production of zinc the starting material is zinc oxide which has been produced from sphalerite, ZnS, by roasting. The impure roaster product is leached with sulfuric acid to obtain zinc sulfate. The purified electrolyte is filtered and sent to concrete or wooden tanks containing aluminum cathodes and lead anodes. The electrodes suspended in the tank are about $\frac{9}{16}$ in. thick, spaced from $\frac{5}{8}$ in. to 2 in. apart. Generally, the decomposition potential of water is lower than that of zinc sulfate, but the high hydrogen overvoltage on aluminum allows zinc to be deposited on that metal without hydrogen evolution. Oxygen is evolved at the cathode and a frothing agent is added to prevent acid spray. The cells are operated at a current density of 30 to 35 amp. per square foot. The zinc is stripped from the cathodes after about 24 hours of operation and is generally remelted and sold as ingots. Current efficiency for the process is usually 92%–94%, the remainder of the current going toward the production of small amounts of hydrogen.

Electrorefining.—*Copper.*—A typical electrorefining process can be described by the production of electrolytic copper. More than 90% of the copper produced in the U.S. is electrolytically refined. Copper is refined in the multiple or series system. Multiple tanks employ fire-refined copper slab anodes and thin pure copper starting sheet cathodes prepared by prior electrodeposition and stripping. The copper sulfate electrolyte composition is maintained at about 40 g. per liter of Cu and 100 g. per liter of H_2SO_4. In the multiple system the anodes and cathodes in a given tank are in a parallel electrical arrangement while the tanks are placed in series. In the series system each electrode in the tank is anodic on on side and cathodic on the other side, so that only one pure starting sheet is needed as the final cathode in the tank. The advantages of the multiple system over the series system are chiefly in the ability to treat less pure copper and operate at a lower voltage. The series system gives a greater copper recovery per unit of power, requires less heavy bus bars to carry electrical current and produces less scrap.

In either system the copper plated is about 99.25% pure. As the copper is plated on the cathodes, the impurities settle to the bottom of the tank as "anode mud."

Gold.—If gold is at least 94% pure, it may be refined electrolytically by the Wohlwill process. The multiple system is used, similar to the copper process. Anodes are slabs $9 \times 14 \times \frac{1}{2}$ in. and the cathode starting sheets are gold foil about the same size as the anodes. The cell containers are porcelain or stoneware. The electrolyte is a hydrochloric acid solution of 7% to 8% gold chloride. As the gold is plated on the cathode, a slime containing gold, silver chloride, lead sulfate and platinum settles to the bottom. The cells are operated at 60°–70° C. with a current density of 110–120 amp. per square foot, and a potential of about 1.4 v. The gold cathodes produced average 985 fine, or nearly 24 k. (1,000 fine).

Nickel.—In the electrorefining process for nickel, impure nickel is produced by carbon reduction of nickel oxide in a reverberatory furnace and pressed into anode plates. The tanks contain 14 anodes per tank; the anodes and cathodes are separated by canvas diaphragms to prevent the codeposition of copper and iron. Nickel starting sheets are produced by plating nickel on aluminum or stainless steel, stripping the plate and attaching it to frames in the electrolytic cells. The copper in the impure nickel tends to build up in the electrolyte, and so this impurity, as well as iron, is removed by recirculating over nickel power in Pachuca tanks (*see* CYANIDE PROCESS). The pure electrolyte generally contains 20 g. of boric acid, 35 g. of sodium sulfate and 40 g. of nickel per liter. The electrolysis proceeds at a current efficiency of about 94% with a potential drop between electrodes of 2.4 v. and current density of 12 amp. per square foot.

Silver.—Silver is refined electrolytically in the Moebius or Thum cells. In the Moebius cell the anodes are 100-oz. slabs of 950 fine silver; the cathodes are $\frac{1}{32}$ in. silver or stainless steel sheets. The electrolyte is a neutral solution of 15 to 60 g. of silver nitrate per liter. Crystals of silver are scraped periodically from the cathode and fall into a tray at the bottom of the tank. Copper accumulates in the electrolyte and must be removed by a recycle purification. The slime, consisting of gold, platinum and other noble metals, is caught in a canvas bag surrounding the anode.

The Thum cell contains a graphite cathode which is a slab covering the entire cell bottom. Anodes of impure silver rest on a wooden or stoneware basket above the cathode. A canvas diaphragm separates cathode and anodes to catch the gold-platinum slime. Silver is deposited as crystals which are scraped off the cathode periodically. The Thum cell produces less scrap than the Moebius process, but requires a greater floor space and consumes a greater quantity of nitric acid in its operation.

Tin.—Tin is another metal that can be refined in a process similar to that for copper, although the method no longer is used commercially in the United States. The anodes are slabs of 95% tin, which are electrolyzed completely in about 21 days, during which time the cathode starting sheets are replaced twice. The electrolytic tin produced analyzes 99.98% or better. The impurities, chiefly silver, bismuth, lead and gold, show up in the anode slime which is treated by electrolytic and pyrometallurgical processes to recover these metals. The electrolyte most generally in use consists of 8% sulfuric acid, 4% cresolphenol sulfuric acid and 3% stannous tin with a small addition of glue to improve the quality of the deposit. Another type of electrolyte uses fluosilicic acid instead of sulfonic acid. The cells are operated at 35° C., a current density of 10 amp. per square foot and a voltage of 0.35.

CORROSION

The largest single waste in the entire metallurgical economy is corrosion. The principles of galvanic corrosion are simple: when two dissimilar metals or alloys are in contact with an electrolyte, which might be moisture in the air, the less noble one is attacked and the more noble one is protected. The sign convention which has been assigned to corrosion phenomena is that oxidation or "corrosion" occurs at the anode of the galvanic cell. If the hydrogen scale of galvanic potentials is referred to, the following common metals will replace each other or "corrode" preferentially when placed in contact with any other metal lower than they in the series. For example, zinc and magnesium would corrode preferentially to iron (valence of 2+), thus protecting it, while tin would not protect iron. Metals above hydrogen are generally considered electrochemically "active" while metals below hydrogen are generally "inert." The same conventions apply, however, to even the most inert metals and alloys; many dentists prefer not to use silver fillings adjacent to gold fillings because slight electrolytic action can promote toothache. A protective

coating of one metal over another may protect the base metal or may accelerate its corrosion, depending on whether the coating is more or less noble than the base and whether the base is entirely covered. Metal coatings applied by electrodeposition or hot dipping may contain pinholes that can set up local galvanic cells. When the coating is less noble than the base metal and protection should be complete, the surface coating cannot be effective if large areas of the base metal are exposed.

Industrial corrosion problems are often accentuated by stresses applied to the parts in use. Stress promotes corrosion by starting microscopic cracks which can be attacked by electrolytes. Furthermore, stressed areas are anodic to nonstressed areas: corrosion can increase effective stresses by concentrating them at a point where the metal part has been attacked.

The methods for combatting corrosion are simple and threefold. The first method is to paint or spray the metal or alloy to be preserved with a nonconducting medium like paint, asphalt, oil or fat, or to pack it in the medium. The second method is to apply a potential, by means of an external generator, which has an equal and opposite sign to that of the galvanic cell in action. This second method is generally reserved for use on large ferrous structures. The third and most common electrolytic method of cathodic protection is to provide a cheap active metal which will corrode preferentially to the base structure which is to be protected. This technique is employed through electroplates and slabs in contact with the base structure. Parts like storage cans, radiators, boilers and even immense pipelines have been successfully protected in this way.

See also ALUMINUM; CADMIUM; CORROSION AND OXIDATION OF METALS; ELECTROCHEMISTRY; TIN; etc.

BIBLIOGRAPHY.—Samuel Glasstone, *Introduction to Electrochemistry* (1942); *Principles and Applications of Electrochemistry*, vol. i, *Principles*, by H. J. Creighton (1943), vol. ii, *Applications*, by W. A. Koehler (1944); Donald M. Liddell (ed.), *Handbook of Nonferrous Metallurgy*, 2 vol., 2nd ed. (1945); U. R. Evans, *Metallic Corrosion, Passivity and Protection*, 2nd ed. (1946); John L. Bray, *Non-Ferrous Production Metallurgy*, 2nd ed. (1947); Samuel Field and A. D. Weill, *Electroplating*, 6th ed. (1951); Hellmuth Fischer, *Electrolytic Separation and Crystallization of Metals* (in German) (1954); A. K. Graham and H. L. Pinkerton (eds.), *Electroplating Engineering Handbook* (1955); A. G. Gray, *Modern Electroplating* (1953); S. Wernick, *Electrolytic Polishing and Bright Plating of Metals* (1951); American Society for Testing Materials, *Specifications and Tests for Electrodeposited Metallic Coatings* (1958); American Society for Metals, *Corrosion of Metals* (1946).
(R. E. GE.)

ELECTRON. One of the characteristic features in the 20th-century picture of the material universe is the analysis of matter into constituents smaller than the chemical atoms. The 19th century saw the rise and, finally, the general acceptance of the atomic theory of matter, according to which there are unique and distinguishable atoms of each of the chemical elements. The combination of these atoms in various ways was understood to produce all of the various chemical reactions and the vast variety of kinds of matter. The 20th century, on the other hand, saw the decomposition of each of these chemical atoms into constituent parts, and one constituent common to all atoms is the negative electron or, briefly, the electron. Historically this was the first subatomic constituent to be identified.

The first hint of the existence of a particle with much smaller mass than the lightest atom came from the study of the cathode rays that are produced when an electric discharge is passed through an evacuated tube. There was much controversy at the end of the 19th century as to whether these rays were waves or streams of small particles. However, the work of J. J. Thomson around 1897, and of numerous later experimenters, gave decisive support to the particle concept. Extensive study led to the conclusion that the particles, regardless of origin or means of production, all had the same ratio of charge to mass, and that the mass of each particle was only about 1/1,837 that of a hydrogen atom. Hence there could be little doubt but that such particles, eventually called electrons, were constituents of all atoms.

About 10 to 15 years later, but conceptually paralleling the work on the electron mass, came the final proof that electric charges are made up of discrete units, also designated as electrons. In view of the identification of negative electrons as particles with a definite though very small mass, it was necessary to conclude that an electrically neutral chemical atom possesses also a positive charge. This was located by Sir Ernest Rutherford on a heavy nucleus at the atom's centre. As a consequence of this work the term electron has been used in two different senses. On the one hand it refers to the elementary unit of electricity, and means an amount of electrical charge, either positive or negative. In this sense it was first suggested by G. Johnstone Stoney in 1881. On the other hand, it is often used to refer to the combination of electrical charge and mass that appears in cathode rays, and can be produced in many other ways. In this sense it is often pictured as a small material particle of definite mass, carrying a definite electrical charge, but poorly defined as to geometrical size and shape.

In 1932 C. D. Anderson announced the observation of positively charged particles with the same ratio of charge to mass as the negative electrons. These soon came to be known as positive electrons, or positrons. The positrons can be produced in the free state only by rather violent processes such as occur in cosmic ray phenomena or in the very high-energy accelerators of modern laboratories. Hence, they can be studied only under special circumstances. Nevertheless, the positive electrons take their place along with the negative electrons as elementary units of matter.

Before 1927 all of the experimental evidence was consistent with the idea of electrons as small pieces of electrically charged matter. The shape was not well defined, it is true, but in the absence of other information electrons were usually regarded as spherical. The size also was only vaguely known, except that it was believed to be much less than that of chemical atoms. The process of giving a negative charge to a body was understood to be the attaching to it of the requisite number of such electrons. On the other hand, giving the body a positive charge required the removal of a sufficient number of these material electrons. Thus a negatively charged body was expected to have a greater mass than the same body when positively charged, although the difference was too small to be detected.

In 1927, however, there were announced the famous experiments on the diffraction of an electron beam at a crystalline surface. It was shown that the electron beam behaves in just the same way as a beam of X-rays. It was also shown that the behaviour of the electron beam can be accounted for quantitatively, if there is attributed to it a wave length inversely proportional to the supposed speed of the electrons.

This result recalled the famous controversy at the beginning of the 20th century as to whether cathode rays were waves or particles. The difference was that, whereas the earlier discussion had been a controversy between differing interpretations of the few known facts, the situation after 1927 presented a dilemma recognized by all concerned. The evidence for both points of view was convincing, but the two points of view appeared to be mutually exclusive.

The resolution of this paradox was to be found in the theory of quantum mechanics (*q.v.*), which was in process of formulation during the 1920s. Concerning this particular problem, it was finally recognized that no entity so fundamental as an electron can be described by saying that it is something else more familiar, such as a particle, or a train of waves. An electron must be defined and described in terms of its own properties, and in terms of its behaviour under various conditions. The principal properties may be listed as the electron charge, the electron mass, the electron wave length and the electron spin.

THE ELECTRON AS AN ELEMENTARY CHARGE

The idea that electricity occurs always in multiples of an elementary unit of charge is closely allied with the idea of the atomic constitution of matter. The experiments of M. Faraday (*q.v.*) in 1833 showed that in electrolysis a definite quantity of material is deposited for every coulomb of electricity that passes through the solution. These experiments can easily be interpreted by the statement that in electrolysis each atom of material carries a number of electrons of charge equal to its valence. Nevertheless,

the experiments were not immediately so interpreted. The atomic constitution of matter, as well as the electronic constitution of electricity, were ideas that developed slowly, and had to be established on firm experimental grounds before they were generally accepted.

It was half a century after Faraday's work that Stoney recognized this atomic interpretation of Faraday's laws, and introduced the term "electron" to designate the elementary unit of electricity. Since Faraday had shown a charge of some 96,500 coulombs to be associated with the liberation of 1 gram of hydrogen, and since it was estimated that there were some 10^{25} atoms of hydrogen in a gram, Stoney was able to estimate the electronic charge as 10^{-20} coulombs or 3×10^{-10} electrostatic units. This is somewhat less than $\frac{1}{10}$ as large as the later values, but the method was entirely correct. The result was incorrect because of the erroneous idea of the number of hydrogen atoms in a gram.

During the first decade of the 20th century, an increasing belief in the physical reality of the atomic nature of matter, and the consequent atomic nature of electricity, led to several attempts to measure the electronic charge. These gave results considerably larger than the early estimate of Stoney, but they were all average measurements. It was not until 1911 that R. A. Millikan reported on the first of his series of famous experiments that showed in a direct and convincing fashion the electronic character of electricity. This experiment was repeated many times and the rapid development of atomic physics after 1910 was based on the almost universally held firm conviction that electricity is composed of indivisible electrons.

Proof of the Electronic Constitution of Electricity.—The most direct method of measurement of an electric charge is the measurement of the force exerted upon it by other charges. If the other charges are suitably distributed on two parallel conducting plates, the force will be proportional to the charge to be measured and will be the same in magnitude and direction at all points between the plates near their centre. Hence the arrangement used by Millikan, and schematically indicated in fig. 1, is suited for an accurate measurement of the electric charge on a small drop of oil.

Two circular brass plates indicated in the figure were kept charged by a 10,000 volt battery, so that the electric field between them was equal to this potential difference divided by the distance between the plates. The direction of the field could be reversed by means of a switch so that the field could be used to move the charged oil drop up against the force of gravity, or to move it downward faster than it would move under its own weight. The plates could also be short-circuited so that no electric field was present. The plates were enclosed in a tank from which the air could be evacuated, and the tank was immersed in an oil bath to keep the temperature constant.

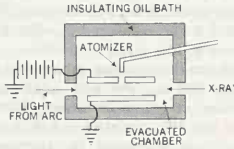


FIG. 1.—MILLIKAN'S EXPERIMENT (*See* TEXT)

Oil was blown into the space above the plates by means of an atomizer, so that a cloud of drops was formed. A few of these slowly fell through a small hole in the upper plate and entered the electric field. They were then illuminated by a strong arc lamp, and could be seen as bright "stars" in a telescope focused on the central region between the plates. When the plates were short-circuited, so that no field existed, all of the particles fell under gravity, and since the residual air acted as a viscous fluid they fell with speeds proportional to their weights.

When the electric field was applied, those charged with one sign slowed down or actually moved upward, while those of the other sign fell at an increased rate. As the heavier drops fell out of the field of view, it became possible to concentrate attention on only one drop, and to watch it rise with the field and fall under gravity repeatedly. Furthermore, it was possible to measure the time necessary for the image to move between two cross-hairs in the eyepiece of the telescope, and so to determine by direct measurement the speeds with which it rose and fell. These speeds provided the data from which the magnitude of the electric charge could be calculated and, perhaps even more important for the development of atomic physics, they provided direct proof that all of the charges were integral multiples of an elementary electron.

Table I gives some results obtained by Millikan in this type of experiment. The second column gives the time necessary for the particle under observation to fall a distance corresponding to the separation of the cross hairs in the eyepiece, or 1.0220 cm. In all of the ten round trips recorded in this table the time of fall was approximately the same, and had an average value of 51.48 seconds. This corresponded to a speed of 0.01985 cm. per second.

TABLE I

Times of fall and rise in seconds
Distance travelled = 1.0220 centimetres

Trip	Fall	Rise
1	51.13	30.55
2	51.25	21.86
3	51.19	50.72
4	51.32	148.63
5	51.53	147.46
6	51.69	50.29
7	51.55	50.25
8	51.54	50.39
9	51.98	49.70
10	51.64	146.41

Average time of fall = 51.48 sec.
Average speed of fall, $v_1 = 0.01985$ cm. per sec.

However, the times of rise were not all the same. Between the first and second trips the charge on the oil drop was changed by passing a beam of X-rays between the plates. The same was done between the second and third trips and the third and fourth.

The striking and important thing about the times of rise shown in Table I is that, although they are not all the same, they fall into a number of well-defined groups. There is no possibility that a time of rise of 30.55 belongs to the group of times near 50 seconds. These groups are collected in Table II, and the average values are found. From these the speed of rise, v_2, can be computed. Since the electric force on the charged drop must overcome the weight before the upward motion will start, it can be shown that it is the sum of the upward and downward velocities that measures the electric force, and therefore the magnitude of the charge. This sum (v_1+v_2) is recorded in the fourth column.

The significant feature of the fourth column is that the values all have a common factor, and that this common factor is contained only a small integral number of times, n, in each value. The conclusion must be that on trips 4, 5 and 10 the oil drop experienced 2 times the minimum electric force. On trips 3, 6, 7, 8 and 9 it experienced 3 times the minimum force, while on trip 1 it experienced 4 times the minimum force, and therefore carried 4 times the minimum charge.

An extensive series of experiments, in which hundreds of droplets were watched as they carried out repeated round trips, always led to the same conclusion. For each drop there was a minimum amount by which v_1+v_2 could change, and each observed value

TABLE II

Trip	Time of rise	Speed of rise, v_2	v_1+v_2	n	v_1+v_2/n
2	21.86	0.04675	0.06660	5	0.01332
1	30.55	0.03345	0.05330	4	0.01333
3	50.72	...	...	...	...
6	50.29	...	...	...	...
7	50.25	...	...	...	...
8	50.39	...	...	...	...
9	49.70	...	...	...	...
Average	50.13	0.02039	0.04024	3	0.01341
4	148.63	...	...	...	...
5	147.46	...	...	...	...
10	146.41	...	...	...	...
Average	147.50	0.00693	0.02678	2	0.01339

of this quantity was an integral multiple of the minimum. Of course the value of this quantity (v_1+v_2) was different for different droplets. The small droplets changed more in velocity than the larger ones, and the magnitude of the velocity changes depended also on the viscosity and density of the gas in which the droplet was falling. Furthermore, there were a few cases in which the charge on a drop changed while it was rising in the electric field. Such a change could easily be detected, however, since the drop could be seen to speed up or slow down while it was rising.

Extensive results of the type shown in Tables I and II gave convincing evidence that the net charge on one of the oil drops was always an integral multiple of an elementary electronic charge.

Oil Drop Measurement of *e*.—The oil drop experiment was also suited to the measurement of the actual magnitude of this electronic charge, but this demanded additional information. It required a method of determining the viscous resistance of the air to the motion of the drop through it. A law governing the rate of fall of a sphere in a viscous medium had been worked out by Sir G. G. Stokes in 1850, according to which the velocity of fall of a drop could be calculated in terms of its radius, its density and the coefficient of viscosity of the gas through which it was falling. Hence from the known rate of the fall, the known density of the oil and the viscosity of the air, the radius of the drop could be determined, and from this the magnitude of the resistance when the drop was moving upward.

To get good results by this method, it was necessary to modify Stokes' law so that it applied to very small droplets and also to make extensive new measurements of the viscosity of air. But when all corrections were properly taken into account, the result was a value of the electron charge e in satisfactory agreement with values obtained later in other ways.

X-Ray Measurement of *e*.—During the years after 1920 there was a great deal of interest in the exact values of various physical constants stimulated by attempts to establish theoretical relationships between them. Because of this interest, other methods of measuring the electronic charge were considered, and one was developed to a point where it exceeded the oil-drop method in accuracy.

This method goes back to the one originally used by Stoney in 1881, and consists in first determining the number of atoms in a unit weight of some standard crystal.

An accurate measurement of the separation of atoms in a crystal can be made by X-ray diffraction, if the wave length of the X-rays is known accurately. This wave length can be determined from the diffraction pattern of the X-rays when scattered by a ruled grating, since the number of lines per centimetre on the grating can be determined by counting. From the number of atoms in a cubic centimetre of a crystal, and its density, there can be determined the number of atoms per gram atom, N.

By the early 1960s the best value of N was taken to be 6.0250×10^{23}. Hence, 16.000 grams of oxygen, or 107.880 grams of silver, contain 6.0250×10^{23} atoms. The determination of this number was really an accomplishment of a high order, for it is known with greater accuracy than is the population of any large city.

The combination of the value of N with the value of the faraday, or the charge carried by a gram atom of an element in electrolysis, gives an accurate value of e. By the early 1960s the best value of e, determined by various methods, was 4.8027×10^{-10} electrostatic units or 16.020×10^{-20} coulombs.

THE ELECTRON AS A UNIT OF MATTER

The idea that negative electrons are constituents of all atoms arises from the fact that they can be extracted from all kinds of substances, and from the fact that they have mass. This mass is not measured by its gravitational effect, but by its inertia. Since the electron experiences a force in an electric field, and since this force is known when the electron charge is known, the inertia, or mass, can be measured in terms of the acceleration in the electric field. The quantity that is measured is actually the specific charge, or the ratio of the charge to the mass of the particles in an electron beam. The fact that this can be measured with precision indicates that it is the same for all of the electrons in the beam. This, coupled with the idea of an elementary charge e, gives the picture of identical electrons as fundamental components of all matter and all atoms.

Specific Charge (e/m) for Free Electrons.—As early as 1890, A. Schuster made rough measurements of the specific charge of electrons by observing the behaviour of cathode rays in electric and magnetic fields. More accurate measurements were made by J. J. Thomson in 1897, and during the succeeding 50 years methods were developed for the precision measurement of this important quantity. Although the detailed arrangements used in each case were different, the general principles can be illustrated by describing one of them which is shown schematically in fig. 2.

In this arrangement two metal electrodes are sealed into a glass tube from which the air has been evacuated to such an extent that the pressure is below 1/10,000 atmosphere. The cathode is connected to the negative terminal of a battery or other source of potential difference of several thousand volts, and the anode is connected to the positive terminal. Under these conditions a beam of cathode rays is emitted from the cathode toward the anode, and a portion of the beam passes through a small hole in the anode and causes a spot on the fluorescent screen to glow. The spot indicates the point at which the cathode rays strike the screen. When the pressure of the residual gas is properly adjusted, the path of the electrons through it can be seen faintly. It is then apparent that in the absence of an electric or a magnetic field between the anode and the screen, the beam travels in a straight line.

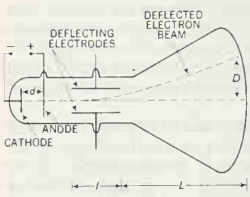


FIG. 2.—EXPERIMENT FOR DETERMINATION OF e/m (THE SPECIFIC CHARGE OF ELECTRONS) (*See* TEXT)

If now a potential difference is applied between the deflecting electrodes so that the upper one is positive, the spot on the screen moves upward. If the upper one is made negative, the spot moves downward. This shows that the beam is negatively charged. If the potential difference between the electrodes is V, and their distance apart is d, the electric field between the electrodes is V/d, and the force on a charge e will be Ve/d.

If V and e are measured in absolute electromagnetic units and d in centimetres, the force will be given in dynes. On the other hand if V is measured in volts, e in coulombs and d in metres, the force will be given in newtons.

Since the electrons composing the cathode ray beam are moving from left to right with some speed, v, they will be between the plates for a time l/v where l is the length of the plates. If, then, the fact that the electric field is not sharply limited at the edge of the plate is ignored, the electrons experience the force Ve/d for the time l/v. During this time they acquire an upward velocity equal to $\frac{Vel}{dmv}$ (since force $\times$ time = momentum, which is mass $\times$ velocity; hence force $\times$ time $\div$ mass = velocity). This velocity acts as though it were acquired suddenly, at the middle of its travel between the plates, so that the deflection D of the spot on the fluorescent screen will be

$$D=\frac{eVl}{mdv^2}\left(L+\frac{l}{2}\right)$$

The deflection D can be measured directly on the screen, and if v is known, e/m can be calculated.

It is important to use a consistent set of units in the preceding equation. If the lengths are all measured in centimetres, the velocity in centimetres per second, and the potential difference V in electromagnetic units, e/m will be given in electromagnetic units of charge per gram of mass. On the other hand, if practical units are used, the lengths are measured in metres, the velocity in metres per second, and the potential difference in volts, e/m will then be given in coulombs per kilogram.

There still remains the problem of finding the velocity v before

the value of e/m can be determined. To find this a magnetic field is produced, perpendicular to the plane of fig. 2, and limited as far as possible to the length l along the path of the cathode rays; its magnitude will be denoted by B. In such a field the cathode ray particles experience a force perpendicular to the direction of the field and to the direction of the beam. This force can be made opposite in direction to the force due to the electric field. When these two forces just balance each other the beam travels in its original straight line and strikes the screen at the central point. The condition for the equality of the forces is

$$B\,ev = Ve/d$$

This gives the velocity v in terms of the electric and magnetic field strengths that just balance each other. When B is measured in gauss with V in electromagnetic units and d in centimetres, v is given in centimetres per second. In the practical system, B is measured in webers per square metre, and v is given in metres per second. The value of v obtained in this way is then used in the previous equation to determine e/m.

This general principle was used with many modifications. Many different arrangements of the electrodes and the fields were used to improve the accuracy of the results and to make use of electron beams produced in a wide variety of ways. Out of a great many experiments carried out by many different persons came several definite conclusions.

1. The value of e/m obtained by measurements such as that described above is independent of the nature of the residual gas in the tube and independent of the material of the cathode. If the vacuum is very high, and ultraviolet light shines on the cathode, a photoelectric current flows in the same way as the cathode-ray beam and gives the same value of e/m. If, instead of a potential difference between the cathode and the anode, a bit of radioactive material is placed on the cathode, the β-rays emitted from it also show the same value of e/m. If the cathode is heated to incandescence, the electrons emitted from it show the same value of e/m.

Out of these experiments grew the conviction that all of these various beams were composed of identical, negatively charged particles, and that these particles were one of the fundamental constituents of all matter.

2. The value of e/m observed for these particles was some 1,837 times as great as the corresponding ratio in the electrolysis of hydrogen. Assuming identical values of e, this indicated the existence of particles of very small mass, so small that they were almost negligible compared with the remainder of the atoms from which they had been removed.

3. The value of e/m is always observed to decrease as the velocity of the beam is increased. The observations are well represented by the equation for the variation of mass given by the theory of relativity. This requires that

$$\frac{e}{m} = \frac{e}{m_o}(1 - v^2/c^2)^{\frac{1}{2}}$$

where m_o is the "rest mass" of the electron and c is the velocity of light. At first, this type of variation was thought to indicate that the cathode rays, or electrons, were "pure electricity," but the point of view of the theory of relativity shows that no such conclusion is necessary. As a matter of fact, it is not easy to define just what is meant by "pure electricity."

4. By the early 1960s the best value of e/m_o, based on all the available measurements, appeared to be 1.7589×10^7 electromagnetic units per gram or 1.7589×10^{11} coulombs per kilogram.

Specific Charge for Atomic Electrons.—The type of measurement of e/m just described applies to "free electrons" (electrons that have been removed from matter and the atoms of which they were once a part). It is also possible, however, to measure this quantity for electrons that are still parts of atoms.

It was observed by P. Zeeman in 1896 that if a source of light is placed in a magnetic field and the light is studied with a spectroscope, the characteristic spectral lines become split and separated by an amount proportional to the magnetic field. H. A. Lorentz immediately calculated the effect to be expected on the assumption that the frequencies of light, emitted under the conditions of the experiment, were the frequencies of oscillation of electrons bound by quasi-elastic forces to positions of equilibrium within the atoms. He showed that the magnetic field would disturb the motion of such an electron so as to give the splitting, as well as the polarization, observed. This gave support to the idea that the free negative electrons of the cathode rays were the same as the electrons within the atoms. Although further experimental work showed that the simple theory of Lorentz was inadequate for the detailed description of the phenomenon, the development of the quantum theory of atomic spectra made possible the use between 1930 and 1935 of studies of the Zeeman effect for a precision determination of e/m_o. This work gave results in substantial agreement with the result quoted above.

QUANTUM THEORY OF ELECTRONS

During the first quarter of the 20th century an electron was generally treated as a small piece of matter, with a known mass and a known electrical charge. It was considered to move under the attractions and repulsions of other electrified particles according to the laws of mechanics; and an atom was often likened to a miniature solar system in which numerous electrons moved in orbits about a heavy, positively charged, nucleus.

About 1925, however, extensive experimental information concerning the behaviour of atoms led to attempts by L. V. Broglie, W. Heisenberg, P. A. M. Dirac and E. Schrödinger to formulate a different picture of atomic structure. The picture at which they arrived suggested that electrons might display the properties of waves as well as those of particles.

Electron Wave Length.—An experimental test was soon made. Two independent pairs of investigators, C. J. Davisson and L. H. Germer of the Bell Telephone laboratories in New York city, and G. P. Thomson and A. Reid of the University of Aberdeen (Scot.), announced the diffraction of electrons in 1927.

Since the days of Thomas Young and A. J. Fresnel at the very beginning of the 19th century, the phenomenon of diffraction has been regarded as conclusive evidence of a wave motion. The wave nature of visible light was established on the basis of various diffraction experiments which enabled the wave length to be measured accurately. M. von Laue, W. Friedrich and P. Knipping observed the diffraction of X-rays in 1912, and so established them as electromagnetic waves and fixed their place at the short end of the wave length scale which included visible light and radio waves. Then, numerous experiments in the years immediately following 1927 seemed to show just as clearly the wave nature of cathode rays, and to permit accurate measurement of their wave length.

The diffraction of a beam of electrons was observed in a number of ways, but in all cases a grating of atomic dimensions was required, as in the case of X-rays. Davisson and Germer observed the scattering from the face of a single crystal of nickel. They found that electrons of a fixed incident velocity were scattered only at certain definite angles. Thomson and Reid used another technique already in use with X-rays. They sent their beam of electrons through a thin film of celluloid or a thin film of metal. A photographic plate behind the film and perpendicular to the beam showed a series of concentric circles similar to those shown by X-rays in the same arrangement.

The results of all the experiments could be described accurately by attributing to the electron beam a wave length $\lambda = h/mv$, where $h = 6.625 \times 10^{-27}$ erg sec. (Planck's constant). When m is expressed in grams and v in centimetres per second, the wave length λ is given in centimetres. An electron that has gained speed in moving between two plates connected to a one-volt battery will have a wave length of 12×10^{-8} cm. This is rather long for X-rays and is about $\frac{1}{500}$ the wave length of red light.

The methods of observing electron diffraction were steadily improved until by 1940 the process could be reversed and the diffraction of electrons used along with that of X-rays to study crystal and molecular structures.

Electron Spin.—Another result of the extensive study of atomic spectra appeared in 1925. In that year G. E. Uhlenbeck

and S. A. Goudsmit suggested that the detailed structure of spectra could best be explained by attributing to each electron a spin and a magnetic moment. Later developments completely confirmed the idea that every electron, positive or negative, has as part of its description (a) an angular momentum and (b) a magnetic moment.

The angular momentum is attributed to an electron since its behaviour can be pictured as something like that of a spinning top or a gyroscope. The magnetic moment of the electrons is considered to be the source of the magnetic behaviour of materials such as iron and nickel.

Dirac's Equation.—In 1928 Dirac proposed a new equation to describe the behaviour of an electron in a given electromagnetic field. In formulating the equation he required that it give the observed wave properties and at the same time be "relativistically invariant." The latter requirement is that the equations be unchanged in form when the co-ordinates of the electron's position, and the time, are subjected to the Lorentz transformation of special relativity theory. Dirac's new equation was thus a generalization of his earlier work and that of Broglie and Schrödinger, to include the variation of the mass with velocity as mentioned above.

As a consequence of this requirement of relativistic invariance there came two conclusions which were not anticipated. The first was the property of spin. An electron, behaving as required by the equation, shows a magnetic moment and an angular momentum. The angular momentum amounts to $\frac{1}{2}h/2\pi$ gram-centimetre-centimetre per second, and the magnetic moment is $2e/mc$ times the angular momentum. Until 1948 this value of the magnetic moment agreed with the observations well within the accuracy of measurement.

The second consequence was the apparent conclusion that an electron could exist in a state of negative kinetic energy. Since kinetic energy, as ordinarily understood, is an essentially positive quantity, this second conclusion was regarded for a time with some skepticism. However in 1933, at about the same time that Anderson observed positively charged electrons, the negative energy states were interpreted as giving a basis for understanding the behaviour of positive as well as negative electrons. On the basis of the theory it could be predicted that a positive and a negative electron, upon colliding, will annihilate each other and emit all of their energy, including the energy of their masses, in the form of electromagnetic waves of very short wave length. Also it could be shown that a beam of radiation of suitably short wave length can, under appropriate conditions, produce a pair—a positive and a negative electron. All of these predictions were confirmed accurately by experiment. The transmutability of electronic mass into radiant energy, and conversely, of radiant energy into electronic mass, was firmly established.

STRUCTURE OF AN ELECTRON

The normal process of analyzing matter into smaller and smaller parts led naturally to speculations concerning the shape, size and structure of an electron. If the observed mass is associated with the energy in the electric field surrounding the electron, a minimum radius for the charge can be estimated.

The energy in the electric field surrounding a charge e, and outside a radius R, is $e^2/2R$. If this energy is taken to be the source of the electron's mass, and therefore equal to mc^2, it follows that $R = e^2/2mc^2$. The electron radius can be no smaller than this value, which is 1.4×10^{-13} cm., or there would be more mass in the electric field than can be accounted for by the observed electron mass.

Experimentally it might be thought that the radius of electrons could be measured by observing collisions between them, as is done for alpha particles and other nuclei. It might be conjectured that if two electrons are shot at each other with sufficient energy, approaching the value of mc^2, they might collide, and some evidence could be obtained as to a minimum distance inside of which the simple inverse-square law of force breaks down. However, according to quantum mechanics, as verified by observations, such electrons would have a wave length of about 2×10^{-10} cm. This is about 1,000 times the radius anticipated, and consequently the experiment would have to be described as the interaction of waves and would give little or no information about a radius.

Furthermore, the spirit of quantum mechanics suggests that it will be impossible to analyze an electron further. The more recent approach has been to treat the electron as a point charge, but to analyze in more detail its interaction with the surrounding electromagnetic field.

Electrodynamics was first developed in terms of the forces between charges and between currents. Later work by Faraday, which culminated in the theory of Maxwell and the experiments of H. R. Hertz, focused attention on the electromagnetic field. In the electron theory of Lorentz both the electrons and the field were considered, but although the electrons were thought of as "producing" the field, the detailed connection between them was a little vague.

The early forms of quantum mechanics, including Dirac's relativistically invariant equations for an electron in a given electromagnetic field, treated the electron as the principal object of study. The field was merely the source of the prescribed forces. However, the continued study of the process of emission and absorption of radiation led to the application of quantum mechanical methods to the field itself. The next step was naturally a more detailed analysis of the interaction between the electrons and the field. This culminated around 1950 in the theoretical computation of two departures from the predictions of the earlier theories, which have been fully confirmed experimentally.

According to the quantum mechanical treatment, the electric field at a point in empty space is not zero but fluctuates. The interaction of a "point" electron with this field results in an energy different from that computed in the classical way. Hence, although the electron's structure is not specified and it is treated as a point, two observable effects are to be expected.

There are small deviations between the frequencies of the light radiated by a free hydrogen atom and those predicted by Dirac's early equations. Such departures were in fact observed as early as 1934, but in 1947 a new technique permitted their measurement with high accuracy. Also in 1947, it was shown by P. Kusch and H. Foley that the magnetic moment of an electron in an atom, as measured by the Zeeman effect, differs from that required by the Dirac theory by some 0.12%. According to Dirac, the ratio between the magnetic moment and the angular momentum should be $2e/mc$. The detailed treatment of the interaction with the electromagnetic field leads to the value $(2 + e^2/hc)e/mc$, which is in good agreement with the observations.

Thus, the structure of an electron in the simple pictorial sense is relegated to the background. The methods of quantum mechanics, however, lead to descriptions of experimental results in close agreement with the observations.

See ELECTRICITY; ELECTRICITY, CONDUCTION OF; MASS SPECTROSCOPY; PARTICLES, ELEMENTARY; *see* also references under "Electron" in the Index volume.

BIBLIOGRAPHY.—R. A. Millikan, *Electrons (+ and −), Protons, Photons, Neutrons, and Cosmic Rays*, rev. ed. (1947); G. P. Thomson and W. Cochrane, *Theory and Practice of Electron Diffraction* (1939); F. K. Richtmyer, *Introduction to Modern Physics*, 4th ed. (1948); V. Weisskopf, "Recent Developments in the Theory of the Electron," *Reviews of Modern Physics*, vol. 21, pp. 305–315 (1949). (W. V. H.)

ELECTRON DIFFRACTION. Electrons (*q.v.*) traveling in a narrow beam are like X-rays (*q.v.*) in one important property. If the beam falls on a crystal, the electrons are scattered in directions which depend on the positions of the atoms in the crystal. This experiment provides one example of the phenomenon called electron diffraction.

The diffraction experiment may be considered to provide proof of a wave character of electron beams, just as analogous experiments have been considered to provide proofs of the wave characters of X-rays and of visible light (electromagnetic radiation). Although the original study of the properties of electrons begun by J. J. Thomson (1897) had apparently demonstrated that these entities were material particles of a definite mass and of a definite charge, the wave properties were predicted by L. V. de Broglie (1924), who proposed the following relation between the wave

length λ associated with the electron and the product of its mass and velocity:

$$\lambda = \frac{h}{mv} = \sqrt{\frac{150}{V}} \times 1.008 \times 10^{-8} \text{ cm.}$$

where h is Planck's constant (6.625×10^{-27} erg-sec.), m is the mass of the electron, v is its velocity and V is the voltage used to accelerate the electron.

The first experimental tests, performed by C. J. Davisson and C. H. Kunsman in 1923 and continued by Davisson and L. H. Germer in 1927, showed quite clearly that the directions of low-voltage electrons scattered by single crystals of various metals depended on the atomic arrangements in the metal crystals. In 1928 G. P. Thomson, using electrons accelerated by about 40,000 v., *i.e.*, having a wave length of about 0.06×10^{-8} cm. or about one-tenth of the usual X-ray wave length, reported photographs from thin metal foils which showed patterns of concentric rings analogous to those obtained from the same materials with the use of X-rays. The diameters of the rings were dependent on the interplanar distances in the crystals and on the wave length of the electrons according to the Bragg law previously developed for X-rays. New diffraction effects not possible with X-rays were observed by S. Kikuchi (1928), who used high-voltage electrons on mica single crystals. Diffraction by molecules in the vapour state was discovered by H. Mark and R. Wierl in 1930.

Important differences between electron and X-ray diffraction arise from the lower penetrating power of electrons. This means that electrons are much more sensitive to small amounts of matter and are preferred over X-rays for studying gases and surface films, which may be only a few atom layers thick. X-rays are preferred for studying the interior structure of crystals. After about 1950 a third type of diffraction using neutron beams was widely applied for locating atoms in crystals containing hydrogen atoms or pairs of elements indistinguishable by X-rays; *e.g.*, iron and nickel.

The experimental technique when electron diffraction is used for determining the positions of atoms in solids requires the generation of a fine beam of high-speed electrons in a high-vacuum camera. If the specimen is not more than 10^{-5} cm. thick, it is mounted in the path of the beam so that the electrons pass directly through it; on the other hand, if the specimen is very thick, it is mounted so that the electrons fall on, and are reflected from, the surface at an angle of two or three degrees. In the study of gases, a jet of the vapour is directed across the electron beam. The electrons, after having come in contact with the specimen, are registered on a photographic plate, which is affected by electrons in the same way in which it is by light; the plate then records a diffraction pattern which is characteristic of the specimen. The exposures are for a few seconds, in contrast to the exposures of several hours required for X-rays under similar conditions. Special high-vacuum equipment is needed to prevent electron scattering by air molecules along the beam and, with gas specimens, to remove the vapour as soon as possible.

The interpretation of the diffraction photographs leads either to detailed information on the location of the separate atoms or more frequently to an identification of the chemical substance responsible for the diffraction pattern. For example, in a molecule of carbon tetrachloride the carbon atom is located in the centre of a regular tetrahedron of chlorine atoms with a carbon-chlorine distance of 1.765×10^{-8} cm. (accurate to about 0.3%) and with thermal vibrations of about 0.060×10^{-8} cm. in amplitude. Similar results usually less precise have been obtained for about 600 different molecules studied in the gaseous state.

Films on solid surfaces are another type of specimen extensively investigated. The acid derivatives of hydrocarbon oils, such as stearic acid, form films only one molecule thick on smooth surfaces with the zigzag chains of carbon atoms standing normal to the surface. The presence of such films observed by electron diffraction has a profound effect on the ease with which liquids wet the surface and is important in the fields of lubrication, separations of mineral ores by flotation and the cleaning of surfaces.

The identification of substances in minute amounts without a detailed structure determination represents the most frequent application of electron diffraction. The optically transparent layer of oxides on polished surfaces of copper, iron and other metals, the very thin crystallized layers on the surfaces of many glasses, the chemical compounds on catalytic surfaces, the minor phase particles occurring in heat-resistant alloys are systems in which electron diffraction information has been of great importance. *See* also Atomic and Molecular Beams.

Bibliography.—G. P. Thomson and W. Cochrane, *Theory and Practice of Electron Diffraction* (1939); Z. G. Pinsker, *Electron Diffraction* (1953); H. Raether, "Elektroneninterferenzen" in *Handbuch der Physik*, vol. 32 (1957). (L. O. B.)

ELECTRONIC MUSICAL INSTRUMENTS. This term is applied to all types of sound-producing apparatus, directly or indirectly under the control of a performer, in which the vibrations or oscillations generated by the source are electrical in form and inaudible except through the medium of amplifiers and loudspeakers or headphones. To be exact, the term electrophonic (recommended by the council of the Royal Musical association) is more accurate, since some of the methods used do not depend on a source actually emitting electrons. It is also to be understood that the initial sources must generate the oscillations and not merely convert them in the sense, for example, typified by the use of a microphone.

Instruments conforming to this generic name may produce single notes or chords; may be playable from keyboards or may have a continuously variable tuning system; or can be operated by a punched tape or other coding system. Modern methods may employ electron tubes (valves) or transistors in feedback oscillators, or may have rotating or vibrating electromagnetic or electrostatic oscillation generators.

Historical Development of Types of Instrument.—It can be said that real progress in the development of electronic music was only made possible by the introduction of the three-electrode tube in 1908. Experiments go back as far as 1837, and a successful though massive generator was made by T. Cahill in 1897 (on which the Hammond organ is based). Although the thermionic tube was the obvious choice for experiments, the high degree of instability attending such tubes and their associated components was not really overcome until World War II. Therefore the Hammond organ (1935) employed rotating magnetic generators; the Compton Electrone (1932) used rotating electrostatic generators; and the Hoschke instrument (1934) had vibrating wind-maintained reeds and electrostatic conversion. Later this instrument was developed into the Everett Orgatron and later still became the Wurlitzer. All the above are polyphonic, that is, full chords may be played thereon.

A large number of melodic or one-note-at-a-time instruments have appeared, some dating back to 1928. But again, with modern tubes and components only the Clavioline (1949), Univox (1954), Solovox (1946), Neonvox (1956) and a revised version of the Martenot (1928) and Trautonium (1930) remained in use in the second half of the 20th century. The last-named instrument was not available commercially.

Development in tube-generating systems of the organ type was rapid not only because tubes and their associated components became so stable but also because the techniques for forming imitative tone colours from complex wave forms were increasingly understood. Although other methods, such as photoelectric generators and magnetic tape systems carrying preformed tone colours, are of experimental interest, the physical and engineering difficulties are formidable and no satisfactory commercial solution of the problems involved in their use had been reached by the 1960s. Instruments in use were principally of the multi-keyboard organ type.

The RCA music synthesizer, another electronic musical-instrument in use in the 1960s, operated from a coded tape punched with all the fundamental factors required to form a musical sound—pitch, duration, loudness, rate of attack, rate of decay, reverberation, harmonics, vibrato and (if required) electrical "noise." It can be said that, given sufficient complexity and manipulative skill, almost every known and many quite new musical sounds may be synthesized satisfactorily.

Function of Electronic Instruments.—Electronic musical instruments had become generally accepted by the mid-20th century, but it is still interesting to remember the arguments put forward for their introduction and the variety of their uses. In the more comprehensive examples, as the Trautonium, effects can be produced over a far wider pitch and power range than can be attained by any other conventional instrument. This is particularly true of musical percussions. The Trautonium is also able to endow conventional tonal synthesis (*e.g.*, to reproduce the tone of the clarinet) with attributes otherwise unattainable, including either extended portamento or rapid staccato "speech," less noise, more uniform tone quality over the compass, and, in addition, it can add an octave or two to both ends of the tone range. Such properties are not only valuable in themselves but are of great help to the designers of the corresponding conventional instruments as a medium of research and comparison.

There is a particularly good case for the electronic organ type of instrument. Pipe organs are cumbersome and difficult to service; their acoustic efficiency is very low in comparison with the power input supplied to the blowing fan; few people understand their mechanism; tuning is difficult and must be done often; there is usually not enough space to deploy pipes so that they can speak to their best advantage; once pipes have been made, they cannot be adjusted in tone or volume; the over-all loudness expression control is poor; the organ can usually only be placed in one position because of its size; one cannot exchange a rank of pipes for one of some other tone quality; metal for pipes, if of good quality, is very expensive; it is difficult, if not impossible, to protect the working parts against insects and fungus in the tropics.

The electronic organ overcomes all these deficiencies, but it must be remembered that it is not possible to attain the massive bass energy of large pipes by any known type of loud-speaker system. Thus such instruments are at their best when confined to a modest tonal scheme. Here, if properly designed, they score heavily, for example, as in the three-manual organ for Khartoum cathedral. By omitting the 32-ft. tone, the cost, compared with a corresponding pipe organ, was halved. The instrument was able to be fully tropicalized, and tuning is required only at intervals of two to three years. The largest electronic organs use tubes exclusively. Modern tubes can be relied on for a life of 5,000 to 10,000 hours.

Principles of Operation.—Many methods may be used to generate suitable oscillations, for musical synthesis; some of these are now only of academic or experimental interest.

The rotating magnetic generator is only to be found in the Hammond organ. A synchronous (constant speed) motor drives a set of toothed iron wheels through gears so proportioned that each wheel is advanced in pitch by one semitone of the equally tempered scale. The pitch range is a little over eight octaves. The iron wheels rotate in front of permanent magnets, each of which carries a coil of wire, and in so doing generate a voltage in the coils. These voltages are used to form the pitch note for each key, while other wheels provide a series of harmonics which can be added to the pitch note to produce complex musical tone colours. The number and loudness of these harmonics are under the control of the player by electrical switches. An amplifying system raises the volume to any required extent. The instrument is portable and permanently in tune. Its peculiarly rapid attack and unique harmonic structure make it most suitable for rhythmic playing.

The Compton Electrone has belt-driven electrostatic generators on which various wave forms are engraved. The pulleys are of such dimensions as to make each successive disk increase in pitch by one semitone. There is a much more extended range of harmonics than on the Hammond, and the wave forms are more pure; consequently the synthesis is much better. More important still, since the disks are supplied with direct current to produce the electrostatic field, they are in effect capacitors. Therefore, if supplied through suitable resistors, the rate of charge can be slowed down as required; so the way the sound starts can be made to correspond very closely with that of pipes. In another model with smaller disks, the reverse effect can be obtained by the movement of a switch; thus percussions can be obtained.

In the Wurlitzer organ, similar properties are exhibited by vibrating reeds acting as variable capacitors, but since the natural harmonic development is limited, only a few different tone qualities can be synthesized. This instrument is suitable for home use and holds its tune for long periods.

In the extensive class of tube organs, two methods apply. One school of thought contends that the most satisfactory tonal synthesis results from the addition of as many harmonics as thought desirable, following the theories of Hermann von Helmholtz and Jean Fourier. While this is true, the harmonics must be true multiples of the pitch note. Using the equally tempered scale keyboard, this is an impossibility, because the only true exact frequency multiples are octaves. Therefore adding even and odd harmonics to any extent produces a discordant sound. The German A. W. B. organ provides the sole example of this technique.

The great majority of instruments generate a complex wave form having many harmonics for each pitch note, which of course are correctly related to that frequency. Then, by means of electrical filters which remove or accentuate certain bands of frequencies, excellent synthesis of most legitimate sounds can be realized. Such filters are of three fundamental types; low pass, which attenuate high harmonics; high pass, which reduce the lower frequencies; and resonant filters, which produce peaks at certain frequencies. Any or all of these can be combined, but many precautions are necessary to prevent one filter from interacting with others.

There is one further possible refinement of the electronic organ. In a pipe organ, each and every pipe is a complete tone generator in itself; that is, it produces its pitch note and all the related harmonics. Because of the many small differences in tuning, regulating, mass of the metal and harmonic content from pipe to pipe, chords played have a peculiar richness of sound. This is attributed to minute changes in phase of the different pipes sounding. It is a chorus effect and changes with every chord. In the electronic organ it is possible to obtain this elusive but characteristic effect if one oscillator is provided for every note. All these oscillators require tuning, but if one goes off tune no others are affected. This method is known as the free-phase system.

An alternative is the frequency divider method. Here there are 12 oscillators for the top octave. All other octaves downward are obtained from these oscillators by circuits which divide the frequency by two in each successive octave. Thus, if the top octave is accurately in tune, all other notes will be. It is only necessary to tune the 12 oscillators to tune the whole organ. All frequencies are locked in phase, so there is no chorus effect. If one oscillator fails, then all octaves deriving from this will fail. The wave forms from dividers are not ideal for tone forming. The circuity is, however, much less expensive than in the free-phase system, and therefore these instruments are cheaper. From an imitative point of view, since all oscillators and dividers must operate all the time, it is a matter of great difficulty to slow down the method of keying so that the tonal build-up resembles that of pipes. Most divider organs are unsatisfactory in this respect.

See A. Douglas, *The Electronic Musical Instrument Manual*, 3rd ed. (1957), *The Electrical Production of Music* (1957). (AL. D.)

ELECTRONICS, the branch of science and engineering concerned with the theory, design and use of devices utilizing electron emission or absorption, such as electron tubes, cathode-ray tubes, photoelectric cells, transistors and the circuits in which these devices are used. Electronics is best described in terms of its applications, but basic information is found in such articles as ELECTRON, which deals with the characteristics and behaviour of this elementary particle, and in ELECTRICITY; ELECTRICITY, CONDUCTION OF; PHOTOELECTRICITY; and THERMIONICS.

Applications of electronics in communications are described in RADIO, which deals with the fundamentals of radio transmission, while BROADCASTING; RADIO RECEIVER; TELEVISION and TELEPHONE are concerned with special aspects of communications. Examples of the use of electronics in industry can be found in COMPUTING MACHINES, ELECTRONIC, and in such research instruments as the electron microscope (*see* MICROSCOPE: *The Electron Microscope*) and the CATHODE-RAY OSCILLOSCOPE. ELECTRON TUBE and TRANSISTOR are devoted to the devices common to almost all electronic equipment.

See also references under "Electronics" in the Index volume.

ELECTRON PARAMAGNETIC RESONANCE. When paramagnetic substances are placed in a stationary magnetic field and subjected to electromagnetic radiation, effects are often observed that depend critically upon both the field strength and the radiation frequency (*see* MAGNETISM: *Paramagnetism*). A commonly observed effect is the absorption of the radiation by the paramagnetic substance at particular values of the frequency and the field strength. Such phenomena are caused by resonances between the frequencies of the electromagnetic waves and the natural frequencies of the electron motions in the substance. The absorption of energy and related phenomena that occur when these frequencies are equal or nearly equal are called electron paramagnetic resonance.

Electron paramagnetic resonance is commonly observed in beams of atoms or molecules, in high-vacuum systems and in solids and normal gases. Attention is here restricted to liquids and solids. Electron paramagnetic resonance in condensed phases (solids) was first observed by E. K. Zavoisky in 1945.

Electron paramagnetic resonance has proved to be an important research tool for the investigation of many physical and chemical problems involving paramagnetic substances. A few examples of its range of application are: the determination of the electronic structures of ions in crystals; the detection of paramagnetic free radical species that are involved in many chemical reaction mechanisms, both inorganic and organic; the study of the products formed when X-rays, electrons, gamma rays, neutrons and other radiations interact with many forms of matter; the measurement of properties of atomic nuclei in paramagnetic substances; and the investigation of the properties of semiconductors.

Paramagnetic substances contain atoms, ions or molecules whose electrons move with nonvanishing total angular momentum. Such electrons produce magnetic fields and may be likened to tiny bar magnets that tend to line up parallel to an external static magnetic field. Following the laws of classical mechanics, if they are disturbed from this equilibrium position they strive to realign themselves with the static field, and they thus exhibit an oscillatory motion whose frequency increases with increasing strength of the static field. This motion is a precession similar to that of a spinning top in the earth's gravitational field. When these elementary "magnets" are disturbed by the alternating magnetic field of an electromagnetic wave, the wave tends to drive them at its own frequency. When the frequency of the wave equals the frequency of their natural oscillatory motion in the static magnetic field, a large interaction between the wave and the paramagnetic system occurs, and considerable amounts of energy may be absorbed from the radiation.

Electron paramagnetic resonance may also be considered as a spectroscopic phenomenon from the standpoint of the quantum theory. Each elementary magnet (atom, ion or molecule) can point in only a discrete number of ways in an external static magnetic field. A different magnetic energy is associated with each direction, the energy differences increasing with increasing field strength. When the quantum of energy of an electromagnetic wave just matches the difference between two magnetic energies, a resonance absorption can occur in which the quantum flips the elementary magnet from one of its positions into another.

Both points of view show that the resonances occur at only certain ratios of static field strength to frequency. The quantum of energy of electromagnetic radiation is proportional to its frequency, ν, the proportionality constant being Planck's constant, h. The magnetic energy separations produced by the static field are proportional to the product of the magnetic moment (magnetic moment is a measure of the strength of the elementary magnets) and the field strength, H. The atomic unit of magnetic moment is the Bohr magneton, β. The resonance condition is therefore described by equating the quantum of electromagnetic wave energy to the magnetic energy separations in the paramagnetic substance, giving $h\nu = g\beta H$ or $\nu/H = g(\beta/h)$. In this relation g is a proportionality constant, usually referred to as the g-factor for the paramagnetic system. The values of the g-factor range widely, but commonly have the free-electron value, 2. This value of g, when used in the resonance relation, gives a value of 2.8 (mc. per second) per gauss for ν/H, the ratio of the frequency to field strength required for resonance. A magnetic field strength of 3,500 gauss, easily obtainable with an iron electromagnet, will therefore require a frequency of 10,000 mc. per second, corresponding to a wave length of 3 cm. Such microwave (radar) frequencies are frequently used in the investigation of paramagnetic systems.

When one of the variables, frequency or field strength, is changed and the other kept fixed, the measurement of radiation absorbed, as a function of the changing variable, gives an electron paramagnetic resonance spectrum. The apparatus used is referred to as an electron paramagnetic resonance spectrometer.

See also ATOMIC AND MOLECULAR BEAMS; NUCLEAR MAGNETIC RESONANCE; SOLID STATE PHYSICS. (C. A. HU.)

ELECTRON TUBE. This is the generic name for a class of devices which includes vacuum tubes or valves, phototubes or photoelectric tubes, thermionic tubes, gas-filled tubes, cathode-ray tubes and the like. It is an electrical circuit element consisting of a partially or wholly evacuated enclosure which contains two or more electrodes; its operation depends upon an electric current flow due to the motion of electrons in the space between the electrodes. An electron tube contains a source of electrons, known as the cathode; usually one or more metallic electrodes to which the current may flow, known as anodes; and often other electrodes for controlling or directing the current flow. Although X-ray tubes and gas-filled tubes that are intended only for illumination can be included under the definition, these types are discussed in ELECTRICITY: *Radio Tubes* and in LIGHTING.

Devices that perform functions similar to those of the electron tube, but are based on different physical principles, are discussed in the articles MASER and TRANSISTOR.

Electron tubes have been devised to change alternating to direct current, to amplify very small electrical voltages or currents, to generate power at frequencies up to about 60,000,000,000 cycles per second, to convert light into electrical currents, or to reverse this process so as to change electrical impulses into light. They have been widely used in communications and in the science which has come to depend on them, called electronics, because of their sensitive and almost instantaneously controlled response.

After an introduction, this article is divided into the following sections:

I. History
II. Emission of Electrons
III. General Theory
IV. Grid Control of Electron Streams in High Vacuum
V. Deflection Control of Electron Streams
VI. Velocity Control and Traveling-Wave Tubes
VII. Special Electronic Devices
VIII. Gas-Filled Tubes
IX. Photoelectric Devices
X. Technique and Construction

In addition to the laws of electricity, an understanding of electron tube behaviour requires consideration of the ultimate composition of matter into atoms, each having a nucleus and associated electrons. Insulating materials have electrons so closely bound to their nuclei that passage of a steady electric current is greatly retarded; in the electron tube, glass, ceramic or mica insulators are used and are chiefly of mechanical importance, serving as strong, electrically inert means for holding parts together. Conductors, usually metals, are of more basic importance, since their "free" electrons can move from atom to atom to permit passage of an electrical current. These electrons are also the source of the electron stream which leaves the cathode of the tube. At room temperatures, the average energy of these minute particles of negative electric charge is too small to permit many to overcome the surface forces of the material. As a result, they are bound within a conductor but, by the addition of energy from outside, or by other internal or external forces, may be released in substantial numbers. In an electron tube, the electrons are released into a vacuum or highly rarefied gas in order that their usefulness is not impaired by too numerous collisions. The methods of releasing electrons, in order of importance, are as follows:

1. Thermionic emission consists of electrons released by energy in the form of heat. It is used in most radio tubes.

2. Photoelectric emission is caused by exposure to light or other electromagnetic radiation of sufficiently short wave length. Although the energy of a light beam may be small and the electrons released comparatively few, this form of emission is of great importance in television and other electronic devices.

3. Collision of an electron with an isolated gas molecule may release an electron, leaving the molecule ionized with a positive charge. This mechanism is used in certain gas-filled electron tubes to increase the available current from one of the other types of emission.

4. Secondary emission is due to bombardment of a solid with charged particles, usually electrons, which may release one or more secondary electrons for each impinging particle.

5. Field emission is caused by the presence of such a strong electric field that the surface forces of the metal are overcome and electrons are released. A strong field may arise from a high potential difference at a distance or from a small potential difference over atomic distances such as in contact devices.

6. Radioactivity provides a source of high-energy electrons from nuclear disintegrations.

Control of the electron stream is possible by either an electric or a magnetic field. An electric field results from an applied voltage between electrodes in the tube. An electron, as a small negatively charged particle, is attracted and speeded up by the electric field from a positive electrode and is repelled and slowed down by the field from a negative electrode. Thus, the electric field can be used (1) to change the path of electrons; (2) to change the number of electrons flowing in a tube so as to change the current; or (3) to change the speed of electrons. All three effects are found useful in electron tubes.

Considering control by the magnetic field, a moving electron near a magnetized piece of iron, or near an electromagnet, is deflected and changes its path. The change is such that the electron tends to move both at right angles to its original path and at right angles to the magnetic field. This control permits either deflection of an electron beam or, when the path of the electron is nearly parallel to the magnetic field, a focusing effect similar to the focusing of light rays by a lens.

The above principles indicate how the functions of various electron tubes are performed. Changing an alternating to a direct current (or rectification, as it is called) requires a copious source of electrons, such as a heated thermionic cathode, and a nearby anode; these elements form a two-electrode tube, or diode, whose parts are shown in fig. 1. When these are sealed into an envelope which is properly evacuated, and an alternating voltage is applied between cathode and anode, the electric field periodically reverses; and current flows only when the anode is positive. The electron current is then a pulsating direct current. If this current is passed through a load resistor, as in the schematic diagram, and the pulsations smoothed out by a capacitor, a smooth direct current flows through the load.

The current passed by such a tube is usually less than the full emission of the cathode because of the repelling force of electrons on each other. The current may be increased, at the expense of rapidity of control, by the addition of a small amount of gas to produce considerable ionization.

In the simplest thermionic tube amplifier, a metallic grid is interposed between the cathode and the anode, so as to be closer to the cathode. This is now a three-electrode tube, or triode,

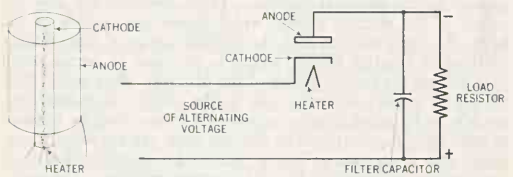


FIG. 1.—INTERNAL PARTS AND SCHEMATIC RECTIFIER CIRCUIT OF THERMIONIC DIODE

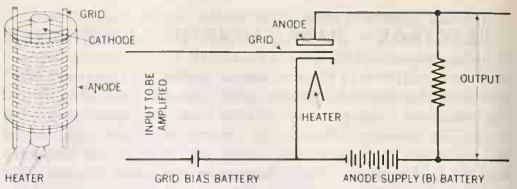


FIG. 2.—INTERNAL PARTS AND SCHEMATIC AMPLIFIER CIRCUIT OF THERMIONIC TRIODE

shown in fig. 2. If a control voltage is connected between grid and cathode, because of their proximity, a strong electric field is produced. At the same time, the grid reduces the field because of the anode-to-cathode voltage. The current flow is, therefore, strongly dependent on the control-grid voltage and weakly dependent on the anode voltage. Furthermore, by use of a bias battery whereby the grid is held at a negative voltage with respect to the cathode, the electrons cannot strike it but, with a sufficient positive anode voltage, will penetrate the grid and reach the anode. In this way no control power is required, but a small control voltage alters the current to a high-voltage anode, producing substantial available output power. Addition of a second grid, called a screen grid, at a positive voltage between control grid and anode reduces the already weak dependence of the electron current flow on anode voltage and so increases the available output power and the amplification. It is often desirable in such tubes to prevent the loss of electrons from the anode, because of secondary emission from it, by using a third grid adjacent to the anode and operated at cathode potential. Such five-electrode tubes are called pentodes. Triode or pentode amplifiers may be made to generate oscillations by feeding back part of the amplified output into the control circuit.

Other thermionic tubes make use of the controlled deflection of a beam of electrons to different positions on an anode screen which luminesces when struck, so as to emit light. These tubes are called cathode-ray tubes or oscillograph tubes when used for measurements, but are also used in a slightly different form to portray a television picture (sometimes called kinescopes). (*See* TELEVISION and RADAR.) Still other tubes, known as velocity-modulation tubes, or klystrons, make use of controlled speed of a stream of electrons and are particularly useful at extremely high frequencies. In special tubes known as magnetrons, a combination of high-frequency electric control fields with a steady magnetic field causes complex electron motions; these tubes are also of great importance in generating extremely high frequencies. Magnetic fields, variable or constant, are also used to deflect or focus beams in television picture tubes.

The photoelectric tube, preferably called the phototube, in its simplest form consists of a cold cathode in a position such that light may reach it, and an anode. A positive voltage between anode and cathode provides a field to support the flow of electrons, if they are released from the cathode by the light. In darkness, substantially no current flows, the photoelectric emission being proportional to the light intensity. For television pickup, more complex tubes are needed in which photoelectric emission effects are combined with a deflectable thermionic electron beam to provide a signal which, at any instant, depends only on the light on a single small spot of the photocathode. Other tubes, known as image tubes, use an electric or magnetic field to focus the photoelectrons onto a luminescent screen; in this way a radiation image of one wave length (such as infrared) may be converted to a radiation image at another wave length (such as in the visible range).

I. HISTORY

The development of the electron tube started from physical investigations of thermionic and photoelectric emission. Frederick Guthrie in 1873 observed the discharge of an electroscope by a nearby heated metal ball, now known to be a thermionic emitter. In the 1880s, Julius Elster and Hans F. Geitel and, independently,

Thomas A. Edison, observed electric currents in an evacuated enclosure passing in one direction from a hot filament to a cold metallic plate, as if negative electrical particles were being emitted by the filament. Heinrich R. Hertz (1887) was the first to notice photoelectric emission, by observing the effect of light on an electric spark. Wilhelm Hallwachs (1888) and A. G. Stoletow (1890) showed that the photoelectric effect was also an emission of negatively charged particles. After Sir J. J. Thomson in 1897 had shown that the particles in these experiments were indeed electrons, of a definite electric charge and mass, irrespective of their origin or the methods for their release, the fundamental principles later used in electron tubes were established.

The aforementioned experiment by Edison led J. A. Fleming, in the first decade of the 20th century, to use such a tube as a means of changing the small alternating currents received by radio to direct currents capable of actuating a meter or telephone receiver. Then known as a Fleming valve, the device was later called a diode. The most important advance in the art of thermionic tubes was made by Lee De Forest, about 1906. He introduced a third electrode, in the form of a grid, between cathode and anode, thus making the first amplifier, which he called the Audion (later called a triode). Within a few years this triode made possible detection of weaker radio signals than ever before and enabled amplifiers to be built for long-distance telephony. Between 1910 and 1915, a number of independent workers (including E. H. Armstrong, De Forest himself, A. Meissner in Germany, and C. S. Franklin and H. J. Round in England) found special advantages in a feedback circuit which produced either a regenerative action or, if the feedback was sufficient, a self-oscillatory condition. In the latter case, the thermionic tube efficiently transformed the steady power, supplied by a battery, to an alternating power of frequency considerably higher than possible with rotating machinery. The thermionic tube became a great stimulus to the development of the higher radio frequencies for communications and radio telephony.

World War I (1914–1918) provided a great stimulus to radio, and triode tubes were made in large quantities for the first time. The improved high-vacuum techniques of Irving Langmuir showed the disadvantages of residual gas and enabled tubes to be made which matched theoretical predictions. High-power triodes were used in the first successful transatlantic radio telephony from the U.S. to France (1915) and the availability of low-power types made practicable the superheterodyne method of reception, invented independently in Germany and the U.S. Walter Schottky, in Germany, also studied extensively the utility of an additional grid, forming a four-electrode tube or tetrode.

Undoubtedly the greatest advances in thermionic tube technology occurred because of the introduction of sound broadcasting about 1922. Within the next two decades home radio receivers came into widespread use. In the U.S., for example, tube production rose from about 1,000,000 in 1922 to an annual average of about 100,000,000 tubes between 1936 and 1941. At the same time the price of the typical receiving tube fell to about one-sixth of the 1922 price, while its life expectancy increased by several times. Tetrodes, pentodes and tubes with even more grids were introduced in the early 1930s. Research and development were actively supported and hundreds of different tube variations were devised in the U.S. and Europe. Perhaps the most significant advances were made in special tubes for the generation and detection of radio frequencies above 30 mc. (*i.e.*, below 10 m. in wave length). The use of these high frequencies increased the possible speed of transmission of intelligence to permit television and, in the late 1930s, led to applications of radio to air navigation, distance determination and object location (radar). Cathode-ray tubes, useful for converting electrical impulses into visible light, were by this time widely used for television and in the laboratory. Between 1920 and 1940, new phototubes and other photoelectric devices were also commercialized under the stimulus of sound motion pictures and television; and the science of electronics became an important one.

The electron tube industry, grown large in peacetime, became tremendous in World War II (1939–1945). The increased tempo of military activity required highly specialized communications and electronic equipment, all dependent on the electron tube. In the U.S., the tube industry expanded from an annual business of around $40,000,000 to one of $400,000,000. Because military tubes were relatively expensive the annual number of tubes produced did not increase in the same proportion, although there was an increase of about 60%. This war led to exploitation of grid-controlled tubes for the ultrahigh-frequency (10 to 100 cm. wave length) region and to development in Great Britain and the U.S. of magnetrons and klystrons for the superhigh-frequency (1 to 10 cm. wave length) region. These previously purely experimental tubes were greatly improved, were made in mass production by 1943, and led to extremely effective military radar equipment. They also led to development of sharply beamed radio relay systems destined to play an important part in peacetime communications. Advances in the commoner thermionic tubes for lower frequencies were also significant. Tubes were developed for high-power television transmission; receiving tubes were reduced in size and, along with cathode-ray tubes, were improved in performance and better adapted for mass production. Even tubes employing photoelectric phenomena shared in the advances; infrared-sensitive image tubes and supersensitive television pickup tubes were developed for war applications.

Fortunately, many of the advances in tube technology were useful in peacetime, and the decade following World War II saw their application to television broadcasting, radio relays and many other communication and industrial projects. At the same time, continued government sponsorship of new military equipment resulted in better tubes for radar and ever shorter microwaves. Electronic computers, using large numbers of tubes, also became a reality. In the U.S. widespread sales of television receivers stimulated tube production to several times prewar volume, and the electronics industry in the mid-1950s became one of the top ten industries. Nevertheless, there were few revolutionary changes in the development of electron tubes; perhaps the most notable were the traveling-wave microwave tube (1946) and the colour-television picture reproducer tube using cathode-ray principles (1950). On the other hand, the invention in 1948 of the transistor (*q.v.*), a nonvacuum electron device which fulfills functions similar to those of the electron tube, introduced a new era based on advances in solid-state physics. Within a few years, production of transistors had become a substantial fraction of that of vacuum electron tubes, and a number of other competitors to the electron tube were emerging from the research laboratory.

II. EMISSION OF ELECTRONS

The phenomena of emission are given separate treatment (*see* THERMIONICS) and so will be reviewed only briefly here. Thermionic emission, or the release of electrons from within a metal by virtue of their heat energy, has been qualitatively understood only since 1928 by application of the wave-mechanical theory of electrons in metals. On purely thermodynamic grounds, however, O. W. Richardson by 1921 had derived an equation of sufficient validity to be widely used. His result indicated the total emission from a unit area of a metal, at a temperature T, would be

$$I_s = AT^2\epsilon^{\frac{e\varphi}{kT}} = AT^2\epsilon^{-11,600\varphi/T} \qquad (1)$$

where ϵ is the base of the Napierian logarithms (2.718), k is Boltzmann's constant (1.38×10^{-23} watt-seconds per °K.), e is the electron charge ($e = -1.60 \times 10^{-19}$ coulombs), and A and φ are constants which depend on the metal and its surface conditions. In the equation, $e\varphi$ represents the energy an electron would lose in passing through a potential difference φ, and, in the derivation, represents the energy needed to extract an electron at zero temperature. Such an energy is often expressed in electron volts (ev), defined as the gain or loss of energy of one electron passing through a potential difference of one volt (thus 1 ev = 1.60×10^{-19} watt-seconds). In the Richardson equation, the voltage φ is called the thermionic work function of the emitting material and is the most important factor determining the emission at a given temperature. Its value ranges from as low as one volt, for the best

composite-surface emitters, to perhaps six volts, for poorer emitters. The constant A is theoretically about 120 amp. per square centimetre per $(°K.)^2$ but actually varies over many orders of magnitude for composite surfaces.

The most striking characteristic of thermionic emission as observed and predicted by equation (1) is its strong dependence on temperature. In many thermionic tubes the electron current is limited by space charge (see *General Theory*, below) to values of only a small fraction of the available emission, so that the temperature of the emitter is not critical as long as it is sufficiently high. Pure metals have relatively high values of φ so that high-operating temperatures are needed (*e.g.*, 2,500° K.) and only high-melting-point metals (tungsten or tantalum) are usable. The oxide-coated emitter, which operates at as low as 1,000° K., consists of a mixture of barium and strontium carbonates on a base metal such as nickel; in processing, the carbonates are heated to decompose them into the oxide and carbon dioxide. The problems of such low work-function emitters are not yet fully understood and their wide practical use has been made possible by application of empirical knowledge.

Furthermore, it was often demonstrated that the emitting properties of composite-surface emitters may be quite different when the current is drawn in short pulses as against a steady current; higher emissions are often possible with the short pulses. Because of the importance of cathode heating power, cathodes are often compared for emission on this basis. When special curved coordinates are used, as in fig. 3, various emitters give straight-line plots and can quickly be compared for relative efficiencies. Wide variations are found in oxide-coated emitters of the same type, and the curves on fig. 3 are representative of commercially processed emitters.

In photoelectric emission, the currents are directly proportional to the light intensity and are usually small compared with those produced by thermionic emission. The velocities of the emitted electrons can be considerably higher than with thermionic emission; the emitted electron energies depend on the frequency of the light but are independent of its intensity. It was the prediction by Albert Einstein of these facts which was of such great importance in establishing the quantum theory. By the quantum hypothesis, light energy is absorbed only in discrete quanta of size $h\nu$ where h is Planck's constant (6.62×10^{-27} erg $\times$ sec.) and ν is the frequency. Since the energy needed to extract an electron is of the order of $e\varphi$ (as with thermionic emission), it is clear that almost no electrons will be emitted if the light has a frequency less than ν_0, given by

$$h\nu_0 = e\varphi \qquad (2)$$

If the light quanta of energy are greater than this (*i.e.*, $\nu > \nu_0$), an excess of energy is available to each emitted electron so that it may be emitted with an energy as high as

$$\tfrac{1}{2} mv^2 = h\nu - e\varphi \text{ watt-seconds} \qquad (3)$$

where m is the electron mass in kilograms and v its ejection velocity in metres per second. This is Einstein's famous equation, postulated in 1905, and later experimentally confirmed. Although a low value of work function φ is necessary for emission by long wave length light, the emission current is affected by other considerations as well. The maximum possible photoelectric emission would result if each light quantum ejected an electron; yields up to 25% of this theoretical maximum can be obtained in photoelectric tubes. The most sensitive materials are semiconductors consisting of complex cesium compounds. The spectral response of some useful photocathodes is shown in fig. 4. Pure metals are so low in response that on the scale of the figure they lie nearly on the axis of abscissas.

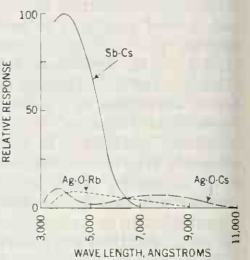


FIG. 4.—SPECTRAL SENSITIVITIES OF COMPOSITE PHOTOELECTRIC CATHODES IN LIME GLASS BULB

Since electron release through ionization of gas molecules releases both a positive and negative charge in space, it is not, in the strictest sense, a phenomenon of electron emission in itself. Its importance in the gas discharge, however, makes it worthy of mention.

Secondary emission is the term used to describe the release of electrons from solids, usually by impinging primary electrons. It is not always possible to distinguish reflected primary electrons from true secondary electrons, since both are present. A few of the primary electrons are reflected with their full energy in all cases and, when the incident primaries strike at energies below about ten electron volts (*i.e.*, 10 ev), this group is predominant. At higher primary energies a large number of the released electrons are true secondaries having release energies corresponding to only a few electron volts. For many clean metal surfaces, the secondary emission ratio (ratio of secondaries to incident primaries) increases with increasing primary velocities up to a maximum of between one and two at around 500–1,000 ev primary energy. At higher primary velocities, the ratio decreases slowly.

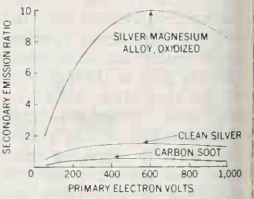


FIG. 5.—SECONDARY EMISSION PERFORMANCE. THE CLEAN SILVER CURVE IS TYPICAL OF MANY PURE METALS

Contaminated surfaces behave similarly except that the maximum ratio is much higher and, in specially prepared composite surfaces, may reach ten or more. The effects of relatively small amounts of oxidation or other contaminating surface conditions are so marked that for many years poor correlation was found between results of different observers. Fig. 5 shows typical results for several materials. Secondary emission from insulators also has been observed, and is qualitatively similar to that from metals except, of course, that steady currents cannot be drawn. Bombardment of surfaces by positive ions or metastable atoms will also release secondary electrons; this phenomenon is of great importance in gas discharges between cold electrodes. Although the phenomena of secondary emission are predictable from wave mechanics, the basic theory for secondary emission from the composite type of surface is even less well understood than that for

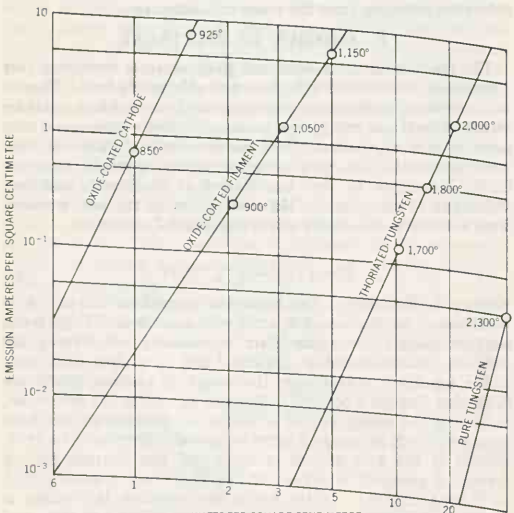


FIG. 3.—CHART SHOWING THERMIONIC EMISSION AS A FUNCTION OF CATHODE HEATING POWER. TEMPERATURES ARE IN DEGREES KELVIN

thermionic and photoelectric emission, and progress has been almost entirely experimental.

Field emission may be thought of partially as a lowering of the work function by the action of a strong external electric field, and partially as a "tunneling-through" the potential barrier by electrons, as calculated from wave mechanics. It is, therefore, closely allied to both thermionic and photoelectric emission. The reduction of work function can be approximated by subtracting from the normal work function an amount which varies as the square root of the external field; the "tunneling-through" gives currents which occur only at the very highest fields, and which vary more rapidly with field. Quantitatively, the field must be of the order of 10^9 v. per metre for appreciable field emission, so that this phenomenon is not present in most electron tubes.

Electron release through radioactive decay or other nuclear process is entirely unrelated to the other types of emission, and is not used in electron tubes.

III. GENERAL THEORY

An electron in a vacuum is subject to the electrical force determined from Coulomb's law that the repelling force between two like charges is proportional to their product and varies inversely as the square of their distance apart. By defining the electric field intensity as the force per unit charge on an infinitesimally small charge, the force on an electron is equal to the product $e\mathcal{E}$ where e is the electronic charge and $\mathcal{E}$ is the electric field intensity. This force has the opposite direction to the electric field because e is negative, and both the force and the field can be considered vectors (*i.e.*, quantities having both magnitude and direction). In most electron tubes, the electronic mass can be considered as if it were a constant, a relativistic correction being negligible for the velocities ordinarily encountered. Thus one may predict the motion of the electron using Newtonian mechanics, the force being equal to the product of mass and acceleration. Mathematically this gives the differential equation

$$m\frac{dv}{dt} = e\mathcal{E} \tag{4}$$

where v is the electron velocity in metres per second in the direction of $\mathcal{E}$ and m is the electron mass (9.107×10^{-31} kg.). $\mathcal{E}$ is in volts per metre, t is in seconds and $e = -1.60 \times 10^{-19}$ coulombs. Following convention, $\mathcal{E}$ is related to the voltage dE, between two points ds apart, by $\mathcal{E} = -\frac{dE}{ds}$.

In a magnetic field, the electron experiences a second force, but only when it is in motion. This force can be derived from Ampère's law and may be thought of as similar to that which drives an electric motor, since an electron in motion represents a flow of current analogous to the current in a motor armature winding, which is subject to a sidewise force. The force is proportional to the charge, to the velocity component perpendicular to the field, and to the magnetic flux density. The force is directed perpendicular to both the field and the electron velocity; *i.e.*, if we represent the magnetic flux density B and the electron velocity v as vectors in space, the force is a vector perpendicular to the plane containing B and v. Thus an electron traveling at right angles to a uniform magnetic field is deflected in the direction perpendicular to both. An electron traveling parallel to a uniform magnetic field, on the other hand, is unaffected; but any departure from parallelism gives rise to a perpendicular component of velocity and thus a force. This force gives the nearly parallel electron a spiral motion about the direction of magnetic field, keeping it from diverging far from the parallel path. Mathematically, the equation of motion in any of these cases is

$$m\frac{dv_p}{dt} = Bev \sin\theta \tag{5}$$

where v_p is the velocity of the electron in metres per second in the perpendicular direction to the plane of B and v, and θ is the angle between the directions of B and v; e is in coulombs and B is expressed in webers per square metre (1 weber per square metre $= 10^4$ gauss $= 10^4$ maxwells per square centimetre $= 10^7/4\pi$ amp. per metre).

Equations (4) and (5) would be sufficient to solve for the path and time of transit of electrons in an electron tube except that they require $\mathcal{E}$ and B to be known, and these may depend, in turn, on the presence of electrons or ions. The currents in electron tubes are small enough so that their effect on the magnetic field, B, is usually negligible. The cumulative effect of the electron or ion charge (called space charge) on the electric field cannot always be neglected, however, and this introduces a serious computational difficulty unless the geometry is extremely simple. Furthermore, the electrode currents are so dependent on space charge that the characteristics of the electron tube in a circuit are thereby largely determined. The electric field with or without space charge is given by an application of Gauss's theorem of electrostatics and leads directly to the partial differential equation of Poisson which, in Cartesian co-ordinates, is

$$\frac{\partial\mathcal{E}_x}{\partial x} + \frac{\partial\mathcal{E}_y}{\partial y} + \frac{\partial\mathcal{E}_z}{\partial z} = \frac{\rho}{\Delta} \tag{6}$$

where ρ is the electric charge density in coulombs per cubic metre; $\mathcal{E}_x$, $\mathcal{E}_y$ and $\mathcal{E}_z$ are the three electric field components in volts per metre; and $\Delta = 8.85 \times 10^{-12}$ farads per metre.

The current entering any surface, such as that of an electrode in the tube, is the time rate of change of charge at that surface. This current is the sum of two components, one representing the actual arrival of electrons at the surface or electrode (known as the convection current) and the other due to the change of induced charge by any change of electric field with time (known as the displacement current). Thus the current per unit area i is found from the charge density ρ, the electron velocity v and the field $\mathcal{E}$ as follows:

$$i = \rho v + \Delta\frac{\partial\mathcal{E}}{\partial t} \tag{7}$$

At low frequencies of operation or under steady (static) conditions, the second term (displacement current) is not of importance. The contrary is true at ultrahigh frequencies and above. Equations (4) to (7), inclusive, are fundamental to all high-vacuum electron tube problems and sufficient for their theoretical solution. Unfortunately, it is only rarely that they can be completely solved in practice.

When space charge is negligible (*i.e.*, when currents are very small) the path taken by an electron in a vacuum tube has a very simple gravitational analogy. This analogy is essentially a solution of equations (4) and (6) under special conditions. It can be shown that a tightly stretched rubber membrane, which is displaced from normal, obeys equation (6) above, if ρ is put equal to zero and if the field can be expressed in two dimensions. The slope of the membrane corresponds to the field, and the height of each point to the electric potential. Pieces of wood or metal are shaped like the cross section of the electrodes of a tube, and made to a large scale. When these are placed under and above the rubber, so that each piece displaces it by a distance proportional to the electric potential, with positive potentials downward, one obtains a contour corresponding to the potentials in the actual tube. A small steel ball, released gently from the edge of the piece corresponding to the cathode (which is displaced upward), will then roll down the rubber membrane and follow, on the model, the path taken by an electron in the tube. The rolling ball model provides a simple physical analogy to the electron ballistics.

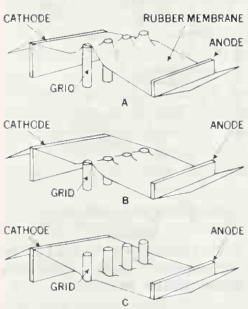


FIG. 6.—RUBBER MEMBRANE MODEL OF PARALLEL-PLANE TRIODE. ELECTRON PATHS ARE SIMILAR TO THOSE OF SMALL STEEL BALLS ROLLING ON MEMBRANE. (A) GRID NEGATIVE (RODS HIGHER THAN CATHODE); (B) GRID SLIGHTLY POSITIVE; (C) GRID MORE POSITIVE

Thus, an attracting (negative) electric field is equivalent to a

downhill slope while the repelling (positive) field is an uphill slope. If a hilltop is higher than the cathode, it corresponds in the tube to a negative voltage and, of course, the ball cannot reach it. An example of the rubber-membrane model is shown in fig. 6 in which a triode with three different grid voltages is modeled. Convergent focusing by the valleys between the grid wires would be found in two of the cases, while a diverging action occurs in the case of the most positive grid. The rubber-membrane model is not suited for showing the forces attributable to a magnetic field; it also does not show the effect of space charge and so does not permit current flow to be evaluated. Nevertheless, it is a useful tool for rapid solution of two-dimensional trajectory problems which are too difficult to solve mathematically.

When current flow must be evaluated and space charge is important, one is restricted to rather simple geometries for treatment. The simplest solution applies to a highly evacuated electron tube with parallel-plane geometry. In this case, the electric field has only one component, normal to the electrodes. If one surface is an electron emitter (cathode), releasing an indefinitely large number of electrons at zero emission velocity, equations (4), (6) and (7) are readily solved to give the Child-Langmuir 3/2-power law for the convection current per unit area in a diode

$$I_0 = \frac{4\Delta}{9}\sqrt{\frac{2e}{m}}\frac{E_b^{3/2}}{d^2} = (2.33 \times 10^{-6})\frac{E_b^{3/2}}{d^2} \qquad (8)$$

where E_b is the potential difference between the two electrodes in volts and d is their distance apart. If d is in centimetres then I_0 is in amperes per square centimetre, or if d is in metres then I_0 is in amperes per square metre. The physics of the solution shows that, if the cathode emission is gradually increased, when the limiting current I_0 is reached, the repelling forces of the electrons are just sufficient to reduce the field at the cathode to zero. Thus, no more current can flow than this value. Langmuir showed that the 3/2-power relationship holds for any two-electrode tube geometry. An exact solution for the proportionality constant (called perveance) is not, in general, possible. An important result of the 3/2-power law is the implied nonlinearity of all such electron tube devices; they do not behave like ordinary resistors in which the current is proportional to the first power of the voltage. It is also clear that the current may be limited by space charge rather than by the ability of the emitter to release electrons. In all but the highest-power transmitting tubes, the space-charge-limited currents are but a very small fraction of the available emission.

The Child-Langmuir relation is an approximation, particularly at low anode voltages, because it neglects the electron emission velocities, which allow a small current to flow even at zero anode voltage. A correction can be applied to E_b and d which gives a more accurate result. When the thermionic emitter is a filament, heated by passing current through it, the voltage drop along it also must be considered in applying theoretical considerations to the device.

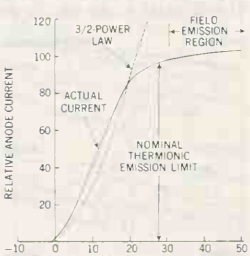


FIG. 7.—CHARACTERISTIC OF THERMIONIC DIODE ASSUMING OXIDE-COATED EMITTER

Fig. 7 shows a typical characteristic of anode voltage v. anode current for a parallel-plane diode with an oxide-coated cathode heated by a separate heater (equipotential cathode). As the anode voltage is increased, the current rises until it approaches the full thermionic emission of the cathode. A further increase of anode voltage causes only a small rise in current because of field emission. Except in the highest-power transmitting tubes, this region of the characteristic is never reached and, to all practical purposes, the emission is much higher than the normal currents drawn. In measuring diodes such as that used for fig. 7, it is often found that the curve and its intercept with the voltage axis are shifted to the right or left of that shown. Such a shift is caused by a small inherent potential difference between cathode and anode, which exists even if no voltage is applied. Called contact potential, the effect is caused by a difference in thermionic work function of cathode and anode; it may amount to a volt or two and can vary considerably with minute surface contamination.

Fig. 7 also shows the theoretical 3/2-power law (equation [8]). It is seen that, in comparison with the true curve, it gives too small a current at the lower voltages and does not apply above the emission limit. Since the theoretical curve can be shifted to the right or left by subtracting or adding a small constant voltage to E_b of equation (8), this may be done to approximate initial velocity and contact-potential effects.

The extension of equation (8) to tubes with more than two electrodes is done by replacing the combination of electrodes beyond the cathode by an equivalent single electrode. This will be discussed in more detail below. It should be noted that solution (8) is for the convection current only; *i.e.*, the static or D.C. part of the current. This is a sufficiently close approximation as long as the time of transit of the electrons is short compared with the time taken for a change in the potentials. The displacement currents are then just those of the capacitances between the various electrodes, modified slightly by the presence of the electron stream. When the time of transit is not negligible, as at frequencies above about ten megacycles per second for many tubes, the time lag between the applied voltage change and the resulting current introduces modifications, some of which will also be discussed below.

IV. GRID CONTROL OF ELECTRON STREAMS IN HIGH VACUUM

The current in an electron stream flowing in a high vacuum may be varied by a voltage by making use of space-charge control. The simple two-electrode (diode) device, whose characteristic is approximated by the Child-Langmuir relation (8) above, is suitable as a rectifier. An amplifier, however, operates by using a small input to control a larger output, so that some electrode arrangement is needed in which too much current does not flow to the input electrode. This may be done for a space-charge-controlled tube by the introduction near the cathode of a permeable electrode, or grid, which controls the electric field in the tube and hence the current. The effect of the grid on the electric field depends on its potential and is very easily understood from the rubber-membrane model of fig. 6. With only cathode and anode present, the rubber-membrane equivalent forms a slanting hill, down which the steel balls (electrons) roll from cathode to anode. If the grid structure is at a positive potential which is just equivalent, in the model, to the height of the hill at the grid location, the grid rods of the model will just touch the rubber; so that the electron paths and the current are unchanged except that the grid now catches a proportion of the total current, depending on the size of its wires. If the grid is more positive (in the model), the rubber is depressed by each wire, so that the hill at the cathode is steeper (corresponding to a stronger electric field) and the small individual depressions in the rubber around each grid wire cause balls rolling nearby to be deflected slightly toward the wire, so as to increase the proportional current to the grid. When the grid is negative in potential (in the model) the rubber is pushed up at the grid region so that the hill at the cathode is less steep and the individual grid wires protrude out of the membrane as small hills, higher than the cathode. The balls (electrons) now cannot strike the grid but pass through between the grid rods and roll down to reach the anode. The model shows that a small change in grid voltage (height on the model) will cause a marked change in the cathode field (steepness of the hill). Space charge in the actual tube changes the potential distribution somewhat from that of the rubber-membrane model, but primarily near the cathode; so that one may actually approximate the effect of the grid by thinking of an equivalent or effective potential which is an average between the actual potential of the wires and the region between the wires, as determined by heights on the rubber-membrane model or by calculation from electrostatics.

Since most of the space-charge forces lie close to the cathode, an approximation to the tube characteristic is given by finding the effective voltage at the grid and applying the Child-Langmuir

relation (8) to the cathode-to-grid region. The effective voltage between grid and cathode is approximated by

$$E_{eff} = \beta\left(E_c + \frac{E_b}{\mu} + \delta\right) \tag{9}$$

where E_c is the grid-cathode voltage, E_b is the anode-cathode voltage, and δ is a small correction voltage for contact potentials. The constants β and μ depend on the tube geometry, although β is usually a little less than one and varies only slightly with changes in geometry. Using the 3/2-power law, therefore, an approximation to the cathode current I_a from a triode is

$$\begin{aligned} I_a &= KE_{eff}^{3/2} \\ &= K\beta^{3/2}\left(E_c + \frac{E_b}{\mu} + \delta\right)^{3/2} \end{aligned} \tag{10}$$

where K is the proportionality constant called perveance. The perveance is usually inversely proportional to the square of the cathode-to-grid distance. When the grid is negative, the anode and cathode currents are identical and (10) applies to the anode current I_b as well. Examination of (9) and (10) shows that a change in grid voltage is μ times as effective in changing current as a change in plate voltage, and μ has been called the amplification factor. Furthermore, it is seen that when $E_c = -\left(\frac{E_b}{\mu} + \delta\right)$ the current is zero; this condition is known as cutoff. Many theoretical calculations have been derived for μ for parallel-plane and cylindrical geometry, but a simple empirical one gives results within a few per cent of the more exact ones for most practical values. This empirical formula applies to a grid consisting of parallel wires of diameter b spaced uniformly N per unit distance, and is

$$\mu = Ns\left[1.25 + 42.7Nb + 4{,}270(Nb)^6\right] \tag{11}$$

where s is the grid-to-anode spacing for parallel-plane geometry. For cylindrical geometry,

$$s = r_g \log_e \frac{r_p}{r_g} \tag{12}$$

where r_p and r_g are anode and grid radii, respectively. In an actual triode with geometry intermediate between plane and cylindrical, the s to be used is also intermediate. It is also found that μ may vary somewhat with the operating voltages, but an average value is easily measured by the relative change in anode voltage required to offset a known change in grid voltage. A more exact formulation than equation (10) for the cathode current of a tube was given by G. Liebmann (1946).

A typical set of characteristic curves for a triode is shown in fig. 8. The curves at the left show anode current as a function of control-grid voltage. When the voltage drop in the anode load resistance or impedance is appreciable, the effect of anode voltage must be included and the right-hand set of curves is more useful. For example, if an anode load resistance of 10,000 ohms is used, with an anode voltage of 300 v., one may draw line B to show that, as the anode current rises, the voltage on the tube drops, along line B. The effect can be shown on the curves at the left by characteristic A, which shows the effect of progressively dropping anode voltage. A quantity of great importance in tubes for small-signal amplification is the sensitivity of anode current to small changes in grid voltage. This quantity, known as the transconductance, is given by

$$g_m = \frac{\partial I_b}{\partial E_c} \approx \frac{3}{2} KE_{eff}^{1/2} \tag{13}$$

The transconductance is made high in low-power tubes by a high value of the perveance K, since large values of E_{eff} lead to excessive current. The transconductance is easily measured by the change in anode current for a small change in grid voltage. In fig. 8, it is given by the slope of the left-hand set of curves.

The triode has a serious disadvantage because of the reaction of the output anode voltage when an output circuit is connected to the tube. When the control grid is made more positive, the anode voltage drops because of the increased current and the greater voltage drop in the output circuit, as illustrated by line B and characteristic A of fig. 8. Thus, E_{eff} is not increased by the full change in E_c, since E_b has also changed in an opposite direction. This has the effect of decreasing the "effective" transconductance from the value given by (13) and of limiting the maximum output current in power tubes. The decrease in "effective" transconductance is clearly shown in fig. 8 by the decreased slope of A. By building into the tube a second grid called the screen grid, between the control grid and anode, and operating this grid at a fixed positive voltage comparable with that of the anode, the effect on E_{eff} of a change in anode voltage is greatly reduced and the tube is thereby greatly improved. A small loss in anode current, because of absorption by the second grid, is encountered, however. The tube characteristic is still given by an expression of the form of (10) except that E_b is now replaced by the screen-grid voltage. Furthermore, in a calculation of μ for insertion in (10), the control-grid to screen-grid distance is used for s in equations (11) and (12). The measured value of the change in anode voltage to offset a given change in control-grid voltage now has little significance and may approach infinity. It will be noted that the screen grid is not used as a control on the electron stream but only as a fixed potential screen which allows the major portion of the current to pass through. Although the construction is more complex, the screen-grid tube is simpler to understand in its circuit operation because of the substantially complete independence of the anode current on anode voltage. It is to be noted that certain screen-grid tube types, known as variable-μ tubes, have a variable-pitch control grid which leads to an advantageous and smoother control of amplification. With such tubes, equation (10) does not apply except when μ is made a function of the applied voltage.

Screen-grid tubes are particularly advantageous in high-frequency amplifiers because the capacitance between control grid and anode is greatly reduced by the shielding or screening action of the interposed grid. This prevents feedback effects which could otherwise cause instability or oscillation.

In practical screen-grid tubes means must be employed to minimize secondary electron emission from the anode when it is lower than the screen grid in potential. The means most commonly employed is a third, coarse-meshed suppressor grid adjacent to the anode and operated at a low voltage (it is usually connected to the cathode). Such a grid has an effective potential below that of the anode but well above the cathode. It can be designed to interpose no obstacle to the cathode-to-anode electron flow but to prevent the relatively low-velocity secondary electrons from leaving the anode. The three-grid tubes, known as pentodes, are most commonly used in high-frequency amplifiers and in power amplifiers. Alternative means, using special anodes or making use of space-charge effects in electron beams, have also been used singly and in combination with suppressor grids, as in the type called beam power tube.

The characteristic curves of a pentode of the type used for power outputs of a few watts are shown in fig. 9. The curves for all pentodes are qualitatively similar. In distinction to the triode curves of fig. 8, it is now seen that, once the anode voltage is above 50 v., it has a negligible effect on the current. The left-hand set of curves is, therefore, almost independent of anode load.

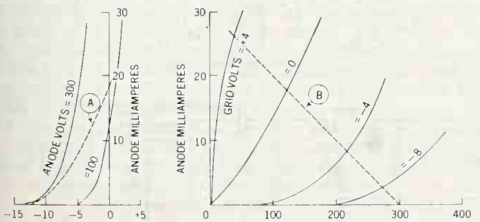


FIG. 8.—TRIODE CHARACTERISTIC CURVES. A AND B SHOW EFFECT OF 10,000 OHM ANODE LOAD RESISTANCE

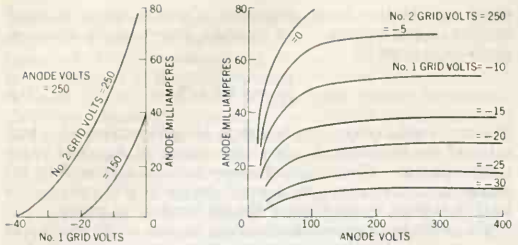


FIG. 9.—PENTODE CHARACTERISTIC CURVES

Instead, these curves depend only on screen-grid (no. 2 grid) voltage which, in normal operation, is held constant. The pentode is thus free of the disadvantages of the triode.

Since the control of electron flow by a grid is possible by interposition of the grid at any point in the electron stream, the suppressor grid may be used as a second control electrode. Alternatively a special grid between a first screen grid and a second screen grid may be used as a control. In either of these cases, the action is quite different from the space-charge control typified by equation (10). Instead, the anode current is controlled by a process of division, with the control grid acting as a gate to allow the current flow either to reach the anode or to be reflected back toward the first screen grid; *i.e.*, dividing the current between screen grid and anode. Tubes of the kind with two control grids and two screen grids are called hexodes and, when the fifth grid is added as a suppressor, are called heptodes. These tubes are used in superheterodyne receivers as frequency changers or converters, where they are sometimes called pentagrid converters.

The variety of circuits with which grid-controlled electron tubes can be used is so great as to defy enumeration. A few illustrative circuits are shown in fig. 10 and are typical of those used in radio receivers. For simplicity, the heater circuits of the thermionic cathodes are not shown, and the D.C. anode and grid voltages are indicated only by their polarity with respect to ground. In a multigrid tube, it is conventional to denote the grid-to-cathode voltages by E_{c1}, E_{c2}, etc., for grids 1, 2, etc., counting outward from the cathode. In all the circuits, the capacitances C_b are to bypass alternating currents; all other unmarked capacitances are parts of resonant circuits. The triode oscillator circuit is of the type known as the Hartley, after its inventor, and is used for the generation of alternating power up to frequencies of hundreds of megacycles per second. It will be noted that no external grid bias voltage is used, because automatic stability is most simply achieved by use of a grid leak resistor. In such a circuit the grid current drawn provides a bias across this resistor which is automatically proportional to the strength of oscillation. In the pentode radio-frequency amplifier circuit, an input and output transformer is tuned to the frequency to be amplified and provides selectivity against undesired signals. Such a circuit is used at frequencies from as low as a few hundred kilocycles to about 200 mc. per second. The amplification can be controlled by variation of E_{c1} and when this is done, the pentode is preferably of the variable-μ type.

The superheterodyne type of receiver requires a change in frequency from an incoming high frequency to a lower fixed frequency, known as the intermediate frequency. Such a change is accomplished by a heterodyne process whereby a second, locally generated high frequency (known as the local oscillator) modulates the frequency to be changed. The process produces a "beat" at the difference frequency by means of a tube known as a mixer. In the pentagrid converter circuit shown, a single tube is used for both oscillator and mixer, and is the lowest-cost frequency changer with satisfactory performance. The remaining circuit of fig. 10 is one used for low-frequency amplification, such as in the audible range. The first stage uses a resistive load in the triode anode circuit coupled to the grid of the power pentode. An iron-cored transformer is used in the pentode anode for delivering the alternating power at high efficiency to an output circuit such as a loud-speaker.

An interesting feature of the circuit shown is that no external grid bias voltage is used on either tube; the resistors R_k and R_k' in the cathode provide a voltage drop on the cathode which is equivalent to a grid bias in its effect. When these resistors are not bypassed, they also have an alternating voltage across them providing what is known as inverse feedback, which helps reduce the small distortion resulting from a nonlinear tube characteristic.

The limitations on tubes with grid control are of great importance in understanding their usefulness. One of the primary limitations is because of the finite velocity of the electrons. An integration of equation (4) shows that the velocity of an electron which has been accelerated from rest by a voltage E is

$$v = \sqrt{-\frac{2e}{m}E} \tag{14}$$

so that its transit time in passing a distance d through a field-free space is

$$\tau = \frac{d}{v} = \frac{d}{\sqrt{\frac{-2e}{m}E}} = \frac{d}{5.93 \times 10^5 \sqrt{E}} \tag{15}$$

where d is in metres and E in volts. The negative sign within the square root occurs because, if E is positive, only a negative charge such as the electron will have a real solution. In a tube in which the electron is continuously speeded up, as in a triode cathode-to-anode traversal, the average velocity is only one-half to one-third of the final velocity, so that the transit time is two to three times the value given in (15), depending on the degree of space charge and the geometry. For the dimensions and voltages ordinarily used, τ is of the order of 10^{-9} seconds, so that such tubes have substantially instantaneous action up to frequencies of some tens of megacycles. At higher frequencies than this, both the convection and displacement currents (equation [7]) show frequency-dependent effects. Of these, a minor effect is the time delay in the output, which is seldom important. Of a much more serious nature is the delay in the normal capacitive displacement current, which no longer leads the applied voltage by a 90° phase angle at the higher frequencies. As a result, the input grid, which draws no direct current, appears nevertheless to have a resistive component of impedance which takes power from the input. The effect

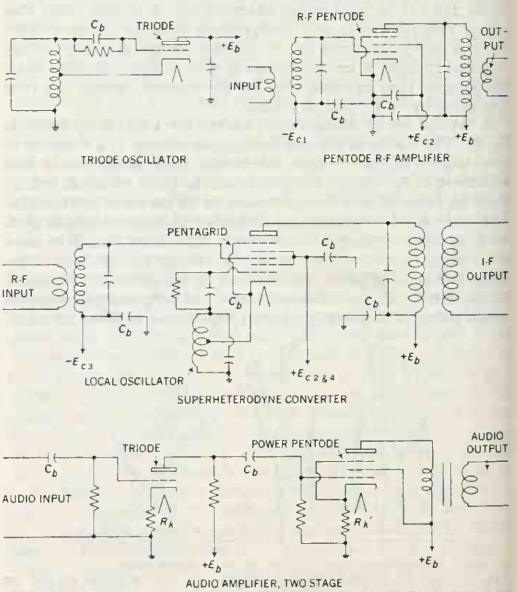


FIG. 10.—TYPICAL CIRCUITS FOR GRID-CONTROLLED THERMIONIC TUBES

gets worse as the frequency gets higher and is one of the chief limiting factors in amplification and oscillation. By use of very close interelectrode spacings, or very high voltage, special triodes have been made operative up to about 4,000 mc. per second, but at poor efficiency. The close spacings combined with circuit difficulties at such high frequencies have made other means more attractive.

Another limitation on conventional grid-controlled tubes appears when very wide frequency bands must be amplified, as in any system for the very rapid transmission of intelligence (*e.g.*, television). It is then found that the input and output capacitances of the tube become very important and must be minimized. The amplification can be shown to be proportional to the tube transconductance and inversely proportional to the product of bandwidth and capacitance. Considering the grid-cathode region, equation (13) shows that a high perveance, K, is needed for high transconductance. Since K varies roughly as $1/d^2$ where d is the cathode-grid distance, and since the input capacitance varies as $1/d$, the ratio of transconductance to capacitance improves as d is made smaller. Tubes for wide-band amplification, therefore, also require the very close spacing needed for high-frequency operation, although for a different reason. A practical limit is soon reached and secondary emission amplification must be resorted to beyond this point. In high-power transmitting tubes for wide bandwidths, it is found that the finite cathode-current density and limited anode dissipation reduce the power attainable with small areas (*i.e.*, low capacitance), so that a limit is imposed by the available cathode materials and anode cooling means. In addition, space-charge effects limit the current density at a given voltage and hence limit the power for a given area.

In the amplification of very minute signals, it is found that a lower limit on the detectable input is imposed by random electrical fluctuations (noise) of the tube and circuit. Because the individual electrons arrive at an anode at random times, there is a small fluctuation in the steady current of a tube. The effect is known as shot effect because of its similar nature to the sound effect produced in dropping large numbers of small lead balls from a tower, as in making shot. Since the fluctuation is in many cases proportional to the current in the tube, the lowest noise tubes will have a high ratio of transconductance to current. The highest ratio is obtained at low currents, when electron initial velocities are important. It is found that the theoretical maximum ratio is e/kT, which has the dimensions of reciprocal volts (volts^{-1}). (Here T is the absolute temperature and k is called Boltzmann's constant.) With the oxide-coated emitter, the value is about ten volts^{-1}. A higher ratio requires a thermionic emitter at lower temperatures, *i.e.*, lower initial velocities of emission. Thus the initial emission velocities play a part in limiting the amplification of minute signals. At low frequencies, there are additional sources of noise known as microphonics and flicker effect which limit amplification.

Nonlinear response of electron tubes is to be expected but is fortunately small enough so that distortion of amplified signals is not too serious and, even without special circuits, can be held below 10% harmonic content for a sine wave. The lower limit of distortion has been tremendously reduced by special circuits using inverse feedback. In such a circuit, a part of the slightly distorted output is recombined with the input signal in such a way as to reduce its effect on the tube. The distorted parts of the output are thereby greatly reduced in proportion to the undistorted, since they are effective in their entirety. When enough care is taken, the distortion can be reduced indefinitely by such means and inverse feedback circuits are common in audio amplifiers, radio broadcast transmitters, etc. The limitation because of nonlinearity has, therefore, been overcome in many important applications.

A very practical limitation on the power which can be dissipated in a given tube is imposed by the softening point of the glass envelope or seals. A second limitation in this respect comes from the high-temperature operation of the internal electrodes themselves. Obviously, their melting point cannot be exceeded but, long before this is reached, serious effects are found from thermionic emission from such electrodes as grids. Grid emission has been a serious problem in all power tubes, and many advances have been made possible only by radical improvement in cooling means for these electrodes.

V. DEFLECTION CONTROL OF ELECTRON STREAMS

A second important control mechanism is the deflection, or change in path, of an electron. Deflection is possible either by electric or magnetic fields at right angles to the electron flow. When the perpendicular fields are symmetric with respect to the axis or centre of a beam of electrons (*i.e.*, they pass through zero on the axis and are directed either toward or away from the axis on each side), the deflection occurs only on electrons displaced from the axis; and a lenslike action is obtained which can be used to converge or diverge the beam. The techniques are similar to those in optics and the science is called electron optics; the electron paths are analogous to the light rays and a change in potential of the field is analogous to a change in refractive index. Perpendicular deflecting fields which are unsymmetric with respect to the axis of an electron beam, on the other hand, deflect all the electrons in the beam and so change all the directions in the same way, as with an optical prism. Such deflection may be used for oscillography; for scanning in television tubes; or to deflect an electron beam on and off one or more anodes, either for amplification or for switching purposes.

Lenslike deflection occurred in connection with the rubber membrane model of the triode (fig. 6) in which it was seen that electron deflections could take place between grid wires. In the usual grid-controlled tube, little or no use is made of this deflection. In particular screen-grid tubes, however, the effect is used to focus the electrons between the screen-grid wires, so that their current is greatly reduced. In other tubes, the cathode is curved so as to focus the electrons between the control-grid wires to reduce their current at positive voltages. The electron optics of these examples is that of cylindrical lenses, which can be analyzed qualitatively on the rubber membrane. The spherical type of electron lens is widely used on tubes in which a finely focused pencillike beam is desired. In its electrostatic form, the spherical lens may consist of a set of circular apertures or adjacent cylinders through which the beam passes. Placing the successive electrodes at different potentials sets up perpendicular components of electric field, near the transition region, which are symmetric about the axis and so have a lens action. The magnetic spherical lens consists of a short current-carrying coil along whose axis the beam passes. The perpendicular field components exist at entrance and exit.

Returning to unsymmetric-field deflection, fig. 11 illustrates the deflection of a beam of electrons passing through two parallel deflection plates between which an electric field is set up by a voltage between the plates. Such a deflecting system is called electrostatic and is widely used in thermionic cathode-ray tubes and oscilloscope tubes, such as those used for radar. The electron source, not shown, includes the necessary electron lenses, whose action is separate from the deflection. It is found by integration of equation (4) that the electrostatic deflection is

$$D = \frac{1}{2}\frac{E_d}{E_0}\frac{L_1 L_2}{d} \qquad (16)$$

where E_0 is the acceleration voltage of the beam, E_d is the deflection voltage and the other distances are as shown on fig. 11.

With magnetic deflection, the electron beam is deflected at right angles to the deflecting field. Thus, in fig. 11, if the two plates are considered poles of an electromagnet, the beam will not be deflected as shown but in a plane coming out of the drawing. By in-

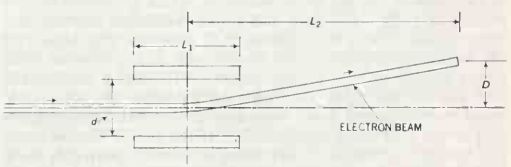


FIG. 11.—DEFLECTION OF AN ELECTRON BEAM BY TWO PLATES BETWEEN WHICH A VOLTAGE IS APPLIED

tegration of equation (5) it can be shown that

$$D = 0.296 \times 10^6 B \frac{L_1 L_2}{\sqrt{E_0}} \qquad (17)$$

where B is in webers per square metre, the lengths are in metres and E_0 is in volts. Magnetic deflection is common in picture tubes for black-and-white and colour television.

Magnetic deflection has another extremely interesting property. Since it is at right angles to the field and direction of motion, electrons injected into a uniform field will be deflected continuously so as to rotate in a circle. This circular path represents a balance between centrifugal force and the magnetic force (equation [5]) given by

$$r = \frac{mv}{Be} \qquad (18)$$

Since the time T taken to complete the circle is $2\pi r/v$, the angular frequency of rotation, $2\pi/T$, is therefore given by

$$\omega_c = \frac{eB}{m} = 17.95 \times 10^{10} B \ (B \text{ in webers per square metre}) \qquad (19)$$
$$= 17.95 \times 10^6 B_g \ (B_g \text{ in gauss or lines per square centimetre})$$

known as the cyclotron frequency. It is striking that this frequency is independent of the velocity of the electron and the radius of the path. The relation is of importance in some types of magnetron oscillators for generating very short wave lengths, since it represents a form of electron resonance. The combination of electric and magnetic deflection, as used in the magnetron, is discussed in the article RADAR, as are typical cathode-ray tubes for oscillography, television and radar display. (*See* also TELEVISION.)

VI. VELOCITY CONTROL AND TRAVELING-WAVE TUBES

The velocity of electrons in a constant current stream may be varied by means of a control voltage. Since there is no direct variation of the current as with grid control, velocity modulation (as it is usually called) is somewhat more free of transit-time effects on the input electrodes than grid control by means of space charge. Of course it is ultimately necessary to convert the velocity modulation into a current modulation, or no useful output would be available. Conversion of velocity modulation into a current modulation can be done by a variety of methods. The simplest conversion method allows the electrons to drift in space after modulation, until groups of fast electrons have caught up to the slow ones which just preceded them, forming a series of bunches. These bunches are equivalent in every way to the bunches which would have been formed by grid control yet the control method is far more advantageous for ultrahigh-frequency use.

Certain forms of thermionic-cathode, velocity-modulation tubes are known as klystrons and are particularly designed for operation in the frequency range from a few thousand to perhaps 50,000 mc. per second. To this end, klystron tubes often contain input and output circuits of the hollow-cavity resonator type. One form of such a tube is the two-cavity amplifier tube shown in cross section in fig. 12. The tube has circular symmetry about the axis and uses a cylindrical type of electron beam, often formed by an electron-optical system. An input is used to excite the input cavity through a coaxial transmission line terminated in a small coupling loop. The input-cavity resonator is adjusted so that its natural resonance occurs at the input frequency, presumably some thousands of megacycles per second. The high-frequency alternating electric field is maximum at the gap in the resonator, which is covered by two permeable grids. The input and output resonators and the final collector are all at a high positive D.C. potential with respect to the thermionic cathode of the electron gun, so the electron beam passes through at high average velocity. In passing through the grid gap of the input resonator the electrons are alternately speeded up or slowed down, depending on the instantaneous direction of the alternating electric field. If the transit time is short, substantially no control of the current is normally exerted by these grids because of their high positive potential. As the velocity-modulated beam passes along the drift space, it is "bunched" by

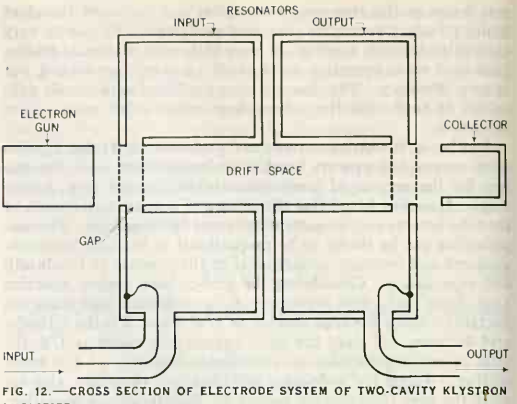


FIG. 12.—CROSS SECTION OF ELECTRODE SYSTEM OF TWO-CAVITY KLYSTRON AMPLIFIER

the faster electrons catching up with the slower ones just ahead. The bunches or groups of electrons then pass through the grid gap of the output resonator and, by changing the induced charge, cause a current to flow inside. The electrons are finally caught by the collector. The amplified output is taken from a coupling loop inside the output resonator.

It should be noted that the electrons need not strike the output resonator grids of fig. 12 to give an output. The klystron tube is an example of the use of the displacement-current term of equation (7) to the exclusion of the convection current. There are disadvantages of this type of klystron in that the resonators limit the bandwidth of the signals to be amplified. As an amplifier of very small signals, the klystron of this kind has excessive fluctuation noise compared with grid-controlled devices. Nevertheless, the klystron with two or more cavity resonators gives a very high-power amplification for narrow bandwidths. A variation of the klystron is the three-cavity type, in which an additional cavity is placed between input and output resonators to increase amplification. The reflex klystron uses only one resonator and a beam-reflecting electrode to obtain a very useful microwave oscillator (*see* RADAR: *Radar Components*).

The narrow bandwidth of the klystron is caused by the resonators employed; a different form of velocity-modulation tube eliminates these altogether and uses continuous interaction of a beam with a traveling wave on a helix. This device, shown in fig. 13, is called a traveling-wave tube and has comparatively unlimited bandwidth, a two-to-one frequency range being common. A beam of electrons is made to travel down the axis of a helix of wire at a velocity determined by its potential. An incoming signal, of a frequency in the thousands of megacycles, is brought in by a wave guide and impressed at one end of the helix to velocity-modulate the beam through the axial component of electric field. The helix, acting as a coiled transmission line, also transmits this signal toward its other end at about the wave velocity of light along the wire, or at an axial velocity reduced by the ratio of pitch to circumference of the helix. Adjustment of the helix potential makes it possible to adjust the electron velocity to nearly the same value as the axial wave velocity, so that a continuous interaction takes place down the tube and leads to a greatly amplified output. The currents induced in the helix are again displacement currents, so that the electrons need not actually strike any part of the helix.

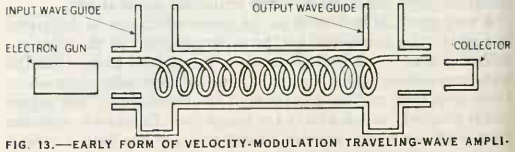


FIG. 13.—EARLY FORM OF VELOCITY-MODULATION TRAVELING-WAVE AMPLIFIER

A magnetic field the full length of the helix is used to keep the electron beam from diverging along the long path, although this is not shown in the figure. High amplification in the centimetre wave region, and bandwidths of some hundreds of megacycles, are made possible by such a traveling-wave velocity-modulation device.

The combination of electric field control with a deflecting magnetic field leads to a number of devices which may be described under the generic name magnetron. Originally a simple cylindrical thermionic diode whose anode current could be controlled by a magnetic field, the magnetron was later found to be useful as a generator of oscillations at centimetre and millimetre wave lengths. Best oscillator operation was achieved by a multiple-anode electrode, using a traveling wave around the periphery; high efficiency, high peak output power under pulse conditions, and operation to as high as 100,000 mc. per second were achieved. A form of magnetron is described in RADAR.

VII. SPECIAL ELECTRONIC DEVICES

There are some electronic tubes and related devices which cannot be readily classified in the categories discussed up to this point. One class of these consists of thermionic electron tubes which are particularly valuable in producing what is known as negative resistance. Negative-resistance devices have a falling characteristic of current versus voltage and are divided into two groups based on whether the current decreases as a result of increase in voltage (voltage-controlled) or whether the voltage decreases as a result of an increase in current (current-controlled). The electric arc and the glow discharge were the earliest negative-resistance devices and are of the current-controlled type. The more commonly used negative resistances in radio communications are of the voltage-controlled type, however.

Negative resistance is achieved in a tube known as a dynatron, which utilizes the principle that many metal surfaces increase in secondary emission when their voltage rises, thereby producing a falling current and hence a negative resistance. In other tubes, a negative transconductance is attained between a control grid and a positive grid that precedes it in the electron stream; negative resistance is produced by a direct feedback connection between the two. Most negative-resistance effects are achieved with more conventional grid-controlled tubes, by a phase-inverting feedback circuit. Negative-resistance effects are chiefly used to produce oscillations; *i.e.*, to generate high-frequency alternating currents.

Another important group of special electronic devices makes use of secondary emission for current multiplication. It is clear from fig. 5 that if secondary emission ratios of from three to ten can be achieved, the secondary emitter provides a simple and inexpensive means of increasing amplification by the same factor. Fortunately, the time taken to emit secondary electrons is so short that operation is possible up to thousands of megacycles per second. Furthermore, no circuits are involved to limit the frequency band in any other way. The secondary emitters can be used in cascade, with secondary electrons from a first emitter going to a second emitter to be multiplied again to a third, etc. Each emitter must be at a higher potential than its predecessor in order that it may collect secondaries at a high enough energy to emit more. This is usually done by increasing the D.C. voltage on each stage. Cascade multiplication can also be done by a dynamic method in which a high-frequency A.C. is used between only two emitters, using the transit time of the electrons, to allow the field to reverse for successive transits. Multistage secondary-emission multipliers are of great value in photoelectric tubes and television pickup tubes, where the initial electron currents are very small. With grid-controlled thermionic tubes, which employ much larger currents, the secondary-emission multiplier proved less successful because of changing secondary-emission ratio during operation.

Another type of tube which found moderate special purpose use was the storage or memory tube, forms of which are called graphecon, radechon and metrechon. With such tubes, an electrical signal is stored by means of a focused electron beam which forms a charge pattern on an insulating target. Either the same beam or a separate one can then scan the target many minutes later and retransmit the signal in its original form, speeded up, slowed down or even reconstituted in an entirely new form. Means to "erase" are also provided. Storage tubes are closely related to television pickup tubes, except that the latter use storage of a light image rather than of an electrical signal. A storage picture-viewing tube has also been devised, wherein an electrical signal is simultaneously stored and viewed as a bright visual image.

Other special tubes worthy of mention are those used for "counting" impulses. They utilize a thermionic emitter and a series of anodes; *e.g.*, ten for a decade counter. Each successive impulse signal applied to a control electrode causes the electron current to jump from one anode to the next one; in a decade counter, after the tenth impulse, the current has returned to the first anode and the tube is ready to repeat the sequence.

VIII. GAS-FILLED TUBES

In any so-called high-vacuum tube, there are many gas molecules remaining. However, the probability of an electron collision with a gas molecule is sufficiently slight in well-evacuated tubes so that no noticeable effect resulting from this gas is observed. If the gas pressure in a thermionic tube is high enough to support a column of perhaps 10^{-3} mm. of mercury, ionization by electron collision becomes noticeable. The electrons released are negligible in proportion to those comprising the currents producing the ionization. The positive ions, however, can cause a small current flow to negative electrodes which, in high vacuum, would be free of current. This is the major effect of very small quantities of residual gas. If the negative electrode in question is connected to a very high external circuit resistance, a loss in bias can occur which may shift its operating point unfavourably. There is also an increase of fluctuation noise in the tube at the frequencies below a few megacycles. At still higher gas pressures, sufficient ionization and recombination may be present so that visible light is radiated by the change in energy levels of the atoms concerned. When this much ionization is present, the tube is already out of the high-vacuum category and the positive ions appreciably affect the space-charge forces. In an oxide-coated cathode tube, for example, the electron space charge is partially neutralized so that abnormally high current can flow, up to the very high emission limit imposed by the cathode. The cumulative ionization, overheating of electrodes and release of more gas are often destructive.

When a gas-filled thermionic tube is properly designed and operated to take gas ionization into account by limiting the maximum current flow, the neutralization of the negative space charge by positive ions can be used to advantage. The gas used may be mercury vapour, argon, xenon, or even hydrogen. The large mass of the ions makes their motions slow and once ionized, they remain in that state for an appreciable time. As a result, intermittent operation is only possible at low frequencies, up to perhaps some thousands of cycles per second. Gas-filled thermionic rectifiers, however, are widely used to permit high anode currents at comparatively low voltage for power purposes. The current flow is almost entirely from the thermionically produced electrons, the ions being used only to cancel space charge. However, since the ions produced in the discharge may be drawn to the thermionic cathode, at higher operating voltages they may strike it with sufficient velocity to cause appreciable local heating. Some gas-filled thermionic tubes, once started, can therefore be operated with the cathode-heating power turned off and, with a cathode of small heat capacity, can even be made self-starting. The anode voltage needed to maintain the discharge is higher than for separately heated emitters.

Grid-controlled, gas-filled thermionic tubes, called thyratrons, came to be widely used in industrial control application. The grids used differed from those in high-vacuum tubes and often consisted only of a metal plate with one or more holes in it. The characteristics of such tubes are quite different from high-vacuum grid-controlled tubes, chiefly because the control is not continuous. The control-grid voltage may be used to prevent initiation of the gas ionization and consequent high electron current. Once the grid is made more positive, to initiate the discharge, a high current flows and the grid loses control. When the anode voltage

is alternating at low frequency, the tube is recycled at each alternation and the grid may be made more negative at any time to turn the tube off. By adjustment of the phase of an alternating grid voltage with respect to an alternating anode voltage, a continuous change of average anode current is made possible, even with such inherently discontinuous tubes. The triggerlike action and high current flow because of a very small control voltage have made the thyratron an extremely valuable tool in control applications even though its operation is limited to low frequencies.

In the extremely high-power applications, rectifiers with and without grid control are made by using a mercury arc between an anode and a cathode spot on a pool of mercury in an evacuated envelope. The rapidly vapourizing mercury at this point is ionized, and the heavy currents are entirely the result of the arc discharge rather than thermionic emission as in most of the lower-power devices. In one form of device, known as the ignitron, a third rodlike electrode is introduced at the mercury surface to initiate the discharge. Variation of potential on this igniter controls the current in very much the same way as the grid of the thyratron. Mercury-pool rectifiers are used up to powers of thousands of kilowatts.

A discharge is also possible in a tube which contains rarefied gas and cold electrodes. The current flow in this case is largely because of the ionization products (electrons and ions) and to secondary emission from the cathode. A very much higher voltage is required to initiate the discharge than in thermionic tubes, so that such devices are relatively inefficient as power rectifiers. In such a cold-electrode discharge, the positive ions strike the cathode with sufficient velocity to release secondary electrons and these, in turn, produce further ionization of the gas. The current must be limited by external resistance, or the cumulative ionization is disastrous to the tube. Such tubes have been made with grid control, as with the hot-cathode thyratron, and are used where the stand-by cathode power of the hot-cathode type is too large and the high voltage drop is not objectionable.

The voltage drop to maintain the cold-cathode discharge is largely a function of the secondary emission characteristics of the cathode. A pure metal, such as nickel, will have a high voltage drop, perhaps 150 v. Specially activated cathode surfaces may maintain the discharge at as low as 60 v. When a large cathode surface is used, such as a cylinder surrounding a central anode wire, it is found that the voltage drop is maintained accurately over a wide range of currents. Such tubes are used as voltage regulators, the regulation voltage depending on choice of cathode.

Gas-filled phototubes are commonly used to increase sensitivity over vacuum types and will be discussed below. For more detailed discussion of gas discharges, *see* ELECTRICITY, CONDUCTION OF: *Conduction in Gases*.

IX. PHOTOELECTRIC DEVICES

The simplest phototube consists of a photosensitive cathode and an anode in a vacuum; such tubes have been made in a great variety of shapes and sizes. Because the photocurrents are so small, space charge is negligible, and the anode may be a small wire spaced relatively far from the cathode. The cathode is ordinarily large enough to receive as much incident light as is convenient. The choice of cathode for a particular application depends on the nature of the light source. Typical examples are: antimony-cesium (Sb-Cs) cathodes for scintillation counters because the phosphors used emit chiefly blue light; bismuth-silver-oxygen-cesium (Bi-Ag-O-Cs) or antimony-potassium-sodium-cesium (Sb-K-Na-Cs) cathodes for television pickup tubes where panchromatic response is essential; silver-oxygen-cesium (Ag-O-Cs) cathodes for detection of infrared radiation. All the materials in fig. 4 are sensitive to ultraviolet radiation, but to utilize this property special glass or quartz windows must be used because ordinary glass does not transmit radiation below wave lengths of 3500 Å. A typical phototube and circuit are shown in fig. 14. The load resistor R_L is usually very high so that the maximum voltage is realized for the small photocurrent.

In using a phototube with a tungsten-filament light source, the increasing light output of the source, with wave length, must be considered in estimating sensitivity. The sensitivity is measured in such cases by the current in microamperes per lumen of incident light from a tungsten filament of specified temperature. Vacuum phototubes have from 20 to 200 microamperes per lumen with a filament of 2,870° K. It is possible to increase the sensitivity by a factor of more than five by use of a small amount of inert gas in the envelope. The gas pressure is well below that which will cause a discharge, but high enough so that appreciable ionization takes place. The positive ions, which are produced by the ejected photoelectrons, bombard the photocathode and release a number of new electrons by secondary emission. The gas-filled phototube uses, therefore, a very simple form of secondary emission multiplication. Because of the limitation imposed by the slow ion mobility, the gas amplification falls off with increasing frequency and, with modulated light, can only be employed with frequencies in the audible range.

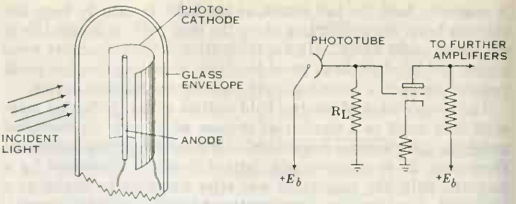


FIG. 14.—PHOTOTUBE AND CIRCUIT

Secondary emission multiplication, which was discussed above in connection with special grid-controlled tubes, is particularly valuable with phototubes because their currents are so small, and the formation of a highly sensitive photocathode and a good secondary emitter are very similar in technique. A tube using a nine-stage secondary emission multiplier is shown in fig. 15. The multiplier electrodes (dynodes) have been carefully shaped and positioned so that the forces on the secondary electrons draw them to the next stage, where they are again multiplied by the secondary-emission ratio. Such a tube, using an over-all voltage of about 1,000 v. divided equally between stages, has a current amplification of about 1,000,000 and can attain the remarkable sensitivity of 100 to 1,000 amp. per lumen to tungsten light at 2,870° K. The tube can be used to detect very small light intensities but is easily overloaded at higher light levels. The lower limit of light which can be detected is imposed by the masking effect of residual electron emission from the photocathode, usually by thermionic emission at room temperature. Cooling of the tube to liquid-air temperatures, or below, reduces the residual current without affecting the photoelectric emission. In this way the most sensitive light detector known has been made possible.

Photoelectric surfaces can be prepared which are semitransparent and yet electrically conducting. Such surfaces are used in some multiplier tubes, in image tubes and in television pickup tubes. If a scene to be viewed is focused on the one side of a flat transparent photocathode, the electrons are ejected from the other side to form an electron image with a current distribution in exact proportion to the light intensities of the scene. By use of a magnetic or electrostatic lens system, the electrons can be held in their correct relative positions but increased in their velocity so as to essentially amplify the available energy. In infrared image tubes, whose purpose is to convert an infrared image to visible light, the electrons are accelerated by some thousands of volts and then caused to strike a luminescent material. A visible image of the invisible infrared radiations then appears on the luminescent screen.

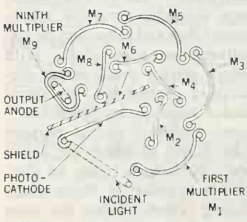


FIG. 15.—CROSS SECTION OF NINE-STAGE MULTIPLIER PHOTOTUBE

In the sensitive television pickup tube known as the image

orthicon, a similar electron image is used to eject secondary electrons from a very thin, slightly conducting glass target, leaving it charged positively and in proportion to the light distribution in the original scene. A cathode-ray beam scans the other side of the glass target to discharge it, giving rise to signals which, after secondary emission multiplication, are used in the television system. *See* also TELEVISION.

X. TECHNIQUE AND CONSTRUCTION

Electron tubes have reached such a state of refinement that the materials used, the technique of evacuation, activation of sensitive emitting surfaces, and the mechanical construction are extremely important in maintaining uniformity and low-cost mass production. The high-vacuum pumps used consist of mercury or oil diffusion pumps which exhaust into the cruder vacuum of rotating mechanical oil pumps. In addition, when the best vacuum is desired (in particular for all phototubes), a liquid-air vapour trap is used in order to reduce the vapour pressure of condensable substances such as water, mercury or oil, which would otherwise limit the pressure. In mass production, the pumps and traps are connected to large rotating turrets whose periphery holds 14 or more rubber ports into which the exhaust tubulations of the tubes to be exhausted are placed. As the turrets revolve, each tube is connected in turn to a rough vacuum and then to progressively better vacuum systems. At the same time, the cathodes of thermionic tubes are heated above normal operating temperature, following a special schedule which activates the cathode material.

The technique of obtaining a high vacuum is by no means limited to the exhaust system. It is found that glass and metal surfaces hold quantities of gas both in adsorbed form and in chemical combination. If such gas is not removed, it would slowly be released and limit the life of the tube. Before use, therefore, all metal parts of the tube are heated to incandescence in an atmosphere of hydrogen. This reduces oxides, removes contamination and drives out some of the heavier gases to be replaced by the more volatile hydrogen. During the actual exhaust process, the glass or metal envelope of the tube is raised in temperature as high as possible to drive out gas. Whenever possible, high-frequency induction coils are placed over the tube and the internal electrodes heated to incandescence. The rotating-turret, mass-production, exhaust machine is, therefore, a complex and costly piece of apparatus.

In all the highest-production, lowest-cost tubes, the vacuum is tremendously improved by use of a special substance called the getter. The getter is usually barium metal which is released from a chemical compound by induction or other electrical heating, after the tube has been exhausted and sealed off. The barium is normally deposited on the glass or metal envelope. The action of a getter is to combine chemically with active gases and to hold inactive gases by adsorption. It not only cleans up the rather high gas pressure remaining after the one- to two-minute exhaust on high-speed systems, but it remains active in the tube to clean up any gas released during operation. Because it is most active with ionized gases, its effectiveness is greater when the tube has voltages on it, and currents flowing, than when the tubes are not operating. In power tubes for high voltages and frequencies, the danger of electrical leakage resulting from the getter deposit or from high-frequency losses in the getter is so great that a getter can not always be used. This is a serious drawback and such tubes must therefore be exhausted longer and more carefully.

The materials used in electron tubes must be chosen carefully to meet the requirements of freedom from contamination, ability to withstand high temperatures and maintenance of proper electrical properties. The best insulating materials are glass, mica and ceramics. Since it is always necessary to make good electrical connection to the internal electrodes, methods of making vacuum-tight seals between a metal and glass or ceramic are of great importance. The technique has been to employ special metal alloys whose coefficient of expansion roughly matches that of the glass to be used, or, conversely, to develop a glass which matches the expansion of a desirable metal. For tube electrodes, tungsten, molybdenum and tantalum are valuable for their high melting point and strength. These metals are difficult to fabricate and high in cost; they are not used in low-cost tubes when steel, nickel or chromium alloys can be substituted.

The sensitive and active surfaces, such as thermionic and photoelectric cathodes and secondary-emitting materials, are of basic importance. Because they are so sensitive to surface conditions, their preparation is done with extreme care and is still, to a great extent, an art which has been developed by trial-and-error methods. The very important oxide-coated thermionic cathode, for example, is found to be greatly affected by minute impurities in the nickel base metal. Certain types of surface contamination on other electrodes in the tube have also been known to "poison" the cathode emission. Theoretical reasons for these phenomena are not well understood and, in mass production, difficulties are sometimes encountered which are the result of completely unknown causes. The preparation of photosensitive surfaces is an even more delicate task which is done by individual processing of each tube. In these surfaces the cesium or other sensitive surface layer is supplied during processing by release from stable compounds previously placed in the tube. Most photoemissive surfaces are very sensitive to heat and other mild abuses. The very slight variations in technique which are sometimes encountered may cause appreciable change in the sensitivity and spectral response. The most stable secondary-emitting surfaces are often simple oxides of magnesium or beryllium which are present in a thin-enough layer to be conducting. Such surfaces can be exposed to air at a slight sacrifice in ultimate performance and can therefore be processed outside the final tube. In secondary-emission multiplier phototubes, the small currents and freedom from overheating allow the use of otherwise less stable and higher-ratio surfaces, which are often sensitized almost exactly like the photoelectric cathode.

See INSTRUMENTS, ELECTRICAL MEASURING; PHOTOELECTRICITY; RADIO; THERMIONICS; *see* also references under "Electron Tube" in the Index volume.

BIBLIOGRAPHY.—*General References:* P. G. Andres, *Survey of Modern Electronics* (1950); A. W. Kramer, *Elementary Engineering Electronics* (1945); R. G. Kloeffler, *Principles of Electronics* (1942); J. Millman and S. Seely, *Electronics*, 2nd ed. (1951); W. G. Dow, *Fundamentals of Engineering Electronics*, 2nd ed. (1952); W. W. Harman, *Fundamentals of Electronic Motion* (1953); K. T. Compton and I. Langmuir, "Electrical Discharges in Gases, Part I," *Rev. mod. Phys.*, 2:123–242 (April 1930), "Electrical Discharges in Gases, Part II," *Rev. mod. Phys.*, 3:191–257 (April 1931). *Thermionic Tubes:* K. Spangenberg, *Vacuum Tubes* (1948); A. H. W. Beck, *Thermionic Valves* (1953). *Gas-Filled Tubes:* J. D. Cobine, *Gaseous Conductors* (1941); L. B. Loeb, *Basic Processes of Gaseous Electronics* (1955); F. A. Maxfield and R. R. Benedict, *Theory of Gaseous Conduction and Electronics* (1941). *Electron Tube Circuits:* H. Reich, *Theory and Applications of Electron Tubes*, 2nd ed. (1944); L. B. Arguimbau, *Vacuum-Tube Circuits and Transistors* (1956); T. L. Martin, Jr., *Electronic Circuits* (1955). *Special Subjects:* M. Knoll and B. Kazan, *Storage Tubes and Their Basic Principles* (1952); V. K. Zworykin and E. G. Ramberg, *Photoelectricity and Its Application* (1949); J. R. Pierce, *Traveling-Wave Tubes* (1950); J. B. Fisk, H. D. Hagstrum and P. L. Hartman, "The Magnetron as a Generator of Centimeter Waves," *Bell Syst. tech. J.*, 25:167–348 (April 1946); A. L. Reimann, *Thermionic Emission* (1934); A. L. Hughes and L. A. Dubridge, *Photoelectric Phenomena* (1932); V. K. Zworykin *et al.*, *Electron Optics and the Electron Microscope* (1945). *History:* G. L. Archer, *History of Radio to 1926* (1938), *Big Business and Radio* (1939). (E. W. Ho.)

ELECTROPHORESIS is the movement of charged particles in solution under the influence of an electrical field. It is one of several electrokinetic phenomena involving relative mechanical motion in a heterogeneous system containing a liquid phase in which an electrical force field is set up. If the liquid rather than the particles moves, *e.g.*, through a fixed diaphragm, the phenomenon is called electroosmosis. The latter was discovered in 1808 by the Russian physicist F. Reuss, who observed the transfer of water from the anode (or positive) chamber to the cathode (or negative) chamber when an electric current was passed through an earthenware diaphragm. Electrokinetic theory was developed by Hermann von Helmholtz about 1880 with modifications by Friedrich Kohlrausch, William Henry, and others since that time. Thereafter, various electrophoretic techniques were developed, particularly the moving boundary method and the microscopic method. In the former, the motion of a boundary

between a colloid and the solvent is observed; in the latter, the movement of small particles such as bacteria is measured directly under the microscope. Electrophoresis was first developed for the study of inorganic electrolytes, both for low molecular weight ions and for colloids such as clay suspensions, and, indeed, the "electrophoretic effect" is still invoked in the theory of electrolyte dissociation (*see* ELECTRICITY, CONDUCTION OF: *Conduction in Liquids*). However, the most important application of electrophoresis has been to the study of proteins.

The development of electrophoresis as a quantitative analytical method dates from the Ph.D. dissertation in 1930 of Arne Tiselius (*q.v.*), subsequently awarded the Nobel prize (1948) for his contributions in this field. Tiselius modified the moving boundary method to insure stability of the boundary and to permit direct optical observation and photographic recording of its rate of migration. By making use of the refractive index difference between the protein solution and the buffer (the Schlieren optical principle), the need for coloured substances was eliminated, and a technique for qualitative identification and quantitative estimation of colourless proteins was introduced. This led immediately to the discovery of the alpha, beta and gamma globulins of serum and later to widespread application of electrophoresis to the study of changes in the distribution of serum proteins in disease. However, the complexity and cost of the Tiselius apparatus prompted the introduction of the simpler method of zone electrophoresis, that is, electrophoresis in stabilized mediums such as filter paper, agar, starch gels, etc., instead of in free solution.

Moving Boundary Method.—In the Tiselius (or moving boundary) method, a sharp boundary is formed between the solution of a colloidal electrolyte, such as a protein, and a buffer solution against which it was dialyzed. The pH (negative logarithm of the hydrogen ion concentration) is fixed by the buffer solution, usually a mixture of a weak acid and its salt. The boundary is set up in a compartmented glass cell of a rectangular U shape which is immersed in a thermostated bath at about 1° C. The cell is connected through a closed system to reversible silver and silver chloride electrodes in vessels containing a reservoir of buffer. The cell is located in the optical path of the instrument in such a way that a light beam passing through the channel of the cell is deflected at the protein boundary and cannot pass through the lens of a camera focused on the cell. Therefore, the boundary appears as a shadow at the conjugate point in the image plane of the camera. By the use of a cylindrical lens system or by mechanical scanning, the boundary is transformed optically into a pattern having the shape of an ideal distribution curve. Because the refractive index is related to protein concentration, the area under the curve is directly proportional to the amount of protein in solution.

When an electrical potential of the order of several hundred volts is applied, the protein molecules migrate at a rate which is dependent upon their charge, the field strength, etc. Since proteins are amphoteric substances (*i.e.*, capable of having either a positive or negative charge), their charge varies with pH and depends on the buffer selected. No movement of the boundary occurs at the pH of zero net charge (the isoelectric point). However, at a pH above the isoelectric point, the protein migrates as an anion, that is, to the anode or positive pole, and vice versa below the isoelectric point where the protein is a cation. Since proteins differ in their isoelectric points and in their acidic and basic dissociation curves, it is possible to achieve their separation by electrophoresis at the appropriate pH. Under these circumstances each protein component in a mixture may give rise to a separate refractive index boundary and a corresponding peak in the electrophoretic pattern. Since the area under each peak is proportional to the concentration, the relative percentage of each component can be evaluated quantitatively. Thus, in the illustration, blood serum was separated into its protein components, albumin, α_1-globulin, α_2-globulin, β-globulin and γ-globulin, plus an artificial boundary left at the origin (δ and ϵ in the respective limbs of the U tube).

Of course, if a protein is electrically homogeneous, a single moving boundary is obtained at all values of pH, and hence, electrophoresis is widely used as a valuable criterion of purity for proteins.

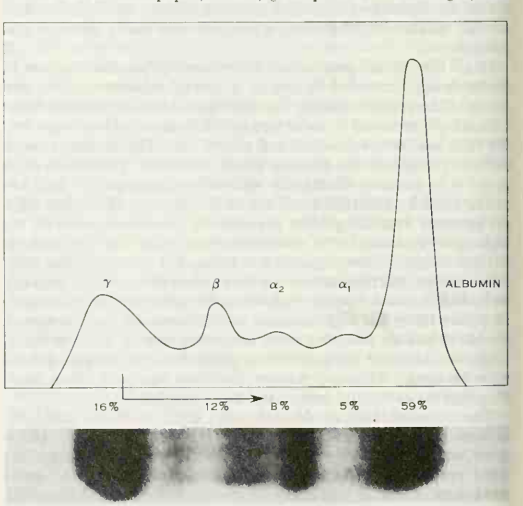
Electrophoretic Mobility.—The electrophoretic mobility u is the velocity of a particle in a unit electrical field (1 v. per centimetre). If the velocity is given by $\delta x/t$, where t is the time in seconds, E_p is the electric field strength in the protein solution in volts per centimetre, and x is the distance traveled in centimetres, the following relation is obtained:

$$u = \frac{\delta x/t}{E_p}$$

For most proteins, u has a magnitude of the order of $0 - 10 \times 10^{-5}$ cm. per second per volt per centimetre at 0° C. The sign of the mobility is determined by the charge on the particle, positive for positive ions, negative for negative ions. In a buffer solution of given pH and composition and at a stated temperature, u is a physical constant characteristic of the protein and only slightly dependent on the particle size and protein concentration. To minimize convective streaming of the boundary, the mobility is usually measured near 0° C. close to the temperature of maximum density of the solution. A plot of mobility vs. pH yields a curve paralleling the hydrogen ion dissociation curve of the protein. The isoelectric point is the pH of zero mobility. The latter is usually about pH 5 to pH 7, *i.e.*, near neutrality, but may be as low as pH 1–2 for acidic proteins such as pepsin or about pH 11 for basic proteins such as ribonuclease.

Many modifications of the original Tiselius apparatus have been introduced, and several commercial models are available. Interference optical systems may replace the refractive index system, etc., but the distinguishing features of the Tiselius method remain as follows: (1) migration of a boundary in free solution; (2) thermostatic control to reduce temperature convection; (3) photographic recording of the pattern; and (4) quantitative evaluation of electrophoretic composition and mobility.

Zone Electrophoresis.—Zone electrophoresis differs from the moving boundary method in that the components (1) migrate as separate zones rather than as advancing boundaries of overlapping zones; (2) are stabilized against convection by movement through the water in the pores or interstices of a semisolid supporting medium such as filter paper, starch, glass powder or silica gel; and



ELECTROPHORETIC PATTERNS OF HUMAN BLOOD SERUM OBTAINED (TOP) BY THE MOVING BOUNDARY (TISELIUS) METHOD IN FREE SOLUTION AND (BOTTOM) BY THE ZONE METHOD OF FILTER PAPER

In the top diagram, each peak is labeled for the component it represents, e.g., α_1-globulin, etc.; and its relative concentration is given for normal serum. In the bottom photograph, the bands of varying intensity correspond to the protein components designated immediately above for the Tiselius method. The vertical tail of the arrow indicates the starting boundary, and the arrow denotes the direction of movement of the protein components

(3) unless coloured, are not detectable until the experiment is completed and the protein is stained with an appropriate dye. Zone methods do not yield accurate mobility or isoelectric point data because of the electroosmotic counterflow of liquid and lack of temperature control. However, with the introduction of photoelectric scanning of the dyed areas, it became possible to obtain electrophoretic patterns approximating those given by the Tiselius method.

Zone methods have found wide favour because they require only simple, economical apparatus that is easy to operate, and also because they permit complete separation of the protein components. The latter property is of advantage in preparative work, but only certain mediums such as powdered starch, latex sponge, etc., are useful for this purpose since irreversible adsorption occurs on many commonly used mediums such as filter paper and starch gel. Although zone methods sacrifice accuracy for convenience, they are more readily applicable to lower molecular weight substances such as amino acids and peptides, and only minute quantities are needed.

Filter Paper Electrophoresis.—Filter paper electrophoresis offhand appears analogous to paper chromatography but is based on electrical charge rather than adsorption phenomena. Because of the variety of simple, commercially available apparatus with a capacity for multiple operation and the requirement of only a drop of serum, this method has been widely adopted in clinical laboratories for measurement of the protein distribution in pathological serums. After electrophoresis, the filter paper strips are dried, stained with a dye and evaluated photoelectrically (*see* illustration). For lower molecular weight substances, two-dimensional analysis may be done, with chromatography in one direction, followed by electrophoresis in another. In this way a resolution is obtained that is based both on charge and adsorption phenomena. High voltage paper electrophoresis at a field strength of several thousand volts per centimetre has been successful for separation of peptides, but great care must be exercised to prevent fires or other accidents.

Starch Gel Electrophoresis.—This method is analogous to that with filter paper except that the migration occurs in a gel of hydrolyzed starch. Although the protein is irreversibly adsorbed and adequate methods for the quantitative estimation of the proteins have not been developed, the method has a remarkable capacity for resolution of the minor components of serum, particularly when done two dimensionally. About 20 spots are obtained for human blood serum instead of the 5 or 6 components revealed by the Tiselius method and by the filter paper technique. In 1956, starch gel electrophoresis gave the first evidence for genetic control of certain minor serum proteins such as the haptoglobins and "prealbumins."

Preparative Electrophoresis.—Preparative electrophoresis is designed for the separation of the individual components of naturally occurring mixtures of proteins such as blood serum and egg white. Although large size preparative cells with sliding compartments have been designed for the moving boundary method, application has not been very successful because of the overlapping of the concentration zones in free solution and the difficulties of sample removal. Several continuous flow "paper curtain" devices have been developed in which the separated components drip from points at the edge of the paper. Other principles that have been employed simultaneously with electrophoresis are convective flow and adsorption chromatography. The most adaptable procedure, however, is zone electrophoresis in unhydrolyzed starch, in latex sponges, etc.; this permits complete separation into zones in the supporting medium which is then cut into segments from which the protein may be eluted. The method in conjunction with an assay is valuable for identification of biologically active substances in a mixture.

Applications.—Although electrophoretic separation is based almost wholly on electrical charge, the amphoteric property of proteins affording a variation in the charge by change of the pH has enabled manifold adaptations of the method to the problems of characterization, analysis and purification of the proteins. Commercial apparatus of many types is available, but the field is not yet stabilized and many ingenious combinations of methods such as immunoelectrophoresis are being introduced. Industrial use of electrophoresis has been limited to process and separation control of a few plant and animal products. Whereas the most extensive application has been made in the study of pathological sera, the most important contributions have been to the field of biochemistry and the better understanding of normal serum proteins, viruses and enzymes.

Electrophoresis has aided in the detection and identification of many new components of plant and animal tissues. Although its use is largely restricted to water-insoluble substances bearing a net electrical charge, applications have been made to many classes of compounds besides proteins. These others include amino acids, peptides, mucopolysaccharides and soaps. Future developments important to biology and medicine may be anticipated.

See M. Bier (ed.), *Electrophoresis* (1959); G. E. W. Wolstenholme and E. C. P. Millar (eds.), *Paper Electrophoresis* (1956). (F. W. P.)

ELECTROPLATING, the art of producing metallic coatings through the action of electric current. Such coatings commonly have the purpose of improving appearance, corrosion resistance, hardness, bearing qualities or other properties of the "basis metal" on which the coating is produced. Typical examples are: plating of steel with copper, nickel and chromium in the automotive field; production of tin-plated steel for food cans; and manufacture of silver-plated tableware. Significant properties of an electroplated coating are: adherence to the basis metal, thickness, smoothness, porosity, density, hardness, ductility, colour and brightness.

Basic Principles.—The essential parts of a typical electroplating system are: (1) the plating bath, which contains a compound of the metal to be deposited; (2) a source of direct current (D.C.) electricity; (3) the object to be coated (the basis metal); and (4) a sheet or bar of the plating metal. The negative terminal of the D.C. source is connected to the object (the cathode), while the positive terminal is connected to the plating metal (the anode). When both these electrodes are immersed in the plating bath, the following phenomena occur: (1) the metal ions (positively charged particles produced by the metal salt in the plating bath) migrate to the cathode surface (under the influence of the electric current) and are there transformed to the metallic state and attached to that surface; (2) the anode metal is dissolved by the action of the current, forming new metal ions (which replace those removed at the cathode). Thus, the combination of these two electrode processes results in the transfer of the plating metal from bulk form (the anode) to the form of a coating on the basis metal (the cathode).

In principle, the thickness of a plated coating is determined by the time of plating and the current density employed. The latter is expressed in amperes per square foot (or amperes per square decimeter) on the cathode surface. In most baths, however, the current does not produce the theoretical amount of metal deposit; some is lost in the deposition of hydrogen gas along with the metal. The ratio of the actual metal deposit to the theoretical (expressed in %) is called the plating efficiency.

Power Sources.—Several types of D.C. sources are used in electroplating, depending on the magnitude of the operation. For small-scale plating (using less than five amperes), such as gold finishing of jewelry, dry cells or storage batteries may be used. For intermediate scale installations (about 5–1,000 amp.), disk rectifiers (copper oxide, selenium, germanium, silicon) are frequently used. For very large-scale plating, where many thousands of amperes are required, motor generators are used almost exclusively.

Cleaning and Pickling.—Before a metal object can be electroplated satisfactorily, it must be thoroughly cleaned so that it is free of all grease, dirt, oxide or other nonmetallic foreign matter. This requirement is of particular importance for the adherence of the subsequent coating; it may also influence smoothness, brightness and porosity.

In practice, grease and dirt are removed from the surface to be plated (the "work") by three means: organic solvents; emulsions of organic solvents in water containing a wetting agent and alkali; and aqueous alkaline solutions. Organic solvents are typified by

trichloroethylene; the work is dipped into the boiling solvent, then into the cold solvent, and finally withdrawn through the hot vapour above the boiling liquid. A typical emulsion cleaner might be kerosene suspended in a soap solution. The work is immersed in the solution (cold or hot) and agitated; it is then rinsed in hot water. Emulsion cleaning is particularly effective in removing grease and dirt, the latter being suspended at the oil-water interface of the emulsion. Simple alkaline solution cleaners are used where the degree of contamination of the work by grease and dirt is not excessive, or where a preliminary solvent degreasing has been performed. The action of alkaline cleaners depends principally upon saponification of oils or greases which may be present on the work. The compounds most used are: sodium hydroxide, sodium carbonate, sodium phosphate, and sodium silicate. Such cleaners are used in two ways: by simple immersion in the usually hot solution; or by immersion with applied current (electrocleaning). In the latter practice, the work may be made cathodic (connected with the negative terminal of the D.C. source) or anodic. In either case, the vigorous evolution of gas at the work surface provides an effective scrubbing action.

After removal of grease and dirt from the work, it is further necessary to remove metal oxide from the surface. This is ordinarily accomplished by pickling the work in an acid solution. The type of acid and its strength are selected on the basis of the metal to be pickled and the amount of oxide present. For iron and steel, the principal acids used are sulfuric and hydrochloric, the latter being somewhat faster in action but more expensive.

A frequent difficulty encountered in pickling (and sometimes in plating) of certain steels is known as hydrogen embrittlement. Part of the hydrogen evolved during pickling (or plating) is absorbed by the steel and is believed to cause the brittleness. Remedies include: short pickling times; use of plating baths which do not cause much hydrogen evolution at the work surface; and baking of the work after plating.

Pickling of copper and its alloys is usually done with mixtures of acids based mostly on sulfuric, but containing nitric and hydrochloric in smaller amounts. Such mixtures are called bright dips, and usually give an extremely bright finish on brass or copper. Other metals and their appropriate pickling solutions are: zinc and die castings, hydrochloric acid; aluminum and its alloys, nitric acid; lead, fluoboric acid.

A special form of metal finishing which in effect combines pickling and bright dipping is known as electropolishing, and involves the anodic treatment of a metal in a suitable solution. The metal is dissolved in such a way that marked smoothing and brightening occurs. Such treatment can produce surfaces comparable to those resulting from buffing—in many cases at much lower cost.

Plating Baths.—Every plating bath must contain a source of the metallic ions which are to be electrodeposited. Some plating solutions employ only simple salts, such as copper sulfate, nickel sulfate, or tin sulfate. Others use compounds in which the metal ions are mainly present as "complex ions," *e.g.*, sodium stannate (for tin), sodium argento-cyanide (for silver), or chromic acid (for chromium).

Other chemicals are frequently employed for the following purposes: controlling the acidity or alkalinity of the bath; improving the electrical conductivity of the bath; implementing dissolution of the anodes (thus replacing the deposited ions); and influencing the smoothness, brightness, grain size, porosity, hardness, ductility, and other properties of the deposited metal.

Selection of Plating Baths.—While it is generally true that most metals can be electroplated on any other metals, certain rules govern the selection of particular plating baths if adherent and otherwise sound coatings are to be obtained. Of major importance is the avoidance of immersion deposits. When a fairly active metal (such as iron) is immersed in a solution containing a high concentration of ions of a more noble metal (such as a copper sulfate plating bath), a displacement reaction occurs spontaneously, resulting in the dissolution of some iron and the formation of a copper coating on the remaining iron. Such coatings are rarely adherent, and lead to poor adherence of any further copper coating produced in the conventional manner. Similar difficulties are met in plating brass or copper with silver or gold.

Fortunately, there are many noble metal baths in which the actual concentration of metal ions is extremely low, most of the ions being in the form of "complex ions." Use of such baths avoids immersion deposits and produces sound coatings of copper on steel, silver on brass, copper on zinc, etc.

Addition Agents.—Of considerable significance to the production of smooth, bright metal deposits is the use of addition agents. These are principally organic compounds, usually colloidal, and have been employed on an almost purely empirical basis. Especially in the early period of development of electroplating, the choice of an addition agent for a particular bath was largely dependent on the imagination and spirit of adventure of the plater. Glue, gelatin, albumin, organic wetting agents, traces of metal ions, aldehydes, ketones and amines are but a few of the substances which have been employed. The mechanism of addition agent action in refining the quality of the metal deposit is not fully understood. Perhaps the most accepted theory is that the addition agent (especially if colloidal) codeposits to some extent with the metal, thus interfering with the regular crystal growth of the metal. Such continuous interference results in a smaller grain size, and hence a smoother deposit.

The concentration of each ingredient of a plating bath is usually quite important. It is customary therefore (especially in large-scale, continuous operations) to perform routine analyses for the significant components of the bath. Similarly, the temperature of operation of the bath is usually critical; most plating solutions are held at the optimum operating temperature by a thermostat.

Testing of Deposits.—Probably the most important properties of electroplated coatings are thickness and adherence. Other criteria for deposits include corrosion resistance, brightness, porosity, smoothness, hardness, etc. Tests for some of these qualities are discussed below.

Thickness of a deposit may be measured in a variety of ways, broadly divided into destructive and nondestructive methods. The destructive methods include: stripping of the deposit from a weighed sample of the cathode, followed by reweighing to determine the amount of metal deposited; exposure of the plated object to a chemical reagent which dissolves the coating, the time for removal being a measure of the thickness; sectioning the plated object and measuring the plating thickness with the aid of a microscope; and "de-plating" the metal with applied current, using the current-time product as a measure of deposit thickness. Nondestructive methods include: direct micrometer measurement of the dimensional change resulting from plating; magnetic techniques, usable where the base metal is magnetic and the plated metal nonmagnetic or vice versa; and X-ray measurement.

Adherence of deposits can usually only be tested qualitatively, at best. Stretching, bending or repeated flexure of the test piece is usually sufficient to detect poor bonding. Attempts to devise direct tensile tests for the plating bond were fairly successful, but sufficiently involved to discourage their general use.

Corrosion resistance testing is usually accomplished by exposure of the plated object to simulated service conditions or to conditions deliberately made severe. Outdoor exposure in rural, urban and marine atmospheres are typical of the former, while salt spray exposure is widely used for the latter.

Typical Plating Processes.—Although thirty-odd metals can be electrodeposited from aqueous solution, only about half of these achieved commercial significance. While the classification of these metals according to their importance in electroplating will not be attempted, some of the commoner plating baths will be described. The most commonly plated metals include: copper, nickel, chromium, tin, cadmium, zinc, silver and gold. Other metals which have more limited use in electroplating are: iron, lead, rhodium, indium, platinum and cobalt. While several alloys can be electroplated, only brass and nickel-tin have extensive use.

Copper.—The two commonest baths are the cyanide bath and the sulfate bath. The former is generally used to produce the initial copper plate on steel, where subsequent nickel and chromium deposits are to be applied. The latter is used chiefly for building up heavy copper deposits over an initial coating of copper

and finds much use in the re-forming of printing rolls. A typical cyanide copper bath is the following: cuprous cyanide, CuCN, 22.5 g./l. (grams per liter); sodium cyanide, NaCN, 34 g./l.; sodium carbonate, Na_2CO_3, 15 g./l.; sodium thiosulfate, $Na_2S_2O_3{\cdot}5H_2O$, 2 g./l. (brightener). Operating temperature is 40° C.; current density, 15 amp. per square foot.

Nickel.—While a very great number of proprietary bright-nickel baths are used commercially, most are based on the Watts bath, which has the following composition: nickel sulfate, $NiSO_4{\cdot}6H_2O$, 240 g./l.; nickel chloride, $NiCl_2{\cdot}6H_2O$, 45 g./l.; boric acid, H_3BO_3, 30 g./l. Without addition agents, a dull gray deposit is obtained, which may be buffed to a bright finish. When suitable addition agents are used, quite bright deposits can be obtained. In addition to brighteners, it is customary to use a wetting agent for the purpose of preventing pits in the deposit.

The principal use of nickel in electroplating is (in combination with copper) as a preliminary corrosion-preventive coating on steel, over which is applied a thin layer of bright chromium.

Chromium.—Commercial chromium plating has two major divisions: decorative chromium plating (thin coatings, 5–30 millionths of an inch), and industrial chromium plating (relatively heavy coatings, frequently several thousandths of an inch). Both employ the same type bath; both yield essentially the same properties in the deposit. A bath which can be employed for either type of chromium plating has the following composition: chromic acid, CrO_3, 250 g./l.; sulfate ion, SO_4^{2-} (as sulfuric acid), 2.5 g./l.

Decorative chromium on automotive hardware was thought for many years to contribute little to the corrosion resistance of the base metal (see *Nickel* above); in the late 1950s, however, it was discovered that 30 to 50 millionths of an inch of chromium gave marked improvement over the usual 5 to 10 millionths, both being plated over the same copper-nickel undercoats.

The current efficiency of chromium plating baths is quite low, seldom over 15%. Unlike most baths, inert lead anodes must be used. The chromium removed during plating is replaced through chromic acid additions.

Tin.—The material of which the common tin can is made is actually steel on which has been applied a thin coating of tin. Early practice in making tinplate employed the hot-dipping method of coating the steel. During the years 1940–45, however (especially in the United States), the electrolytic method was greatly developed. In this process, continuous bands of steel (usually about 0.010 in. in thickness) are cleaned, pickled and electroplated with about 30 millionths of an inch of tin, at speeds up to 2,000 ft. per minute. The baths generally used are of two types: acid (containing stannous sulfate) and alkaline (containing sodium stannate).

Cadmium.—The largest use for cadmium is as a semidecorative and corrosion-resistant coating for steel. The mechanism of its rust-preventive effect lies in its sacrificial action when exposed to corrosive environments. It is actually a less "noble" metal than iron; it is attacked more readily and thus protects the steel. Cadmium is deposited almost entirely from cyanide baths.

Zinc.—Like tin, much zinc coating of steel is done by hot-dipping. There are numerous applications for electrolytic zinc coatings, however, all for the purpose of corrosion prevention. Like cadmium, zinc protects steel sacrificially. Two types of plating baths are used in zinc deposition, one based on acid zinc sulfate, the other on cyanide. The former is used primarily in the plating of steel wire at high speed; the latter, for larger objects, such as electrical conduit, outdoor hardware, etc.

Silver and Gold.—These metals are deposited exclusively from cyanide baths. Their use is largely for decorative purposes, although heavy deposits of silver serve well as corrosion resistant coatings on tableware.

Since sufficiently bright deposits are rarely obtained on plating, it is customary to buff the silver to the desired type of finish.

Periodic-Reverse Plating.—While normally only direct current is used in electroplating, it has been found that greatly improved smoothness and uniformity of thickness results when a periodic-reverse process is used. As the name implies, the work is alternately plated and de-plated. The technique is particularly applicable to cyanide copper baths and offers promise with many others.

See ELECTROCHEMISTRY; *see* also references under "Electroplating" in the Index volume.

See W. Blum and G. B. Hogaboom, *Principles of Electroplating and Electroforming*, 3rd ed. (1949); A. G. Gray (ed.), *Modern Electroplating* (1953). (H. T. F.)

ELECTROSCOPE, an instrument for the measurement of electrostatic charge. The operation of the instrument is based on the fact that like charges repel each other. The force of repulsion is rendered observable by the deflection of a flexible conductor working against a mechanical restoring force. A common type is a pocket electroscope used for the measurement of radiation exposure of persons working with X-rays, radioactive substances and nuclear energy devices. *See* INSTRUMENTS, ELECTRICAL MEASURING; ELECTRICITY: *Electrostatics.* (M. G. I.)

ELECTROSTATIC GENERATOR, a machine which converts mechanical power into high-voltage electrical power through the interaction of electric fields. Most electrostatic generators accomplish this conversion by repeatedly moving electric charge, frictionally or conductively deposited on a suitable conveyor, from a region at low potential to an insulated terminal at higher potential. Other generators are of the variable capacitance type in which the insulated electrode of a cyclically variable condenser receives charge from ground during intervals when the capacitance is increasing and then delivers this charge to a terminal at higher voltage during intervals of diminishing capacitance.

The modern electrostatic generator of the Van de Graaff belt-conveyor type described below is especially effective in producing high constant potentials in the 1,000,000 to 10,000,000-v. range; these are needed in nuclear research for the acceleration of positive ions used in bombarding atomic nuclei. Many atomic research laboratories employ such high-voltage electrostatic accelerators, mounted within steel tanks containing compressed-gas insulation, for the precise determination of nuclear energy levels. Electrostatic constant-potential generators are also used for the acceleration of steady streams of electrons, usually in the energy range from 500,000 to 5,000,000 ev. Such energetic electrons are used to control superficial cancerous growths, to destroy bacteria and viruses in injectable pharmaceutical products such as sutures and antibiotics, to increase the storage life of food products and to improve the properties of certain plastic materials such as polyethylene through cross-linking or polymerization of their long molecules. When megavolt electrons impinge on a tungsten or gold target, a portion of their kinetic energy is converted to penetrating X-rays used in the treatment of deep-seated malignant tumours and in the radiographic examination of thick structural sections. Electrostatic generators are also used in flashover and breakdown studies of dielectric material and in laboratory studies of spark-over and corona in air and other gases.

Principles of Electrostatic Belt Generators.—For the production of high constant voltages the frictional and influence machines of the 18th and 19th centuries have been replaced by a more direct and effective type of electrostatic generator originated by Robert J. Van de Graaff in 1929. As shown diagrammatically in fig. 1, the generator consists of a well-rounded high-voltage terminal isolated from ground on an insulating column, and a charge-conveying system consisting of one or more belts of insulating material running between this terminal and ground. Electric charge of one polarity is sprayed by corona discharge on the belt at its grounded end and is transported by the belt into the hollow terminal, where it is removed. Within the terminal, charge of the opposite polarity may be deposited on the belt and transferred to ground. At any instant the terminal potential (V) is $V = Q/C$, where Q is the stored charge and C the capacitance of the terminal to ground.

The terminal will maintain that equilibrium potential at which the charging current delivered by the belt is just balanced by the load current. The latter may consist of the currents through the accelerating tube and down the column which supports the terminal, as well as the corona current from the terminal itself. These latter two components of load current increase with termi-

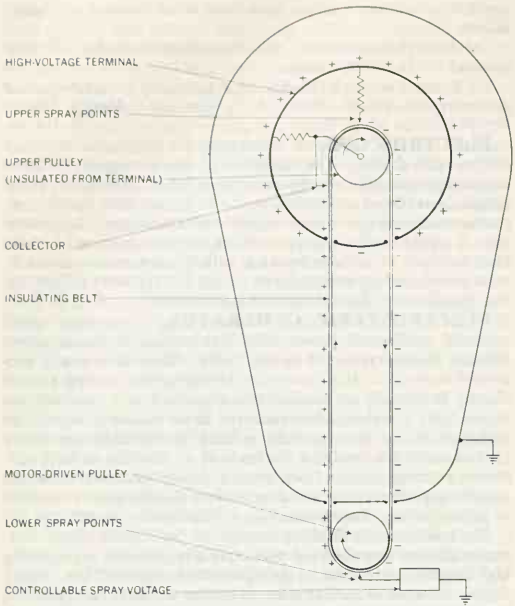


FIG. 1.—SCHEMATIC DIAGRAM OF THE VAN DE GRAAFF ELECTROSTATIC BELT GENERATOR

nal voltage and ensure its stability around the equilibrium value. By adjusting the current carried by the belt and the load current, the terminal voltage may be established at any desired value up to a predetermined maximum. An essential feature of these generators is that the function of belt-charging and charge-removing is performed within the field-free regions below the ground plane and within the high-voltage terminal. In this way, these functions are made independent of the potential difference which may exist between the terminal and ground.

The maximum voltage which can be maintained on the terminal depends only on its geometry and the dielectric strength of the surrounding insulating medium. The terminal shape must be such that, at the desired operating voltage, the surface gradient at all points is well below the ionizing or breakdown value. Ideally considered, an isolated conducting sphere is the most favourable electrode shape and would maintain a voltage Er, where E is the permissible gradient for the insulating medium surrounding the sphere and r its radius. The necessity, however, of physically supporting the high-voltage terminal from ground, as well as of terminating an electrically charged belt and often an acceleration tube inside the terminal, makes it desirable to modify its spherical shape in the region of the column.

Likewise, the supporting column must be modified because of the field-distorting influence of the charged belt. This is particularly true in compressed-gas-insulated generators if maximum advantage of the improved insulating medium is to be obtained. It is best accomplished by constructing the column of a series of spaced conducting planes separated by insulating supports. Corona or resistances between these equipotential planes are used to divide the terminal potential uniformly along this column. In this way the electric field in the region of the belt and tube and in the vicinity of the high-voltage terminal and the ground plane is kept more uniform.

Belt-Charging Methods.—Flat endless belts of insulating material have thus far proved the most effective charge-conveyors for both air and compressed-gas-insulated electrostatic generators. Such belts are usually made of multiple rubber fabric, though woven cotton, nylon and other materials possessing good mechanical and electrical properties have been used. The belt travels between the ground plane and the interior of the high-voltage terminal at constant speed, commonly between 3,000 and 8,000 ft. per minute.

Referring again to fig. 1, electric charge is sprayed on the belts at the grounded end from a row of corona points extended across the width of the belt and directed at the pulley. A corona spray voltage maintained between these points and the pulley results in gaseous ionization at the points and causes a transfer of electric charge toward the pulley. This charge is deposited upon the intervening insulating belt, neutralizing any residual charge of opposite polarity remaining from the downward belt run, and leaving a net charge which is conveyed to the terminal. The charge sprayed on the belt in this manner is controlled by regulating the voltage between the points and the pulley and has a maximum value which depends on the insulating strength of the gaseous medium in which the belt operates. The belt charge may be removed within the terminal from the ascending belt by a corona-point collector connected to the terminal. In this arrangement only the upward belt run is utilized. The generated electric power is equivalent to the mechanical power required to move the belt charge to the terminal potential.

Within the high-voltage terminal a second source of corona-spray voltage may be used to neutralize the charge brought up and to charge the descending run of belt with opposite polarity. A simple self-inducing arrangement may also be used to perform this function. This is accomplished by insulating the upper pulley and maintaining it at a potential higher than that of the terminal by requiring the charge from the collector and pulley to flow through a resistance as in fig. 1. The resistance between the upper pulley and the terminal is so adjusted in value that the current carried on the downward run of belt corresponds approximately to the current arriving on the ascending run. Thus both belt runs can be effective in transferring charge continuously between ground and high-voltage terminal.

Permissible Belt-Charge Density.—The maximum charge density which can be carried by a belt in air at atmospheric pressure is that which will result in a gradient of about 30 kv. per centimetre. This maximum charge density can be shown to be 5.3×10^{-9} coulombs per square centimetre; about 40% to 60% of this value is commonly attained in generators which operate in air at atmospheric pressure. The current of an air-insulated generator is equal to the product of the charge density and the speed, width and number of belts. In the operation of multiple belts, it is necessary that adjacent belt runs move in opposite directions and carry charge of opposite sign.

Compressed-Gas-Insulated Generators.—The most effective way of increasing the voltage and current capacity of an electrostatic generator is to increase the dielectric strength of the insulating medium in which it operates. It had been known since the beginning of the 20th century that the insulating strength of a uniform-field gap increases almost directly with gas pressure over the range from 1 to 10 atm. and thereafter more slowly. This led in 1932 to the use of compressed gases such as air, nitrogen, and carbon dioxide for the insulation of electrostatic belt generators intended for operation at higher voltages and currents. Nitrogen and carbon-dioxide mixtures at pressures up to 25 atm. are now commonly used. The power rating of electrostatic generators, in principle, increases as the square of the dielectric strength of the surrounding gas.

That certain compounds in the gaseous state of chlorine, fluorine, and other negative atoms have a higher insulating strength than air or nitrogen was suggested in 1889. Since World War II, the compounds Freon ($CC1_2F_2$) and sulfur hexafluoride (SF_6) have been found to be particularly favourable as insulating media and are occasionally used either alone or with the addition of nitrogen. These gaseous compounds require only one-third the pressure to attain the dielectric strength of air. Their use may be complicated by gradual decomposition which occurs under conditions of ionization and sparking.

Other Conveyer Types.—Although the flat belt traveling in a gaseous medium has thus far been the most practical charge-

conveyer for high-voltage electrostatic generators, other carriers have been considered and investigated. Among them are the use of an insulating fluid such as oil circulated through an endless insulating pipeline by a pump, a whirling insulating disk either carrying charge directly or carrying charged metallic carriers and charged particles such as dust, droplets or metallic spheroids. The current capacity of these various carriers depends, as with the belt, on the dielectric strength of the medium itself, on the surface area enclosing the charged carriers and on the linear speed of the carriers with respect to the collecting terminal.

A rotating cylinder type of electrostatic generator insulated in compressed hydrogen was developed by N. J. Felici of the University of Grenoble, France. As shown in fig. 2, the upper high-voltage electrode (+U) and the lower charge-inducing electrode (−V) are mounted on a stationary cylinder which has just sufficient conductivity to distribute the electric field fairly uniformly between them. The insulating rotating cylinder has positive electric charge bound on its inner surface by corona induced from the lower points by the small negative potential on the lower electrode. This charge is carried to the region of the upper electrode where the corona points neutralize this positive charge and leave an excess of negative charge to be carried back to ground. Such rotating cylinder electrostatic generators are relatively compact but limited to considerably lower voltages than the belt-type machine.

Variable Capacitance Generators.—Several forms of A.C. and D.C. electrostatic generators have been devised which are essentially variable condensers, usually with interleaving metallic plates. Such generators may attain considerable output power at moderately high voltage if insulated in superior media such as compressed gases or high vacuum. Vacuum-insulated electrostatic generators appear promising for the power sources in outer space vehicle propulsion where this insulating medium is freely available. High vacuum offers superior insulation when voltages below 100 kv. are applied to single gaps. Higher voltages are preferably divided among several gaps to avoid a breakdown process which depends on the gap voltage. Thus 100 kv. can be insulated across a 1 mm. gap between parallel polished steel electrodes immersed in high vacuum but it is difficult to insulate ten times this voltage across a 100 mm. gap.

The forces which may be involved in compressed-gas or vacuum-insulated electrostatic generators are illustrated by the following example: Consider two metallic plates of 100 sq.in. area which are parallel and insulated from each other. If a constant voltage is applied to produce between them the highest electric gradient that can be supported in atmospheric air, 30 kv. per centimetre, the electrostatic force of attraction acting on the plates will be just over ½ lb. If the plates are immersed in high vacuum and the gradient brought to 1,000 kv. per centimetre, the electrostatic force would be over 600 lb.

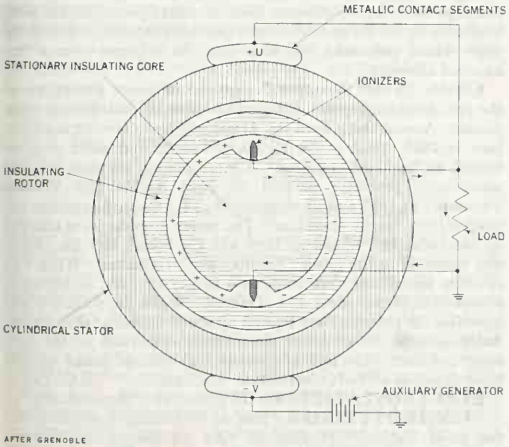


FIG. 2.—ELECTROSTATIC GENERATOR WITH INSULATING CYLINDRICAL ROTOR

An interesting A.C. electrostatic machine consists of a simple variable capacitor of balanced construction across which a high alternating voltage is impressed. Such a machine develops power as a motor during intervals of increasing capacitance, and as a generator during intervals of diminishing capacitance. If the rotor is driven synchronously with the voltage, and at such a phase angle that the capacitance is increasing when the voltage is high, then the generator power will exceed the motor power.

A D.C. electrostatic generator of the interleaving metallic type (fig. 3) may be arranged so that the rotor-stator capacitance varies cyclically between a minimum value C_0 and a maximum value C_m. The rotor is insulated from ground and, to generate negative electric power, maintained at a positive inducing voltage V relative to ground by an auxiliary supply. The stator is connected to the junction of two rectifier tubes (T_1, T_2) which are connected in series across the D.C. line at $-E$ volts. The cycle for negative power generation is as follows: When C becomes equal to C_m, T_2 ceases to be conducting. At this point the stator is at ground potential and holds the negative charge $Q_m = C_m V$ bound by the inducing voltage V. Full line voltage E exists across T_1. As the rotor turns, the capacitance C diminishes. As neither tube now is conducting, the voltage across C increases, the stator becoming more and more negative relative to ground. When the stator attains line potential, T_1 becomes conducting, and further movement of the rotor causes the charge to leave the stator and to flow onto the line. T_2 now is withstanding full line voltage E, and the charge left in the generator system at this point is $Q_0 = C_0\ (E + V) + (C_s + C_v)\ E$, where C_s is the capacitance of the stator to ground, and C_v is the capacitance of T_2 to ground. The average output power is $P_a = \frac{4.63 \times 10^{-15}\ Knsp\ (r_1^2 - r_2^2)\ VE}{d}$ watts, where $K = \frac{1 - C_0\ (E + V) - (C_s + C_v)\ E}{C_m V}$, n is the rotations per minute, s is the number of rotor plates, p is the number of poles, d is the separation between rotor and stator in centimetres, and r_1 and r_2 are the radii of the rotor in centimetres. In a practical machine K would be about 0.7 and the stator would be in phased sections to distribute generation over the whole cycle. Such machines could in principle attain power densities comparable to electromagnetic machines if operated in superior insulating media such as high vacuum at gradients above 500 kv. per centimetre. (*See* also ACCELERATORS, PARTICLE.)

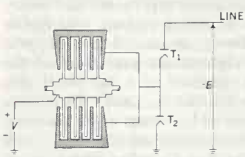


FIG. 3.—CONSTANT POTENTIAL VARIABLE-CAPACITOR GENERATOR FOR OPERATION IN COMPRESSED GAS OR HIGH VACUUM

Historical Development of Electrostatic Generators.—Electrostatic phenomena were the first in the history of electrical science to be observed and investigated. In the 7th century B.C. Thales of Miletus knew that amber, when rubbed, had the property of attracting light objects. Otto von Guericke of Magdeburg, Ger., produced the first generator of electricity in about 1663 by holding his hand against a rotating sphere of sulfur. In the following hundred years a variety of improvements were made on this basic frictionally excited conveyer-type generator. Isaac Newton suggested (1675) a glass globe instead of a sulfur sphere, Francis Hauksbee built a machine with two concentric glass globes rotating in opposite directions, Stephen Gray in England and Charles DuFay in France enlarged on Guericke's discovery of electrical conduction along a linen thread by studying the conductivity of other materials including metals, while George Bose at the University of Leipzig, Ger., found that the addition of a large prime conductor increased the energy of the sparks from such machines. Another increase in the power of electrostatic machines was made when Pieter van Musschenbroek of the University of

Leiden, Holland, invented the Leyden jar around 1746. Meanwhile, Johann Winkler, also of Leipzig, had replaced the hand with a leather pad, which led to the suggestion around 1740 by a Scottish Benedictine monk, Andreas Gordon, that a glass cylinder would be more practical than a globe. The suggestion to use a pointed rod to collect charge from the spinning globe was made by Benjamin Wilson in 1750. The single point gradually became a comb of points. Benjamin Franklin's interest, which resulted in his single fluid theory of electricity, was aroused in 1744 by a popular lecture on the electrical experiments in Europe. The last major improvement in the frictional electrostatic machines was the use of glass disks, attributed independently to Jesse Ramsden (1768), Ingenhousz, and Sigaud de la Fond. John Cuthbertson of Amsterdam and London is credited with building in 1784 the most powerful frictionally excited disk machine of the 18th century, using two 65-in. disks in parallel.

Alessandro Volta's invention of the electric pile around 1800 opened the development of electromagnetic experiment and theory by providing a low-voltage source of steady current. An earlier invention of Volta's, the electrophorus, announced in 1782, profoundly influenced the design of electrostatic generators of the 19th century. The electrophorus consisted of a flat metal disk which could be placed upon a smooth dielectric sheet. To operate the device, the metal disk was removed by means of its glass rod handle and the surface of the dielectric rubbed vigorously with fur. The metal disk was then replaced and touched with a finger. This permitted the induced charge on the top of the disk to be neutralized. Now when the disk was lifted, it carried an electric charge opposite to that bound on the dielectric. This operation could be repeated many times without recharging the insulator.

During the 19th century hundreds of methods were proposed for combining the electrophorus principle with a mechanical charge transfer process. Among the most successful was the "doubler" invented by William Nicholson in 1788, an improvement thereon by C. F. Varley in 1860, the two varieties of induction machines introduced independently by August Toepler and Wilhelm Holtz in Germany around 1865 and the Wimshurst machine invented in 1882.

Wimshurst Generator.—While most of the 19th century influence machines included some elements, such as disk or belt or pointed combs, which are retained in the modern electrostatic generators discussed above, the Wimshurst generator is almost the only one still in some use mainly as lecture demonstration apparatus. This generator uses two contrarotating glass disks on a common shaft carrying on their outer sides many separated sectors of metal foil. Across the outside face of each disk is a stationary neutralizing rod whose metal brushes momentarily connect two diametrically opposite sectors. The two neutralizing rods are at right angles to each other (fig. 4). In addition, there are two sets of pointed combs, occupying an intermediate position, which simultaneously remove by corona the electric charge from the pairs of sectors passing them. The combs may be connected to the inner linings of two Leyden jars to increase the stored energy of the system. The outer linings of the jars would be connected to ground.

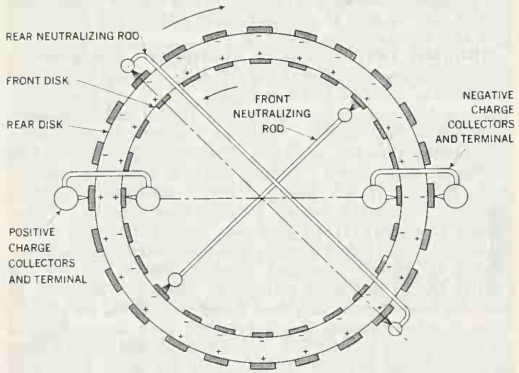


FIG. 4.—PRINCIPLES OF WIMSHURST ELECTROSTATIC DISK GENERATOR (1882)

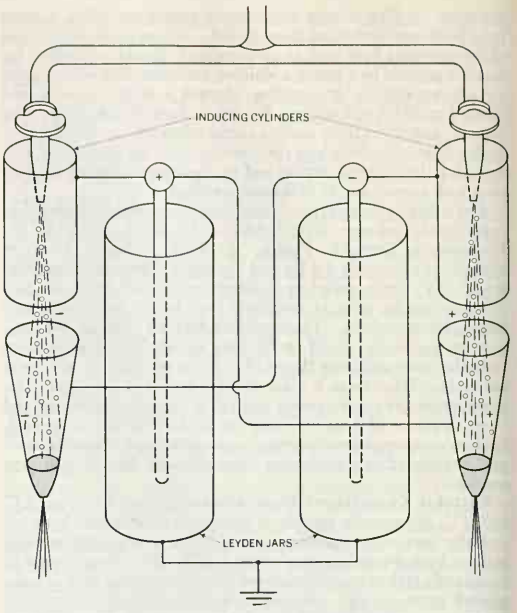


FIG. 5.—WATER-SPRAY INDUCTION GENERATOR OF LORD KELVIN (1867)

To understand the operation of the Wimshurst generator, suppose that two diametrically opposite sectors on the rear disk are positively and negatively charged, respectively. Then, when they pass the front neutralizing rod the two sectors thus momentarily connected on the front disk will continue with equal and opposite charges induced and retained upon them. The rear sectors travel on to the collecting combs to which they yield most of their charge, but the front sectors in a similar manner inductively charge two rear sectors before they yield their charge to the collectors. Thus each metal sector is in turn a field-inducing electrode and a charge-conveyer. The Wimshurst machine was generally more reliable and capable of higher currents than its contemporaries. Its basic weakness lay in the poor electrode geometry involving thin metal edges which prevented the attainment of voltages over a few hundred kilovolts.

Kelvin "Water Dropper."—Several influence generators of the past have employed moving particles as the charge-conveying means. Among these were the Armstrong hydroelectric machine, built in 1840, which projected charged steam and water droplets toward an insulated collector. A more famous machine was the water-spray generator devised by Lord Kelvin (Sir William Thomson) in 1867 and illustrated in fig. 5. Gravity provided the motive force for its operation. The water droplets were charged by the inducing cylinder as they left the nozzle and the charge was collected in the lower "Faraday pail" container. When two of these assemblies were connected to Leyden jars, as shown, a charge-multiplying effect was achieved. While the principle was beautiful the potential was limited by spark discharge through the water spray or by excessive electrostatic dispersion of the light spray particles. The lack of effective high-voltage design was the basic defect of all early electrostatic machines. (J. G. Tr.)

ELECTROSTATICS: *see* ELECTRICITY: *Electrostatics.*

ELECTROTYPING is a process for making duplicate plates for relief (letterpress) printing. An electrotype or "electro" serves on the press in place of the original type, halftone or line

cut, or any combination of these. It also can duplicate and replace linoleum cuts, woodcuts and wood engravings.

An electrotype is made by first depositing electrolytically a thin shell of metal on a mold or cast of the original cut or type form and then backing the shell with metal. Electros are more durable than type and cuts and therefore are used for long press runs. As multiple duplicates, they are employed when matter is to be printed several times on the same sheet, or when identical plates are to be sent to a number of printers. Another advantage of electros is that they make it unnecessary to subject expensive type and cuts to wear and damage.

The electrotyping process was first announced in 1839 by Moritz Jacobi, a German working in St. Petersburg, Russia. Thomas Spencer and C. J. Jordan of England, and Joseph A. Adams of the United States, also produced results in the same year. R. Hoe and Co. of New York, about 1865, was probably the first firm to market the various machines used in electrotyping.

The oldest and cheapest electrotyping method is the wax mold process. The type and cuts to be duplicated are locked in a special chase (frame) and pressed into warmed mineral wax. The resulting mold is sprayed with silver or a mixture of graphite and water to make it a conductor of electricity. It is then connected with a negative electric power source and suspended in a tank filled with a solution of copper sulfate and sulfuric acid. A copper bar is attached in the tank to a positive power source. By the process of electrolysis, a shell of copper from .006 to .008 of an inch thick is deposited on the wax mold. The shell is backed with lead and then usually is mounted type high on a wood or metal base.

Nickel is sometimes substituted for copper in the electrolytic bath. It is used for extra-long press runs, or for printing on rough stock that would wear the softer copper. A common practice is to apply a thin plating of nickel to copper electros. Chromeplated copper electros, which provide the toughest surface of all, are employed for the longest runs; *i.e.*, those of more than 400,000 impressions.

Other molding materials are superior to wax but are more expensive. Those most commonly used are lead, plastics and a wax-rubber compound bonded to a sheet of annealed aluminum. Electros can be made in a curved shape to fit high-speed rotary presses by bending them when they are either hot or cold, or by using centrifugal pressure casting.

See ELECTROCHEMISTRY; ELECTROPLATING; PRINTING: *Letterpress Printing.* (J. KA.)

ELEGY, in classical literature a poem composed in elegiac metre, which consists of distichs, or pairs of lines, the first a hexameter (*q.v.*), the second a so-called "pentameter." The "pentameter" actually consists of two hemiepes (—⏑⏑—⏑⏑—), with diaeresis (word-end) between (*see* PROSODY, CLASSICAL). In the first hemiepes, single long may be substituted for double short (–⏑⏑–⏑⏑–). Thus the following four variations (taken from the first elegy in Propertius' third book) are possible:

— ⏑⏑— ⏑⏑—‖—⏑ ⏑—⏑ ⏑—
non faciet capiti dura corona meo

— — — ⏑⏑—‖— ⏑⏑ — ⏑ ⏑—
exactus tenui pumice versus eat

— ⏑ ⏑ — —‖— ⏑ ⏑—⏑ ⏑ —
quove ped(e) ingressi quamve bibistis aquam

— — — — —‖— ⏑⏑ —⏑ ⏑ —
in vestrum, quaeso, me sinit(e) ire nemus

Writers in later antiquity (*e.g.*, Horace, *Ars poetica* 75–76; Ovid, *Amores* III, ix, 3) associate this metre with dirges. Pausanias (X, vii, 3) mentions mournful elegies composed by the Argive, Echembrotus, in the 6th century B.C., but the earliest extant elegiac dirge is in Euripides' *Andromache* (103–116; 5th century B.C.). Surviving fragments of early Greek elegy (7th–6th century B.C.) deal from a more or less personal angle with various subjects: martial exhortation (Callinus, Tyrtaeus, *q.v.*), politics (Solon, *q.v.*), personal matters, moralizing, love, conviviality (Archilochus, Mimnermus, Theognis, *qq.v.*, etc.). In the late 5th century, Antimachus of Colophon used elegiac metre for a collection of mythological love stories called *Lyde*. Narrative elegy became a favourite genre in the Alexandrian period. Its practitioners included Philetas, Hermesianax (*qq.v.*), Phanocles, Alexander Aetolus. Callimachus (*q.v.*), in his *Aetia,* recounted the origins of cults, etc. ("etiological" elegy). Of one of Callimachus' elegies, the *Lock of Berenice,* substantial fragments (fragment 110 in the edition by R. Pfeiffer, 1949–53) survive and a Latin translation by Catullus (lxvi). From the 7th century B.C. until the 9th century A.D. elegiac metre was used for Greek epigrams. In Latin also epigrams were composed in elegiacs, notably by Catullus (*q.v.;* 84?–64 B.C.) and Martial (*q.v.; c.* A.D. 40–*c.* 104). Catullus also composed longer elegies (notably lxiii, lxxvi). Latin elegy reached its zenith in the love poems of Tibullus (*q.v.;* d. *c.* 19 B.C.) and Propertius (*q.v.; c.* 50–*c.* 15 B.C.), for which no Greek prototype is known.

The works of Cornelius Gallus (*q.v.; c.* 70–26 B.C.), who probably contributed much to the development of Roman love elegy, are lost. Ovid (*q.v.;* 43 B.C.–A.D. 18) employed elegiacs of extreme technical refinement not only for personal poems (*Amores, Tristia,* etc.) and for etiological poetry (*Fasti*), but also for mock-didactic poems (*e.g., Ars amatoria*), and fictitious epistles from mythological heroines to absent husbands or lovers (*e.g., Heroides*). (L. P. E. P.)

Modern Elegy.—From the 16th century in England the word elegy has been used to mean a poem or song in lamentation of death. Theocritus, the inventor of bucolic or country poetry, with his laments for dead shepherds, is undoubtedly responsible for the pastoral convention in English elegy, but the elegiac metre has not been used. Edmund Spenser wrote *Astrophel. A Pastoral Elegy Upon the Death of the Most Noble and Valorous Knight Sir Philip Sidney* on the death of his friend at the battle of Zutphen in 1586, and companion to *Astrophel* was the "Lay of Clorinda"; Spenser also published *Daphnaïda. An Elegie Upon the Death of the Noble and Vertuous Douglas Howard* (1591). Milton himself called *Lycidas* (1638) a monody in which "the Author bewails a learned Friend [Henry King], unfortunately drowned."

Alexander Pope's "Verses to the Memory of an Unfortunate Lady" (1717) and Thomas Gray's *Elegy Written in a Country Churchyard* (1751) on the wider theme of human mortality were followed by Shelley's *Adonais* (1821) in memory of the young Keats who died in that year in Rome; Tennyson's *In Memoriam* (1850), evoked by the untimely death of Arthur Hallam; and Matthew Arnold's "Thyrsis" (1867), lamenting the death of his friend Arthur Hugh Clough. Robert Bridges experimented in prosodic form in "Elegy on a Lady Whom Grief for the Death of her Beloved Killed" and wrote at least one more elegy, "The Summer-house on the Mound."

Among French elegists are Ronsard, and Lamartine who may be remembered for "Le Lac." Camões used the form in Portuguese, and Ramón Jiménez in Spanish. In Russian, Alexander Pushkin's "Elegy on the Portrait of F.-M. Barclay de Tolly" should be mentioned.

In German literature, elegy as poetry of lamentation does not exist, but Goethe's Roman elegies followed Ovid's in form and theme.

See also references under "Elegy" in the Index volume.

(J. W. T.)

See A. Couat, *Alexandrian Poetry* (1931); introduction to H. E. Butler and E. A. Barber, *The Elegies of Propertius* (1933).

ELEKTROSTAL, a town of Moscow *oblast* of the Russian Soviet Federated Socialist Republic, U.S.S.R., lies 58 km. (35 mi.) E. of Moscow on the branch line to Noginsk from the main railway to Gorki. Pop. (1959) 96,922. The town was developed in the 1920s and is one of the most important industrial centres in the *oblast*. It specializes in the production of high grade steels and rolled steel and in heavy engineering. (R. A. F.)

ELEMENTARY EDUCATION refers generally to the first stage of formal education. The term "primary education" is used in many countries. Historically, both terms have been used in the western world to designate the vernacular and terminal program of education for the masses—a program which paralleled secondary education but that had little or no organic connection with it. However, both terms have come increasingly to refer to

the first step on the educational ladder.

Elementary education was once defined as the development of the knowledge and attitudes that should be possessed by all citizens, traditionally identified as reading, writing and arithmetic but including geography and history. Children normally enroll in elementary (or primary schools) at age five or six. The length of the period of elementary education varies from five to eight years, depending in part upon the purposes of the school and in part upon the accepted definitions of elementary and secondary education. For the history and development of national systems of education *see* EDUCATION, HISTORY OF. *See* also *Education* sections of separate articles on various countries, states and provinces, as ARGENTINA; INDIANA; QUEBEC; etc.

In the United States, compulsory schooling most commonly begins at the age of 6 (when kindergarten is provided, attendance begins at age 5) and elementary education describes the schooling provided through the ages of 12, 13 or 14. Primary education usually refers to the education provided in the first three grades of an elementary-school system comprising six, seven or eight grades. In England and Wales, on the other hand, compulsory schooling begins at the age of 5 and "primary education" describes the schooling provided through the age of 11. The last three years encompassed in the definition of elementary education used here are included in secondary education in England, Wales, Scotland and Northern Ireland. In the United States and Canada eight-grade elementary schools embrace the full span of years included in the definition; in the United States, however, most school children reaching the age of 12 or 13 move up into a junior high school.

DEVELOPMENT OF ELEMENTARY EDUCATION

Historically, elementary education has served to preserve the most cherished tribal customs, to inculcate religious doctrines and to promote national causes. In primitive tribes education has been a conservative force, seeking to preserve the traits that give a tribe its unity. Every parent is a teacher, thus assuring the tribe that no child goes untaught. The very thoroughness of the system allows no flexibility, no room for counter teachings and, therefore, little opportunity for improvement. Down through the centuries parental education for individual and group survival predominated over the greater part of the earth and continues to be prevalent in underdeveloped countries, particularly in large parts of Africa and South America. But a few isolated cultures —such as the Incas, who did not survive, and the Samoans, who still thrive—matured beyond mere primitive survival. They educated the young in the folk arts and more elaborate social customs that characterize advanced peoples.

The relatively mature cultures of ancient Egypt, Babylon, Greece and Rome help little in understanding the development of modern systems of elementary education. The emergence of alphabets and of trades and professions called for formal education and for special teachers. But such education was for a favoured few.

These great civilizations and their learned men provided some of the content of modern elementary-school instruction. They provided, too, certain concepts of human value that underlie attempts to provide elementary education for all. But no systems of formal education embracing more than a relatively small number of children evolved in these countries.

In Europe, during the centuries preceding the Renaissance and the Reformation, the spark of education was kept alive in monasteries and cathedral schools. By the 15th century universities were the most important educational institutions, and elementary and secondary education, dominated by the church and with Latin the language of instruction, was mainly for the privileged. Meanwhile, in much of China, India and Japan, formal education for the chosen few prevailed. Among the masses, rudimentary education was largely a parental affair. Education of the masses as a guiding idea was yet to come.

Struggle between and reform within Protestant and Roman Catholic churches marked educational developments in Europe during the 16th and 17th centuries. It is not at all surprising that Martin Luther, seeking church reform, proposed compulsory primary instruction in order that everyone might learn to read the Bible. Religious groups have continued to be influential forces —sometimes with, sometimes against, national and local efforts to finance schools, select subject matter and provide instruction.

The Kindergarten Movement.—Europe during the 18th and 19th centuries was a fruitful source of educational thought, stimulating the movement toward universal elementary education and reforming teaching practices. Rousseau, in France, urged that children be permitted to develop naturally as children, free from restricting adult standards. The ideas and practices of Robert Owen in Great Britain, J. H. Pestalozzi in Switzerland and Maria Montessori in Italy exerted tremendous influence on the kindergarten, or infant-school, movement of the 19th century. Owen, managing partner of cotton mills at New Lanark, established a school for the young children of his employees in 1816. His two ideals—pleasant, healthful conditions and a life of interesting activity—had a profound influence in spite of many setbacks. The circumstances of the industrial revolution in England tended to encourage the provision of schools for young children, known as infant schools, because early legislation prevented their employment while the factories took their parents and older brothers and sisters away from home for long hours. In 1824 the London Infant School society was founded and the Glasgow Normal seminary, in which teachers were trained to serve in infant schools, was established in 1836. Importance was attached to memory drill and moral training, with restricted playgrounds and fixed galleries in large classrooms, reflecting an attitude quite different from that of Owen. In 1837, however, the Home and Colonial society was founded to train teachers in the methods advanced by Pestalozzi, who saw in education a means of elevating the masses. He and his pupil, Friedrich Froebel (*q.v.*), believed in leading children to self-understanding and not imposing the ideas of adults. Pestalozzi's demonstrations of the educability of retarded children led to provision for both physically and mentally defective children in the educational systems of many countries. In 1837 Froebel opened at Blankenburg, in Thuringia, "a school for the psychological training of little children by means of play and occupations"; applying to it the name "kindergarten" in 1840, he sought to convey the impression of an environment in which children grew freely like plants tended in a garden. The kindergarten plan to meet the educational needs of children between the ages of four and six or seven through the agency of play gained widespread acceptance. During the 25 years after Froebel's death in 1852 kindergartens were established in leading cities of Austria, Belgium, Canada, Germany, Great Britain, Holland, Hungary, Japan, Switzerland and the United States. In Great Britain the term infant school was retained for the kindergarten plan. In the United States a private kindergarten was established in Wisconsin in 1855 and kindergartens were added to the St. Louis, Mo., public-school system in 1873, the plan becoming, for children below the traditional school entrance age of six years, a generally accepted first unit of the elementary schools. In Canada, where the first public kindergarten was opened in 1855, in Ontario, it is regarded as a preschool educational institution and is not a part of the elementary system. For educational programs for younger children *see* PRE-ELEMENTARY EDUCATION.

Humanism.—The thought and practices of these educators, together with the ideas of men working in other geographic, social and political arenas, characterized the humanistic reform movement throughout most of Europe. They influenced education around the world. Later, however, the leading figure in the humanistic tradition was an American, John Dewey. His ideas of problem-solving, of making education alive and lifelike, of using children's interests in teaching, were developed and tested in his experimental school at The University of Chicago. His later writing and teaching at Columbia university spread his views around the world. Dewey himself gave three names to his philosophy: pragmatism, humanism, experimentalism. His long life, spanning four decades of the 19th century and five of the 20th, coincided with a growing scientific approach to studying educational problems. World-wide research and child study began to provide teachers with a wealth of data about individual differences

among children, the complexity of learning processes, testing procedures, techniques of classroom management, and so on. The view of children as valuable in their own right, of beginning with the real and moving to the abstract in learning, of leading rather than telling in teaching, became fundamental principles of elementary education.

Two great struggles mark the development of elementary education during the past three centuries. The first has been to establish the principle and then the fact of universal elementary education. The second has been to establish equality of opportunity in moving from elementary to upper units of the educational system. Neither struggle has yet been won.

ELEMENTARY EDUCATION FOR ALL

The expansion of knowledge and the advancement of science are together forcing overhaul in the educational systems of advanced countries around the world. But what might be called an explosion of the human spirit, in many ways an extension of the humanistic movement that stimulated educational reform in previous centuries, is bringing educational interest to places where scarcely any existed before. Nearly half (approximately 250,000,000 of 550,000,000) of the world's boys and girls between the ages of 5 and 14 are yet to be enrolled in any school, public or private.

The world's provisions for elementary education are expanding more rapidly than its population. The most rapid educational expansion is taking place in countries that have literacy as their immediate goal. In these countries—most of them in Africa, Asia and South America—elementary education is viewed necessarily as terminal for nearly all children. Unfortunately, these are also the countries experiencing the most rapid increases in population so that the problems of providing elementary education are aggravated and facilities lag far behind needs. In the educationally advanced countries of northern, western and central Europe and of northern America, where the literacy level is above 95%, and in the countries of Oceania, where literacy is higher than 90%, the population is increasing less rapidly. Expansion in elementary education is proceeding more slowly in these countries, where the emphasis has shifted to expanding free systems of preschool education, improving the relationship between upper and lower units and eliminating educational disadvantages arising from place of residence, wealth and, occasionally, family position.

Problems.—The problem of providing education for all the world's children is in part a problem of providing schools, teachers and at least rudimentary equipment and materials. It is in part, too, a problem of keeping children in school long enough for some permanent value to result. Four or five years of schooling generally are considered necessary to produce functional literacy in the average individual. Yet a survey conducted in the 1950s revealed that in Brazil, for example, only 15% of those enrolled in elementary schools remained until the fifth grade. For large numbers of children to attend school for only a year or two constitutes a shocking waste of teaching and material resources in countries desperately short of both.

The practice of repeating grades likewise is wasteful. A concentration of enrollment in the first grade or two is characteristic of underdeveloped school systems, partly because of the number who remain no longer and partly because of the number who fail to meet requirements for promotion. But even in some sections of western Europe 40% of the pupils repeat a grade or more before completing the elementary unit. In the United States, the amount of grade repetition has been drastically cut from the high levels prevailing early in the 20th century. Some school systems still retain up to 10% of a class, particularly in the lower grades. Most schools retain a few pupils each year but the practice of repeating more than one grade is rare. Several hundred schools across the United States are experimenting with a nongraded plan designed to eliminate the century-old graded structure. In them, a primary unit replaces the first three grades and, less often, an intermediate unit replaces the upper grades of the traditional graded system. Children are encouraged to move at more individualized rates of speed and the problem of grade-to-grade repetition is largely eliminated.

Education for Girls.—Enrollment of girls in elementary schools around the world is increasing steadily, especially in those countries where literacy is the primary aim. Countries with high percentages of children enrolled almost invariably have correspondingly high enrollments of girls. On the other hand, in the early 1960s girls represented less than 12% of the total primary school enrollment in such African countries as Ethiopia, Mali and Mauritania. (*See* also COEDUCATION; WOMEN, EDUCATION OF.)

Costs.—It is extremely difficult to make world-wide statistical comparisons of virtually any aspect of elementary education. Cost figures must be translated into a common currency and even were this done on a broad scale they reveal little about relative purchasing power or expenditures in relation to ability to pay for schools. Some system for computing a ratio between actual provision for education and ability to provide is needed for systematic appraisal of world-wide need. The comparative description and classification of school systems and practices on an international basis lies ahead as a challenge to educators.

Obviously, there is a profound relationship between the educational provisions of a country and both its economic resources and its willingness to use these resources for educational purposes. One would expect, therefore, to find fewer pupils per teacher in well-developed countries but this is not the case. The average ratio in elementary schools, regardless of the relative development of countries, is about 34 to 1. In Denmark, with a well-developed system of elementary education, average class size is 36 or 37 pupils. Many quite primitive countries have ratios of less than 30 to 1. Research into the question of optimum class size is inconclusive. Some school systems in the United States have been experimenting with instructional groups ranging in size from a half dozen pupils to 100 or more, depending on the purpose in view.

Teachers.—World-wide estimates and comparisons of teacher quality are likewise difficult to make. In the United States, completion of a four-year college program is virtually the standard minimum preparation, but a few teachers having less preparation than this are permitted to teach. On the other hand, many persons teaching in the primary schools of selected African countries in the 1960s had completed only an eight-year elementary education. Following World War II, there was a critical teacher shortage throughout most of the world, necessitating short-term programs of teacher preparation. By the 1960s countries such as Sweden and Denmark were able to report abatement of the crisis. The problem of teacher supply remained acute, however, over large sections of the earth.

Goals.—The general objective of elementary education around the world is to provide for the intellectual, moral, emotional and physical development of boys and girls. To achieve this objective schools provide in varying degrees instruction in reading, arithmetic, writing, spelling, rudimentary science, history, geography, the arts and various physical skills. Few people yet realize how recently even the most advanced countries established elementary education as the birthright of all rather than the privilege of a few. Even fewer are aware of how much remains to be done in getting the other (or the second) half of the world's children into schools and keeping them there.

For many countries just emerging as independent nations the goal, beyond the immediate goal of literacy, is to replace opportunity based on class and family privilege with opportunity based on individual aptitude. For those countries that now enjoy universal elementary education, two goals lie ahead. The first is to sweep aside those inequalities of educational opportunity that still exist and which prevent many young people from developing their abilities to the full. These inequalities arise out of differences in wealth, differences in local provision of facilities and inadequate techniques for identifying and stimulating the unique talents of individuals. The second goal for these countries is to assist wisely those countries in which elementary education is less fully developed. To assist wisely is to avoid the temptation to impose. Individual countries must be left free to adapt from other countries what they see to be appropriate. Appropriate systems of education are indigenous to the countries in which they develop, shaping and being shaped by the same geographic, economic, spirit-

ual and humanistic factors that shape the character of a people.

See EDUCATION, HISTORY OF; EDUCATION (ARTICLES ON); and references under "Elementary Education" in the Index.

BIBLIOGRAPHY.—John F. Cramer and George S. Browne, *Contemporary Education* (1956); Harry G. Good, *A History of Western Education*, 2nd ed. (1960); Virgil E. Herrick *et al.*, *The Elementary School* (1956); National Society for the Study of Education, *Early Childhood Education*, 46th yearbook, part 2 (1947); *Primary Education* (H.M.S.O., 1959); United Nations Educational, Scientific and Cultural Organization, *World Survey of Education*, vol. 2, *Primary Education* (1958). This last reference provides both a comprehensive world survey and a detailed country-by-country treatment. (J. I. G.)

ELEMENTS, CHEMICAL, are those substances that are not resolved into simpler entities by chemical processes. They are the simple substances of which the hundreds of thousands of varieties of matter are composed and into which all other substances and mixtures may be ultimately divided. The number of elements definitely known by the early 1960s was 103; of these, about 90 occur in nature either chemically free or in combination with other elements. Only ten of the elements—carbon, sulfur, copper, antimony, iron, tin, gold, silver, mercury and lead—were known in the uncombined state in ancient times and, although they were recognized as distinct varieties of matter, they were not classified as elements. In ancient and medieval philosophy, the elements were earth, air, fire and water, the four simple substances of which all material bodies were supposed to be compounded. In prescientific chemistry, the elements were variously enumerated, the usual number being five or six, such as spirit, salt, sulfur, water and earth, or water, oil, air, salt and earth. The modern use of the word dates from the time of Robert Boyle (1661), who described elements as primitive and simple, or perfectly unmixed bodies, that are not made of any other bodies or of one another. From this time onward the term "element" was reserved for material substances, but as early chemistry was chiefly preparative and descriptive, no criteria were immediately available for determining whether substances were elements or compounds. By the middle of the 18th century, however, chemistry became a quantitative science based on the use of the analytical balance, and the needed chemical criteria were established. The new and clarifying ideas on the nature of matter were published by A. L. Lavoisier in 1789 (*see* below), and thereafter, until the later development of physical methods (*e.g.*, X-rays), substances were classified as elements or compounds solely on the basis of their chemical reactions.

The chemical evidence that a substance is a compound is positive and is found either in its decomposition into two or more other substances or in its formation by the chemical union of two or more different substances. For example, water may be decomposed by electrolysis to yield hydrogen and oxygen, two distinct substances; or water, a single substance, may be prepared by the chemical combination of hydrogen and oxygen. On the other hand, if such evidence is lacking for a substance it is presumed to be an element; neither hydrogen nor oxygen has been decomposed by the use of any of the forms of energy (heat, light and electricity) in quantities and intensities employed in chemical processes. There is another distinguishing chemical characteristic of an element. As the weight of a sample of a compound is the sum of the weights of the elements of which it is composed, an element invariably increases in weight upon undergoing a chemical change; that is, an element can react chemically only by combining with another substance, the latter accounting for the increase in weight.

By the use of data obtained in his experimental studies of chemical reactions, Lavoisier was able to prepare the first scientific list of elements, which he published in his renowned book *Traité élémentaire de chimie* (1789). He included in his list 23 actual elements; the alkalies, although he noted that they resembled certain oxides; a number of other oxides; and, with striking inconsistency, "light and caloric," which have neither mass nor any of the other characteristics he had defined for elements.

The discovery of current electricity by A. Volta (1800) supplied an effective means for separating the more active elements from their compounds. Soon thereafter Sir Humphry Davy prepared the elements sodium and potassium by the electrolysis of their respective molten hydroxides (alkalies), proving that the latter were compounds as Lavoisier had suspected. The formulation of the periodic law in 1869 gave further stimulus to the search for new elements by indicating not only their existence but also their probable properties. By the close of the 19th century the number of known elements had increased to 82, and the conception of the nature and structure of matter was beginning to undergo changes.

Isotopes.—The small particles of the elements that are called atoms are not, as was formerly believed, simple indivisible units but are complex structures, each consisting of a nucleus surrounded by an electron cloud. (*See* ATOM.) The nucleus is generally thought to be composed of protons, which carry unit positive charges, and neutrons, which carry no electric charge. The distinguishing mark of an element and its atoms is the number of protons in each nucleus, designated by the term "atomic number." The number of electrons in the cloud surrounding the nucleus is also equal to the atomic number, and the sum of the number of protons and neutrons is the "mass number," an integer approximately equal to the atomic weight.

Early in the 20th century two presumably different radioactive elements were discovered with identical chemical properties and spectra but with different densities and radioactive properties. The two specimens were found to be variations of the element thorium and were termed isotopes (Gr. "equal places") by F. Soddy; they have the same atomic number and therefore occupy the same position, that of the element, in the periodic table but differ in the number of neutrons in the nuclei. In 1912 Sir J. J. Thomson showed that the stable element neon, atomic number 10 and atomic weight 20.183, is a mixture of isotopes by separating the isotopes Ne^{20} and Ne^{22}. Later, a third isotope, Ne^{21}, was separated. Two or more isotopes of all but 22 of the naturally occurring elements were separated and identified; in addition to the natural isotopes more than 1,000 radioactive isotopes have been prepared artificially. The elements of odd atomic number with one or two exceptions consist of not more than two isotopes, while those of even atomic number have more than two natural isotopes. An element can therefore no longer be regarded as a substance that consists of identical atoms but must be thought of as atoms that have the same atomic number. (*See* also ISOTOPE.)

Molecules and Modifications.—Chemically free elements in the vapour and liquid phases consist of discrete units or molecules that may contain a single atom or an aggregate of atoms bonded together by valence forces. The noble gases, which have unique electronic structure with complete valence shells, are monatomic; that is, the molecules each contain one atom and are represented by the formulas He, Ne, Ar, Kr, Xe and Rn (helium, neon, argon, krypton, xenon and radon). Molecules of some of the common gases, such as hydrogen, oxygen, nitrogen and chlorine, are diatomic, with molecular weights corresponding to the respective formulas H_2, O_2, N_2 and Cl_2. The molecules of phosphorus vapour at fairly low temperatures each contain four atoms (P_4) and at higher temperatures two atoms (P_2).

In the solid state there are a few elements whose crystals are composed of identical molecules in which the atoms are bonded more strongly to each other than to those in adjacent aggregates; *e.g.*, sulfur (S_8) and iodine (I_2). Carbon has four valence electrons and can form four covalent bonds. In diamond, one of the crystalline forms of carbon, each atom is bonded to four other atoms placed about it at the four corners of a regular tetrahedron, and each of those in turn is similarly joined to four others, thus binding all the atoms in a crystal that may be regarded as a giant molecule. Germanium, silicon and gray tin, which also have four valence electrons, crystallize with the diamond structure. Solid metals are crystalline but contain no discrete groups of atoms. The metallic bond is unique and not exactly comparable with any one of the other bond types although it possesses some of the characteristics of ionic and covalent bonds.

A number of elements exist in two or more forms or modifications, a phenomenon that is called allotropy. This phenomenon is due either to the formation of different kinds of molecules as in the case of oxygen, O_2 (ordinary oxygen) and O_3 (ozone), or to the formation of two or more different crystalline forms as in the case of carbon (diamond and graphite) or sulfur (rhombic and

The Chemical Elements

Element	Symbol	Atomic number	Atomic weight*	Terrestrial abundance† (parts by weight per 1,000,000)	Cosmic atomic abundance‡ (silicon = 1 x 10^6)	Year discovered
Actinium	Ac	89	(227)§	3 x 10^{-10}		1899
Aluminum	Al	13	26.9815	81,300	9.48 x 10^4	1825
Americium	Am	95	(243)			1944
Antimony	Sb	51	121.75	1	0.246	B.C.
Argon	Ar	18	39.948	0.04	1.5 x 10^6	1894
Arsenic	As	33	74.9216	5	4.0	1649
Astatine	At	85	(210)			1940
Barium	Ba	56	137.34	250	3.66	1808
Berkelium	Bk	97	(247)			1949
Beryllium	Be	4	9.0122	6	20	1797
Bismuth	Bi	83	208.980	0.2	0.144	*c.* 1739
Boron	B	5	10.811	3	24	1808
Bromine	Br	35	79.909	1.62	13.4	1826
Cadmium	Cd	48	112.40	0.15	0.89	1817
Calcium	Ca	20	40.08	36,300	4.9 x 10^4	1808
Californium	Cf	98	(251)			1950
Carbon	C	6	12.01115	320	3.5 x 10^6	B.C.
Cerium	Ce	58	140.12	46.1	2.26	1803
Cesium	Cs	55	132.905	7	0.456	1860
Chlorine	Cl	17	35.453	314	8.85 x 10^3	1774
Chromium	Cr	24	51.996	200	7.8 x 10^3	1798
Cobalt	Co	27	58.9332	23	1.8 x 10^3	1742
Copper	Cu	29	63.54	70	2.12 x 10^2	B.C.
Curium	Cm	96	(247)			1944
Dysprosium	Dy	66	162.50	4.47	0.556	1886
Einsteinium	Es	99	(254)			1955
Erbium	Er	68	167.26	2.47	0.316	1843
Europium	Eu	63	151.96	1.06	0.187	1901
Fermium	Fm	100	(253)			1955
Fluorine	F	9	18.9984	900.0	1.6 x 10^3	1771
Francium	Fr	87	(223)			1939
Gadolinium	Gd	64	157.25	6.36	0.684	1880
Gallium	Ga	31	69.72	15	11.4	1875
Germanium	Ge	32	72.59	7	50.5	1886
Gold	Au	79	196.967	0.005	0.145	B.C.
Hafnium	Hf	72	178.49	4.5	0.438	1922
Helium	He	2	4.0026	0.003	3.08 x 10^3	1895
Holmium	Ho	67	164.930	1.15	0.318	1878
Hydrogen	H	1	1.00797	1,300	4.00 x 10^{10}	1766
Indium	In	49	114.82	0.1	0.11	1863
Iodine	I	53	126.9044	0.3	0.80	1811
Iridium	Ir	77	192.2	0.001	0.821	1804
Iron	Fe	26	55.847	50,000	6 x 10^4	B.C.
Krypton	Kr	36	83.80	9.8 x 10^{-5}	51.3	1898
Lanthanum	La	57	138.91	18.3	2.00	1839
Lawrencium	Lw	103	(257)			1961
Lead	Pb	82	207.19	16	0.47	B.C.
Lithium	Li	3	6.939	65	1 x 10^2	1817
Lutetium	Lu	71	174.97	0.75	0.050	1907
Magnesium	Mg	12	24.312	20,900	9.12 x 10^5	1755
Manganese	Mn	25	54.9380	1,000	6.85 x 10^3	1774
Mendelevium	Md	101	(256)			1955
Mercury	Hg	80	200.59	0.5	0.284	B.C.
Molybdenum	Mo	42	95.94	15	2.42	1782
Neodymium	Nd	60	144.24	23.9	1.44	1885
Neon	Ne	10	20.183	7 x 10^{-5}	8.6 x 10^5	1898
Neptunium	Np	93	(237)			1940
Nickel	Ni	28	58.71	80	2.74 x 10^4	1751
Niobium	Nb	41	92.906	24	1.00	1801
Nitrogen	N	7	14.0067	46.3	6.6 x 10^5	1772
Nobelium	No	102				1957
Osmium	Os	76	190.2	0.005	1.00	1804
Oxygen	O	8	15.9994	466,000	2.15 x 10^7	1772
Palladium	Pd	46	106.4	0.010	0.675	1803
Phosphorus	P	15	30.9738	1,180	1.00 x 10^4	1669
Platinum	Pt	78	195.09	0.005	1.625	*c.* 1735
Plutonium	Pu	94	(244)			1940
Polonium	Po	84	(210)	3 x 10^{-10}		1898
Potassium	K	19	39.102	25,900	3.160 x 10^3	1807
Praseodymium	Pr	59	140.907	5.53	0.40	1885
Promethium	Pm	61	(147)			1947
Protactiuium	Pa	91	(231)	8 x 10^{-7}		1917
Radium	Ra	88	226	13 x 10^{-6}		1898
Radon	Rn	86	(222)			1900
Rhenium	Re	75	186.2	0.001	0.135	1925
Rhodium	Rh	45	102.905	0.001	0.214	1803
Rubidium	Rb	37	85.47	310	6.5	1861
Ruthenium	Ru	44	101.07	0.004	1.49	1844
Samarium	Sm	62	150.35	6.47	0.664	1879
Scandium	Sc	21	44.956	5	28	1879
Selenium	Se	34	78.96	0.09	67.6	1818
Silicon	Si	14	28.086	277,200	1.00 x 10^6	1823
Silver	Ag	47	107.870	0.10	0.26	B.C.
Sodium	Na	11	22.9898	28,300	4.38 x 10^4	1807
Strontium	Sr	38	87.62	300	18.9	1787
Sulfur	S	16	32.064	520	3.75 x 10^5	B.C.
Tantalum	Ta	73	180.948	2.1	0.065	1802
Technetium	Tc	43	(99)			1937
Tellurium	Te	52	127.60	0.002	4.67	1798
Terbium	Tb	65	158.924	0.91	0.0956	1843
Thallium	Tl	81	204.37	3	0.108	1861
Thorium	Th	90	232.038	11.5		1828
Thulium	Tm	69	168.934	0.20	0.0318	1879
Tin	Sn	50	118.69	40	1.33	B.C.
Titanium	Ti	22	47.90	4,400	2.440 x 10^3	1791
Tungsten	W	74	183.85	69	0.49	1783
Uranium	U	92	238.03	4		1789
Vanadium	V	23	50.942	150	2.20 x 10^2	1830
Xenon	Xe	54	131.30	1 x 10^{-6}	4.0	1898
Ytterbium	Yb	70	173.04	2.66	0.220	1907
Yttrium	Y	39	88.905	28.1	8.9	1794
Zinc	Zn	30	65.37	132	4.86 x 10^2	before 1400
Zirconium	Zr	40	91.22	220	54.5	1789

*Based on the weight of carbon-12 as the standard; by permission of International Union of Pure and Applied Chemistry.
†Parts per 1,000,000=grams per metric ton. Most of data taken from K. Rankama and T. G. Sahama, *Geochemistry* (1950).
‡H. E. Suess and H. C. Urey, *Rev. Mod. Phys.*, 28, 53 (1956).
§Value in parentheses denotes mass number either of isotope with longest-known half life or, for berkelium, californium, lawrencium, polonium, prometbium and technetium, a better-known one.

monoclinic). The more general term "polymorphism" is also applicable to the phenomenon involving the crystalline forms. (*See* also MOLECULE.)

Occurrence.—Of the approximately 90 naturally occurring elements, about 30 are found chemically free on the earth; that is, not in chemical combination with any other element. These elements are obviously not very active chemically and include, for example, nitrogen, gold, platinum, copper and the inert gases. However, oxygen, a very active element, occurs uncombined in large quantities in the atmosphere and combined in the waters, the rocks and the minerals of the earth. The only naturally occurring oxidizable materials remaining are in the living things and the materials that were probably derived from the life of the more recent geologic periods; *e.g.*, oil, gas and coal. The surplus free oxygen in the atmosphere is utilized in animal metabolism and replenished by plant processes.

The estimated average amounts of the elements in the earth's crust are given in column 5 of the table. The eight most abundant elements, oxygen (466,000 parts per 1,000,000 or 46.6%), silicon (27.72%), aluminum (8.13%), iron (5%), calcium (3.63%), sodium (2.83%), potassium (2.59%) and magnesium (2.09%), comprise more than 98.5% of the earth's crust.

There appear to be definite relationships between the abundance of an element and the structure of its nuclei, and a number of generalizations were made by W. D. Harkins and others, as follows:

1. Elements of even atomic number are more abundant and have more isotopes than those with odd atomic numbers. In the earth the elements of even atomic number are estimated to be ten times as abundant as those of odd number; in meteorites the ratio is about 70 to 1. Four of the first five elements mentioned above fall in this category, while those elements whose natural occurrence has been in doubt have odd atomic numbers; *e.g.*, technetium (43) and promethium (61).

2. Atoms whose nuclei have even numbers of neutrons are more abundant than those with odd numbers of neutrons. This relationship is even more striking, for it is estimated that 99% of all the atoms in the earth's crust fall within this category.

3. Nuclei with even mass numbers are more abundant than those with odd mass numbers.

To illustrate, the element neon, atomic number 10, consists of three isotopes whose abundances are: Ne^{20} (90.8%), Ne^{21} (0.26%) and Ne^{22} (8.9%), the two nuclear species with even numbers of protons together being nearly 400 times as abundant as the third. Chlorine, atomic number 17, has two stable isotopes, Cl^{35} (75.5%) and Cl^{37} (24.5%), both of which have an even number of neutrons.

The 272 known stable nuclei have been classified according to whether their proton or neutron content is odd or even and thereby placed in one of the four categories listed below:

Protons	Neutrons	Number
Even	Even	162
Even	Odd	55
Odd	Even	49
Odd	Odd	6
	Total	272

There is another significant structural relationship, the excess in the number of neutrons over the number of protons. Approximately 85% of the atoms on the earth have an equal number of neutrons and protons (that is, the excess, sometimes called the

isotopic number, is zero), while for 99.9% of the earth atoms the isotopic numbers do not exceed four.

The estimated cosmic abundances—the numbers of atoms per 1,000,000 (10^6) atoms of total silicon—are given in column 6 of the table. Hydrogen is by far the most abundant element in the universe, accounting for more than 90% of the total number of atoms and for about three-fourths of the mass. Helium is next in abundance, comprising about 7% of the number of atoms and nearly one-fourth of the total mass.

Properties.—The properties of the elements, to a large degree attributed to the electronic structures and sizes of their atoms, are extremely diverse. Helium, for example, has the lowest melting point (−272° C. at 26 atm. pressure) and boiling point (−268.9° C.) of any of the elements, while tungsten has the highest known melting point (3,387° C.) and boiling point (6,700° C.). Densities range between 8.986×10^{-5} g. per cubic centimetre at 0° and 1 atm. pressure for hydrogen and 22.56 g. per cubic centimetre for iridium and osmium. Most of the properties show fairly regular gradations within the main groups of the periodic system, whereas in the periods there is considerable divergence. In the third period, for example, the melting points vary from 97.9° C. for sodium up to 1,414° C. for silicon and down to −102° C. for chlorine.

Classification.—The elements may be classified into three general types on the basis of their properties and their structure: inert gases, metals and nonmetals.

Noble Gases.—The six elements typified by neon have very little chemistry since the outer valence shells of all of them are filled with electrons, two electrons in the case of helium and eight in the others.

Metals, by far the most numerous (65), are generally characterized by the fact that their valence electrons, which are relatively few in number, are lost with more or less ease yielding positive ions. Their oxides and hydroxides are basic (alkaline), and with negative ions or radicals the metals form numerous salts. The free metals have a high lustre and with the exception of copper and gold are silver white or gray. They are good conductors of heat and electricity because of the unique bonding of their atoms to each other, which gives free play to their outer electrons. Most of the metals are malleable, ductile, strong and have high melting and boiling points.

Nonmetals have few physical properties in common; some are gases, one is a liquid and the others are solids at atmospheric temperatures. They have a relatively large number of valence electrons, and in chemical reactions some gain additional electrons to form negative ions. Others with fewer outer electrons form covalent compounds by sharing their electrons with the atoms of other elements. Carbon, for example, forms several hundred thousand compounds in this manner. The oxides of the nonmetals are acidic and with water form the oxyacids; *e.g.,* sulfuric (H_2SO_4), phosphoric (H_3PO_4) and carbonic (H_2CO_3). The hydrogen compounds of the more active nonmetals are also acids; *e.g.,* hydrochloric (HCl) and hydrofluoric (HF). Although there are only about 15 nonmetals they constitute the greater portion of the outer crust of the earth and are essential to the growth and existence of living things as well as of industrial importance.

Several elements (namely, arsenic, germanium and antimony) are difficult to classify as they have neither pronounced metallic nor nonmetallic properties.

There are other criteria by which the chemical elements may be classified, but from both the theoretical and practical standpoint the periodic classification is by far the most fundamental and important.

See PERIODIC LAW; *see* also articles on the various elements or groups of elements, and references under "Elements, Chemical" in the Index volume.

BIBLIOGRAPHY.—N. V. Sidgwick, *The Chemical Elements and Their Compounds* (1950); T. Moeller, *Inorganic Chemistry* (1952); M. E. Weeks, *Discovery of the Elements* (1956); *see* also bibliography to article PERIODIC LAW. (J. B. Ps.)

ELEPHANT, a mammal of the suborder Elephantoidea, order Proboscidea, characterized by massive size; great strength; a disproportionately large head; long, grasping trunk or proboscis; short neck; and thick, loose, nearly hairless skin. The name pachyderm, referring to the thick skin, is a holdover from an earlier and now obsolete classification in which the elephants were grouped with such unlikely animals as walruses, tapirs, rhinoceroses and hippopotamuses under the term Pachydermata.

Elephants, the only surviving proboscideans, belong to two genera, *Elephas,* the Asiatic elephant, and *Loxodonta,* the African. Although certain superficial features distinguish them, these two forms are very much alike in basic structure, habits and behaviour.

General.—Adults have a tuft of hairs at the tip of the tail and sometimes a small patch on the forehead. The limbs are adapted for bearing a great weight. The legs are straight and pillarlike and the bones of the joints flat at the articular surfaces. The feet are clublike; each toe has a heavy nail, the weight being borne on thick pads behind the toes. The nose and upper lip are extended into a long, prehensile trunk, containing the nasal passages and with nostrils at the tip. Water for drinking is sucked up into the trunk and then discharged directly into the mouth. The trunk also serves as a hand and arm to stuff food into the mouth, as well as for spraying and dusting the body.

The upper incisors are typically developed into ivory tusks, the longest and heaviest teeth of any living animal. Canine teeth are absent. The complex molars are of the high-crowned lophodont type, with transverse rows of enamel ridges on the grinding surface. There is normally only one functional tooth at a time on each side of each jaw; these teeth are replaced from the rear as they wear away.

Natural History.—Elephants are gregarious, traveling slowly in herds of up to 100 animals. Their maximum running speed, when pressed, is from 15 to 25 m.p.h. They inhabit forests or grassy plains and are entirely vegetarian. The Asiatic elephant eats mainly grass, as well as leaves and fruit of wild plantain, young bamboo shoots and the leaves, twigs and bark of various trees. The African elephant seldom eats grass, but is fond of soft fruits and the inner bark of some trees; and chews the roots and branches of many trees, reducing them to pulp, for the sap therein. The proboscis is used for uprooting or stripping the leaves from trees, or for shaking down fruits; the tusks are used for digging up roots or tubers. In captivity an elephant may consume 500 lb. of hay and 60 gal. of water in a day.

An elephant's brain weighs about 10 lb. but is not unusually well developed. The elephant is alleged to have a high order of intelligence; it certainly is amenable to training. Its sense of smell is keen, but vision and hearing are poorly developed. Elephants, though normally timid, can be dangerous when aroused. Old bulls, especially, may become unruly and mentally deranged; in the wild they are solitary outcasts from the herd. Their periods of rage, called must or musth, may be related to sexual excitement or to increased exudation from special frontal glands of the head. Females are usually placid.

Elephants may first breed at from 15 to 20 years, but are not fully grown until about 25. A cow may have four or five calves during her lifetime. The gestation period is 20 to 22 months, and one young (rarely twins) is born at a birth. A newborn calf is about three feet tall and weighs about 200 lb.; it is clad in a short woolly coat, and may nurse for two years. There are but a few records of breeding in captivity.

Elephants are reasonably long-lived; probably a life span of about 50 years is normal in the wild. One captive elephant in India was recorded as living for 130 years; the African species has lived for 80 years in captivity.

The fossil history of elephants is remarkably complete since the late Eocene. The stout bones and teeth have left a well-documented record on all continents except Australia. The early *Moeritherium* of northern Egypt was about two feet high. Since then there have evolved over 300 species of proboscideans, marked by progressive increase in size, development of trunks and tusks and complexity of teeth together with reduction in their number. At least four types of proboscideans inhabited North America during the Pleistocene: the mastodon (*q.v.; Mastodon* [*Mammut*] *americanus*), covered with long, coarse hair; the woolly mammoth

(*Mammuthus primigenius; see* MAMMOTH), contemporary with Paleolithic man; the Columbian elephant (*Elephas columbi*), with great incurved tusks; and the imperial elephant (*Elephas imperator*), standing over 13 ft. high. The gigantic *Loxodonta antiquus* occurred over most of Europe, India and Africa; a specimen from Kent, on display in the British museum, stands about 14 ft. high. Closely related to it were several dwarf forms whose fossil remains have been found in cave deposits on Malta, Crete and Cyprus. All these elephants died out, due perhaps to their massive size, to overspecialization or to changing climatic conditions. Today the only living elephants are the Asiatic and African species.

Asiatic Elephants.—The Asiatic elephant, *Elephas maximus* and its races, is distinguished from the African elephants by being somewhat smaller and by having relatively small ears, a less domed forehead, structurally more complex and high-crowned teeth, a smooth trunk with a single fingerlike process at the tip and four nails on the hind feet. They inhabit the forests of southeastern Asia from India to Ceylon and Borneo. Adult bulls reach a weight of six tons and commonly stand 10 ft. at the highest point. A record specimen in the Calcutta museum stands 11 ft. 3 in. at the shoulder. Only the males develop tusks, which average five feet in length, the pair weighing 70 lb. Tusks nine feet long and 150 lb. in weight have been recorded.

Asiatic elephants are rather intolerant of exposure to heat and resort to deep shade to avoid it. They love bathing, and shower themselves with water sprayed over their backs. They normally rest and sleep lying down.

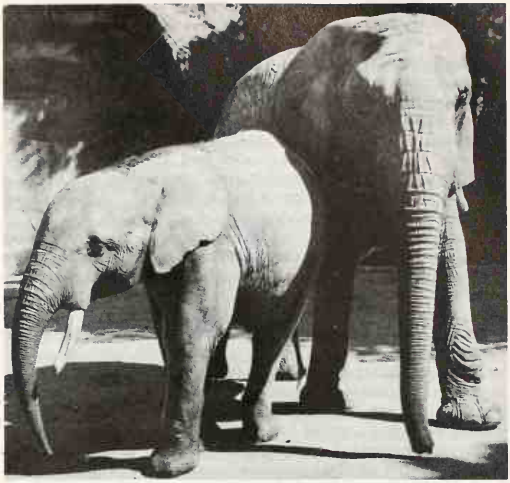
Four geographical races have been recognized: the Indian, Ceylonese, Malayan and Sumatran elephants. Males of the Ceylonese race generally lack tusks, and Sumatran elephants are of slighter build and have longer trunks; but these and other differences are not constant throughout the entire population of the species.

African Elephant.—In contrast to the Asiatic species, the African elephant is generally larger, except for the pygmy form (*see* below), and has tremendously large ears (3½ ft. in breadth), a roughened trunk with two processes (instead of one) at the tip and three (rather than four) nails on each hind foot. The molars are of coarser construction, with fewer ridges and thicker enamel. This species has longer legs, a more convex forehead and relatively larger eyes. Another point of difference from the Asiatic elephant is that the African species usually sleeps while standing.

Bush Elephants.—African bush elephants (*Loxodonta africana*) have been hunted out of large areas in southern and southwestern Africa, but still occur over much of the continent south of the Sahara. The average height of adult bulls is 11 ft. at the highest point, and the average weight six tons. Cows are about two feet shorter. Both sexes possess tusks, which average about six feet long, the pair weighing 80 to 120 lb. A pair of tusks in the British museum weighs 293 lb.; the larger one of the two is 11 ft. 5½ in. long, with a basal circumference of 18 in. The famous "Jumbo," of the Barnum and Bailey circus, was a bush elephant that stood 10 ft. 9 in. in height. The largest elephant on record, a bull bush elephant killed in the Cuando river district of southeastern Angola in 1955, is on display at the Smithsonian institution in Washington, D.C.; it weighed 12 tons when alive and stood 13 ft. 2 in. at the shoulder.

YLLA — RAPHO GUILLUMETTE

FIG. 1.—ASIATIC ELEPHANT (ELEPHAS MAXIMUS) WITH YOUNG



W. SUSCHITZKY

FIG. 2.—AFRICAN ELEPHANT (LOXODONTA AFRICANA) WITH YOUNG

Several varieties have been recognized, differing chiefly in the form and size of the ear and the shape of the skull: the central, Sudan, east and south African races.

Pygmy Elephant.—There is some disagreement about the status of the African pygmy elephant: some mammalogists give it species rank as *Loxodonta cyclotis*, whereas others consider it a subspecies of the bush elephant and name it *Loxodonta africana cyclotis*. It is a forest-dwelling elephant of western equatorial Africa, particularly the Congo region. Adults may be seven feet tall and weigh 2,700 lb., with slender tusks no longer than two feet. The ears are relatively small and smoothly rounded at the margins.

Albino Elephants.—Albino or "white" elephants occasionally appear, especially in Siam and Burma, where they are revered as creatures related to the gods. These elephants have eyes and skin of a pinkish cast, because of the absence of normal pigments; their skin is not actually white, but rather a light gray instead of the usual dark gray or brownish-gray. At the cost of a small fortune, the famous American showman P. T. Barnum procured one of these specimens, but it proved disappointing as an attraction, expensive to maintain and extremely difficult to dispose of—hence our expression, "white elephant," for a possession that is valuable but too burdensome to keep.

Importance of Elephants.—Elephants constitute the chief source of commercial ivory; unfortunately, practically unregulated traffic in this commodity has greatly reduced the African herds. From early times elephants have been used as beasts of burden in India and Burma. Since they do not breed freely in captivity, new stock for domestication is often captured from wild herds. One method is to drive them through a funnel-shaped palisade into a small enclosure; a specially trained tame elephant helps to subdue, separate and train the young captives for service. The training and handling of an elephant are usually entrusted to one man, called the mahout in India and the *oozie* in Burma; an

elephant and its keeper usually become inseparable and work together for the rest of their lives.

Trained elephants carry human cargo in a howdah or miniature hunting lodge on their backs. More particularly, they serve for moving and stacking timber or other heavy materials, using both tusks and trunks in the operation. In the teak industry huge logs are floated down rivers and then hauled out with chains by elephants.

Asiatic elephants have been most used as draft animals, not that they are more tractable but their trainers are more highly skilled. Since 1890, the French and Belgians in the Congo have used elephants for heavy work.

African elephants may have been domesticated by the ancient Carthaginians and used in wars with the Romans. Elephants are depicted on medals of Faustina and of Lucius Septimius Severus—with heads and bodies of the Asiatic, but ears of the African species. In 218 B.C., Hannibal and his army crossed the Alps with elephants. In later times they were used in Roman amphitheatres and for military pageants, a practice that continues to this day in the form of exhibitions at circuses, carnivals and zoological parks. In warfare elephants were used for dragging heavy equipment, especially through mud and up steep slopes. As late as World War II elephants were of considerable value in military movements over the mud-clogged roads of southeast Asia.

Elephants can also be a menace to cultivation. Wild herds that break into a plantation may destroy acres of produce in a single night.

Folklore.—The association between man and elephants goes far back into mythology, and a rich folklore has developed around these animals. The elephant-faced god, Ganesha (*q.v.*), plays a part in Hinduism, as a remover of obstacles. According to Hindu legend, eight bull and eight cow elephants sprang from the two halves of a "cosmic egg" at a chant from the Creator; these first elephants possessed the power of flight, but this they lost from the curse of a hermit when they landed in a banyan tree and crushed his dwelling. In parts of Africa the natives wear bracelets of hairs from an elephant's tail as good luck charms.

Fact and Fancy.—Attempts have been made to locate the legendary "graveyards" to which old elephants resort when their end is near. Groups of buried elephant remains have been found, but they probably represent sites, since elevated and dried, where elephants once drowned in bogs or while crossing rivers, or perished from imbibing at an alkaline water hole. The remains of dead elephants usually disappear in a short time from the action of scavengers and weather. The "dancing" of elephants is probably an attempt to dislodge stinging and biting insects. An elephant is not inherently afraid of mice, but is concerned that these and other small animals might enter its trunk. Of course any such foreign object that does enter is forcibly expelled. There is some truth to the saying that "an elephant never forgets"; it does have a retentive memory.

See also references under "Elephant" in the Index volume.

BIBLIOGRAPHY.—F. G. Benedict, "The Physiology of the Elephant," *Carnegie Institution of Washington Publ.* 474 (1936); Richard Carrington, *Elephants* (1958); G. Blond, *The Elephants* (1961); Sir Gavin de Beer, *Alps and Elephants: Hannibal's March* (1955); Guy Dollman, "Pigmy Elephants," *Natural History Magazine* (London), vol. 4, pp. 266–271 (1934); C. B. Moore, *Ways of Mammals in Fact and Fancy* (1953); R. I. Pocock, "Notes on the Asiatic Elephant (*Elephas maximus*)," *Annals and Magazine of Natural History,* series 11, vol. 10, pp. 273–280 (1943); E. Sunamoto, *The Elephant,* 2 vol. (1931–32); G. H. H. Tate, *Mammals of Eastern Asia* (1947); Rowland Ward (ed.), *The Elephant in East Central Africa* (a monograph by various contributors) (1953); J. H. Williams, *Elephant Bill* (1960).

(R. H. MA.)

ELEPHANTA (Hindi/Marathi GHARAPURI), an island 4½ mi. in circumference on the east side of Bombay harbour, India, 6–7 mi. from the city landing stages, famous for its cave temples or *lenen.* Pop. (1951) 333. The Portuguese and customary modern name comes from the stone elephant that stood at the southern end of the island but was removed to Victoria gardens, Bombay, in 1864. There are on the island eight caves (they are in fact chambers hewn by man from the hard trap rock). Of these, cave I (the Main or Great cave) is the most notable. All were dedicated to Siva, whose story is presented in a series of high-relief wall panels in cave I.

Cave I comprises a main hall 123 ft. 4 in. long, 55 ft. wide, with a central columnated atrium, 48 ft. by 12 ft., and 12-ft.-wide aisles on the long (north and south) sides. The entrance portico, giving access to the north aisle, is balanced by a side chapel lying parallel to the south aisle. Near the western end of the main hall is an 18-ft.-square chamber with four entrances guarded by colossal sculptured doorkeepers, and with space around it for pradakshina, ritual circumambulation. This is the main shrine of the lingam, the phallic symbol of Siva as procreator. East and west of the main hall are pillarless courts, once containing statues of Nandi, Siva's riding bull. From the west court opens northward a large cistern, westward an atrium with a small lingam shrine beyond. A large atrium, 57 ft. 1 in. by 24 ft. 5 in., lies south of the east court and, beyond it, a third, larger, lingam shrine with pradakshina ambulatory. During the cutting of the caves, probably in the 8th century A.D., 26 massive freestanding columns with large Dravidian-type cushion capitals and 8 lesser columns had to be disengaged from the solid rock. The average height of the chambers is 16 ft. At the back of the south-side chapel is the grandly impressive "Mahesamurti Siva," a 20-ft.-high three-faced bust representing the god as destroyer, creator and preserver.

Elephanta cave I is the last major monument of western Indian architectural sculpture, achieving its effect with massive forms and deep undercutting and recessing of the carving. Of the remaining caves, II and VIII were never finished; III and IV still have doorkeeper figures. All were much damaged by water, by cattle and by Portuguese soldiers, who in 1712 even fired cannon into cave I. About 1750 cave VI was being used as a church.

See H. Sastri, *A Guide to Elephanta* (1934).

ELEPHANTIASIS (LYMPHEDEMA), a medical term used to describe a progressive enlargement or hypertrophy of the skin and subcutaneous tissues of one or more extremities, labia, the scrotum and sometimes of the mammae or of the face. It results from the blockage of the venous or lymphatic channels. The nematode round worms are the commonest cause of the blockage. These worms, at one stage of their development, invade the blood and lymph vessels and interfere with the drainage of the affected parts. Other causes are injury, inflammation or neoplasms.

See FILARIASIS; PARASITOLOGY: *Roundworms;* LYMPH AND LYMPHATIC SYSTEM: *Filarial Infestation: Elephantiasis.*

(L. T. Co.)

ELEPHANTINE, the Greek name for the ancient Egyptian Yeb, "Elephant town," modern Jazirat Aswan, an island in the Nile opposite Aswan in upper Egypt. Ruins of the ancient city, so-called probably because it was the mart for Nubian ivory, can still be seen. The 18th- and 19th-dynasty Pharaohs built there a temple to Khnum, the ram-god of the cataract region, to his consort Satis, and to Anukis, lady of nearby Sehel; this and other temples, altered and enlarged by the Ptolemies and Romans, were largely destroyed in 1822 by a local governor in need of building stone. A granite gateway, one of the rare monuments of Alexander IV, the ill-fated son of Alexander the Great, still stands. The necropolis of the sacred rams was found nearby in 1908, their gilded mummies enclosed in small stone sarcophagi. In 1948 the Egyptian Antiquities department excavated a mud-brick temple built by a nomarch of Elephantine, Sirenput, in the Middle Kingdom in honour of an illustrious 6th-dynasty predecessor, Hekayeb. Both these governors and some 34 others are buried in rock tombs high in the cliff on the west bank of the Nile. In the Old Kingdom Elephantine was the "door of the south," the most southerly city in Egypt, and the starting point for Sudanese trade. In the New Empire the region was part of the province of Nubia, but in the Saite period it again became a frontier fortress.

The Egyptians believed that the Nile flowed out at Elephantine from bottomless springs between two rocks, called by Herodotus Crophi and Mophi. In the quay in the southeast of the island is the Roman Nilometer, mentioned by Strabo.

Graffiti of all periods are found on the rocks and in the granite quarries of Syene. In the ruins and rubbish heaps of the ancient town were found a number of Aramaic papyri from the archives of

Jewish mercenaries garrisoned there in the 5th century B.C.
(M. S. Dr.)

ELEPHANT SEAL (*Sea Elephant*), a name applied to seals of the genus *Mirounga* because of their unusual, trunklike snout. There are two living species. The northern elephant seal, *M. angustirostris*, occurs along the Pacific coast from central California south to northern Baja California. Although almost exterminated by the end of the 19th century as a result of commercial exploitation, it has now come back in numbers. The southern elephant seal, *M. leonina*, occurs widely in subantarctic waters, but is most numerous in the Falkland Island area. Both species are about the same size. Males have been recorded up to 22 ft. long and females up to 11 ft. 6 in. *See* also Seal; Sealing.

E. AUBERT DE LA RÜE

ELEPHANT SEAL OF THE SOUTHERN HEMISPHERE (MIROUNGA LEONINA)

(R. T. O.)

ELEPHANT'S-FOOT, popular name for the plant *Testudinaria elephantipes*, a native of the Cape of Good Hope. It takes its name from the large tuberous stem, which grows slowly but may reach a size of more than three yards in circumference with a height of nearly three feet above ground. The stem is rich in starch, whence the name "Hottentot bread," and is covered outside with thick, hard, corky plates. It develops slender, leafy, climbing shoots, which die down each season. Elephant's-foot belongs to the monocotyledonous family Dioscoreaceae (*q.v.*), which includes the yam (*q.v.*).

ELEPHANT'S-HEAD (*Pedicularis groenlandica*), a North American plant of the figwort family (Scrophulariaceae, *q.v.*), native to wet soil from Hudson bay to Alaska and southward on mountains to New Mexico and California. It is a smooth, erect perennial, about a foot high, bearing pinnately divided leaves and a dense cluster (spike), 2 in. to 6 in. long, of crimson flowers, irregular in shape, with the upper lip produced into a long beak, at first bent downward but soon curving forward and upward, the whole strikingly suggestive of an elephant's head. The plant, which blooms in midsummer, is a showy feature of alpine meadows in the national parks of the western United States and Canada.

A related species, *P. attollens*, of high mountains in California and Oregon, with white or pinkish flowers marked with purple, but lacking the long beak, is known as little elephant's-head.
(N. Tr.)

ELEUSIS, an ancient city and now a modern town of Greece, famous as the site of the Eleusinian mysteries. This article is concerned only with the place; for the religious observances *see* Mystery.

The city lies about 14 mi. W. of Athens on a rocky ridge close to the shore of the bay of the same name and is opposite the island of Salamis. Tradition carries its origin back to the highest antiquity. In earlier history it seems to have been independent of Athens, as it was from 403 to 401/400 B.C. when the oligarchs expelled from Athens occupied it. Of its small territory the plain of Thria was fertile. To the west was the Rharian plain, where Demeter is said to have sown the first seeds of corn, and the field called Orgas, planted with trees sacred to Demeter and Persephone. The sacred buildings were destroyed by the Gothic leader Alaric in A.D. 395. The present town of Eleusis with much of the neighbouring plain has become an industrial suburb of Athens.

The Greek archaeological society, excavating the site after 1882, laid bare the whole of the sacred precinct, traced its extensions at various periods and revealed successive stages in the structure of the *telesterion*, or hall of initiation.

In front of the main entrance is a paved area, with the temple of Artemis Propylaea of Roman date; on each side of the Great Propylaea stood a Roman triumphal arch of the time of Hadrian (2nd century A.D.). Just to the left of the Propylaea is an ancient well, almost certainly the Callichoron well ("of beautiful dances") mentioned in the Homeric *Hymn to Demeter* and by Pausanias. The Great Propylaea is of the time of the emperors Antoninus Pius and Marcus Aurelius (2nd century A.D.) and is a copy of the central building of the Propylaea on the acropolis of Athens. The Small Propylaea was built in the mid-1st century B.C. by Appius Claudius Pulcher, the contemporary of Cicero, and forms the entrance to the inner precinct. It is built over a gate of the Pisistratid (6th century B.C.) fortification wall. Thence a paved road of Roman date leads to one of the doors of the *telesterion*. Above the Small Propylaea, partly set beneath the overhanging rock, is the precinct of Pluto with a curious natural cleft behind the small temple. There was thought to be the entrance to Hades through which Pluto carried off Persephone. Farther on a rock-cut platform approached by broad steps probably served for spectators. Beyond this, on higher ground, close to the side of the *telesterion*, stood a small temple, perhaps a treasury.

The *telesterion* was a large covered building about 170 ft. square, surrounded on all sides by seats partly cut in the solid rock where the *mystae* or initiates sat while the sacred rites took place on the floor. Near the centre stood a small chamber, the *anaktoron* or Holy of Holies, where the sacred tokens were kept. There were two doors on each side of the hall, except the north-west, where it is cut out of the solid rock and where a rock terrace at a higher level adjoins it from which there was access to the gallery. The roof was carried by rows of columns in two stories.

Excavation revealed the architectural history of the hall. The earliest building is a small houselike structure with an altar in front of it dating from Late Mycenaean times, before 1000 B.C. This is presumably the temple referred to in the Homeric *Hymn to Demeter*. The first monumental building, of polygonal masonry, was built about 600 B.C. This was succeeded by a square hall of almost the same plan as the later *telesterion* but about a quarter of the size; its eastern corner coincides with that of the later building and it had a portico in front and internal columns. This has been assigned to the time of Pisistratus. It was replaced by the much larger Periclean (mid-5th century) hall designed by Ictinus, who also designed the Parthenon at Athens. This hall is faced with a portico which was designed by the architect Philo (*c.* 330 B.C.). In front of the telesterion was a spacious terrace supported by massive walls and fortification towers which was extended with each successive enlargement of the hall. The precinct was full of altars, dedications and inscriptions. There is a museum on the site which houses the principal discoveries.

Bibliography.—The excavations of Eleusis have been made by the Greek archaeological society and published in its journals *Praktika* and *Ephemeris*. *See* also F. Noack, *Eleusis* (1927); K. Kourouniotes, *Eleusis, a Guide to the Excavations and the Museum*, Eng. trans. by O. Broneer (1936); G. E. Mylonas, *Eleusis and the Eleusinian Mysteries* (1961).
(E. Gr.; E. V.)

ELEUTHERIUS, SAINT, pope from *c.* 175 to 189, is said to have been born in Nicopolis in Epirus, Greece. Hegesippus relates that he was deacon of the Roman church under Anicetus. He was pope in the reign of the tolerant Commodus. He followed the Montanist controversy carefully and received advice on it from Irenaeus, who was sent by the churches of Lyons and Vienne for that purpose. Modern historians agree in rejecting the legend that a king of Britain named Lucius wrote to Eleutherius manifesting a desire to accept Christianity. His feast day is May 30.
(J. V. Hn.)

ELEVATOR (Lift). An elevator is a contrivance for lifting and lowering people or freight from one level to another. There

are two general classifications, passenger and freight. Elevators are used in most multifloor structures, as well as in mines.

An elevator is distinguished from a hoist in that the latter is commonly and primarily used for the handling of material (building material, sand, mortar, etc.). Essentially, an elevator consists of a power-operated machine located usually at the top of a hoistway and which, by means of steel hoist ropes or cables, lifts and lowers an enclosed platform, known as a car, between guide rails.

History.—Lifting loads by mechanical means, using hand power, dates from the Roman empire, as early as 236 B.C. Vitruvius, a Roman architect of about 26 B.C., is credited with describing early devices. Mention is made of manpower, animal power and water power.

The more practical elevator was developed in Europe in the early 1800s and used the hydraulic principle. The basic hydraulic type operates by means of a vertical plunger traveling in a cylinder which extends into the ground approximately equivalent to the height that the elevator travels. The plunger, carrying a car at its top end, is moved by liquid under pressure. Some hydraulic elevators, however, employ a horizontal or vertical piston and a combination of sheaves and ropes to multiply the car motion. Because of the multiplying ropes, this type is known as the rope-geared hydraulic elevator. By means of a hand rope extending through the car, a valve can be operated to control the starting and stopping of the elevator. Electric control is sometimes used.

Steam was employed in the early elevators for operating pressure pumps, and the exhaust was frequently used in the colder climates for heating and other purposes. Electric motors replaced the steam pumps for those hydraulic elevators still in existence in the middle 1900s; these are known as electrohydraulics.

Hydraulic elevators found considerable use in the 19th century and made a valuable contribution to the building industry. Because of physical limitations of plunger length, drilling depth and piston lengths for geared hydraulics, building heights and elevator speeds were limited. The maximum speed is 600–700 ft. per minute. The plunger-type hydraulic elevator is still used to some extent for short travels.

Electric Elevator.—In the U.S. the electric elevator was first used commercially in 1889 when two of them were installed by Otis Brothers & Company in the Demarest building, New York city. These used the principle of a drum driven by an electric motor through a worm-and-gear reduction. Hoist ropes are fastened to a spiral-grooved drum and pass over and around it so as to be wound onto and off the drum as the elevator car travels up and down. A second rope wound on the drum is attached to a counterweight. As the elevator rope winds onto the drum the counterweight rope unwinds. This type is known as a winding-drum machine.

The drum elevator restricted building height to about 150 ft. because the elevator travel was limited by the practical size of the drum and the length of hoist rope which could readily be wound onto it. The power drive was by a direct-current motor with rheostat control. The latter proved somewhat rougher in starting and stopping than the hydraulic type. Speed was limited to approximately 350 ft. per minute.

In 1895 Frost and Strutt (England) invented a hoist called the "Teagle" which used hoist ropes extending from a car up over a sheave and then down to a counterweight. The overhead sheave was connected to a power unit. This principle of using friction between ropes and sheaves is used in practically all modern elevators. It is known as traction drive, and it removed the mechanical limitations imposed by the drum-type machine and allowed higher elevator travel and greater speed.

The gearless-traction type was developed by the Otis Elevator company and the first installation was made in 1904; in this type the shaft of a relatively slow-speed electric motor is connected rigidly (without gears) to a grooved sheave over which the hoist ropes, usually four to six, are double wrapped (using a secondary sheave) to transmit the necessary lifting power through the friction between the sheave and hoist ropes. The usual arrangement is to mount the gearless machine over the hoistway. The ropes are attached to the car at one end and to a counterweight at the other end. This type came into general use by 1910.

For low buildings where higher speed is not a factor, a gear reduction of the worm-and-gear type is used with a high-speed motor and a traction sheave with special grooves and no secondary sheave. Known as the worm-geared traction type, it is used chiefly for low-rise, slow-speed (350 ft. per minute) passenger elevators and for freight elevators. It had generally superseded the drum machine by 1918.

Motive Power.—It was early determined that the direct-current motor and rheostat control offered the best characteristics for attaining acceleration. However, for slow-speed operation in both passenger and freight service, single-speed and two-speed alternating-current motors are frequently used. This was due partly to the A.C. power-distribution system so widespread in the U.S. The use of a D.C. motor and an A.C. power supply requires a conversion unit to obtain D.C. elevator power. Because of its adaptation for controlling a D.C. elevator motor, there has been general acceptance of the motor-generator set for the conversion unit.

Control.—The early use of D.C. motor drive was accompanied by rheostat control, which depends upon the cutting out and in of resistance to accelerate and retard the elevator motor. Change in direction was easily accomplished by reversing the lead connections to the armature. For high speed and smooth acceleration, a large number of resistance steps is necessary. Resistance steps were also used with A.C. motors about 1900. This type of control, with slight modifications, continued generally until 1919.

Resistance control has the inherent characteristic of high power loss and, unless there are many resistance steps, relatively rough acceleration. To overcome this a multivoltage control system was developed whereby a number of fixed voltages are impressed progressively on the driving machine. This, in practice, is accomplished by a motor-generator set with multiple generators, each with a fixed voltage.

Generator Field Control.—Building-height requirements for higher speed presented a definite need for considerable variation in voltage (in infinite increments) if fast, smooth acceleration and retardation of high elevator speed was to be obtained under changing load conditions. The first successful commercial arrangement was placed in operation by Westinghouse Electric corporation (1922) in the Chicago Athletic Club building. In practice, this consists of a D.C. elevator motor and a motor-generator set for each elevator whereby the elevator speed is controlled by varying the voltage applied to the driving machine by adjusting the strength of the generator field. This allows higher speed since stopping can be attained accurately, and both acceleration and retardation are characteristically smooth. It is suitable for all elevator applications and for either D.C. or A.C. power supply. It is commonly known as variable-voltage control. In the 20th century, consequently, three types of equipment came into general use: geared variable voltage, geared A.C. rheostatic and gearless variable voltage.

The general type of car-operating device progressed with elevator development from the early hand-rope type for electric elevators to the car switch or controller which was introduced about 1895 and remained until about 1927 as the commonly accepted type where elevator attendants are employed. By manipulating an electric car switch in the car the attendant controls the direction of the car, its speed, starting and stopping through magnetic switches in the control room. However, with the advent of higher speeds, accurate stopping became difficult and operation inefficient. The practical speed limitation for such operation is 600–700 ft. per minute.

Automatic landing, introduced by Westinghouse in 1925 and used with car-switch operation, overcame many of the limitations in speed. Accurate, smooth landing at a floor was obtained irrespective of elevator speed or load. Automatic releveling could take place in case of rope stretch. This feature started another phase in elevator development by allowing higher speeds and making higher buildings feasible.

Push-button control or automatic operation, whereby an elevator can be operated by pressure of a button in the corridor or in the car, was first introduced by Otis in 1892. Its main advantage was to allow a passenger to operate the elevator and eliminate the attendant. Such applications were restricted primarily to single, slow-speed elevators where landing accuracy was not important. This led gradually by 1925 to collective control, also developed by the Otis Elevator company, which allowed a selectivity of calls to be answered in sequence according to their floor location. The car operates in response to calls registered on the car station or on corridor buttons. Momentary pressure of any of these buttons starts the car in the desired direction. As the car travels, stops are made in the order in which the floors are reached. All calls in one direction are answered before the car reverses automatically to answer calls in the other direction.

This system was later expanded to two- and three-car operation by further development of accessory equipment. With such an arrangement the corridor buttons are common to all cars. Also, dual control was made available whereby both attendant and collective control operation are possible.

Beginning with the late 1920s, electric power-operated doors and other refinements were important contributions to automatic operation.

Signal operation marked a distinct phase in elevator-system control. It was first installed in New York city in 1924, in the Standard Oil building. With this system, the cars stop automatically in response to car and corridor button calls. The attendant's primary function is to close the elevator doors and start the car. The system allows high-speed (1,400 ft. per minute) operation because the human element is eliminated in making stops.

The efficiency of such a system was enhanced considerably by the accompanying development of power-operated doors, dispatching equipment and signals. Dispatching equipment is a device to dispatch a car by operating a signal in the car to indicate when the car should leave a terminal. With later development, the dispatching can be accomplished by actuating the car's starting mechanism. The system functions on an interval basis of usually 20–40 seconds. The cars need not follow any prescribed sequence.

"Selectomatic" and "Autotronic" are names of supervisory controls as installed by Westinghouse and the Otis Elevator company, respectively. With them a bank of elevators operates as a system by being properly dispatched to provide service in accordance with traffic demands as represented by morning up-peak, evening down-peak, normal, night service and other than normal traffic periods. Such innovations as "high call reversal," whereby the cars reverse automatically at the highest call; high-zone and low-zone separation of cars; "instant dispatch," etc., are incorporated to provide service synchronized to the traffic demands. An attendant in the lobby (known as a starter) by means of a switch in the starter's station can place the elevator system into any one of the several traffic patterns so that the system will respond automatically to the type of traffic existing. Cars are dispatched automatically; under up-peak traffic conditions they reverse at the highest call automatically, all in a manner to provide both large handling capacity and efficient quick service to all floors; during down-peak the elevators may be divided into two groups: a low-rise group to serve the lower floors of a building and a high-rise group to serve the higher floors.

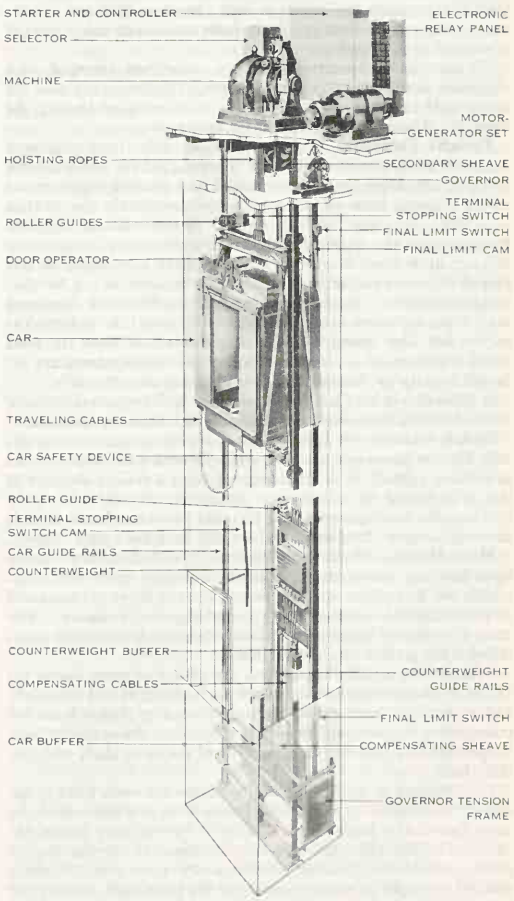
By 1949 additional automatic features were incorporated to allow such systems to operate automatically in the various traffic patterns without elevator attendants. With this arrangement the passengers press the buttons in the car station to register the stops for the floors to which they wish to go. The system will operate automatically in the traffic pattern as selected by the starter. The cars are dispatched automatically and the doors close and open automatically. The proper interval is automatically maintained. The system is highly efficient and also highly economical in that the cost of elevator attendants is eliminated.

A further refinement was the incorporation of automatic traffic pattern or program which automatically changes the operation of the elevators in accordance to the traffic demands or patterns without personal action or supervision by a starter. The controls, by taking note of the location, frequency and direction of calls select the operating pattern just as an attendant would. This system therefore, offers an efficient operation of a bank of elevators without any operating personnel. It is applicable, however, to elevators with or without elevator attendants. Less than 5% of the elevators made today require an elevator attendant to operate the controls.

Application.—Passenger elevators are classified usually as to type of control and as to capacity (pounds of load) and maximum speed (feet per minute). Capacities vary from about 1,500 lb. to 7,000 lb., and speeds from 100 to as high as 1,400 ft. per minute (R.C.A. building, New York city).

The shape of the cars varies with different services. A hospital elevator will normally be deeper than wide to accommodate beds; a department store will have a wide, shallow cab equipped with a wide door opening to allow quick ingress and egress of large numbers of passengers. The usual office building elevator would have greater width than depth. A 7-ft. (width) by 5½-ft. (depth) platform is typical of a general-purpose 3,000-lb.-capacity car.

Passenger elevators are normally counterweighted to the weight of the empty car plus 40% of the rated load. This allows the sys-



BY COURTESY OF OTIS ELEVATOR CO.

ELECTRIC PASSENGER ELEVATOR FOR AUTOMATIC SERVICE

tem to approach a balance in suspended loads and thereby economize in power consumption.

The early collapsible car gates and hand-operated hatchway sliding doors, while still used to an extent have been generally superseded by power-operated sliding doors for both cars and hoistways. The power units are of the electric type employing electric motors connected to a system of door-operating arms and suitable control so that door operation can be synchronized with the starting and stopping of the car. Doors may be of the centre-opening or side-opening and of single-speed or two-speed type. (*See* also ESCALATOR.)

Safety Devices.—To prevent a car from overspeeding in the down direction, a speed governor is provided in the penthouse. A car safety device actuated from a cable which passes over the speed governor sheave, causes powerful jaws to grab the guide rails in case of predetermined overspeed of the car. Buffers (devices for absorbing energy) of the coil-spring or oil-buffer (plunger) type are mounted in the elevator pit at the bottom of the hoistway, under the car and counterweight. The type used depends upon the rated speed of the elevator. If overtravel occurs the car is thereby brought to a safe stop.

So-called limit switches are provided in the hoistway to cut off power in case of overtravel of the car.

Door interlocks are used to prevent the car from starting until the hoistway doors are locked in their closed position and to prevent opening of the hoistway doors from the outside unless the car is stopped or is stopping.

Elevator power-operated doors are sometimes arranged with electronic devices or other means to protect passengers in case the doors start closing while a passenger is entering or leaving the elevator. These devices cause the doors to reopen.

Freight Elevators.—Freight elevators differ from passenger elevators in the size of the car, the more rugged car and platform construction, appearance, lower speed and generally greater capacity. Speeds from 75 to 200 ft. per minute for the geared type (machine), and capacities from 1,500 lb. to 16,000 lb. are common. In some cases the capacity is 30,000 lb. or more. For capacities of 4,000 lb. or more a 2:1 or 3:1 roping of the hoist ropes, as employed in blocks and tackles, is used as compared to 1:1 for passenger elevators. Normally, geared-type machines are employed for freight elevators since their chief characteristic is large capacity and slow speed. However, gearless machines are used where higher speed is a factor. Horsepower requirements are affected directly by the unbalanced load, speed and efficiency.

A sidewalk elevator is a freight elevator which operates between a sidewalk and floor levels inside a building below the sidewalk.

Dumb-waiters.—A dumb-waiter is a hoisting and lowering device, hand or power operated and equipped with a car which moves in vertical guides. It is differentiated from a freight elevator in that it is limited by code to nine square feet of floor area; is of low capacity (not exceeding 500 lb.) and never carries an attendant or passenger. It is normally operated from floor push buttons.

Mine Hoists.—The mining industry depends to a large degree upon hoisting equipment for handling minerals, waste rock, materials and men. For such equipment, electric drive is recognized as providing the required safety, reliability and economy. The chief difference in hoisting operations between coal mines and metal mines is the greater depth of the latter.

Practice has dictated the general use of one or more hoists, referred to as main, ore or skip hoists, for hoisting mined materials, and an auxiliary hoist called the supply, man or chippy hoist for transporting supplies and men. In many mines the auxiliary hoist is operated in a separate compartment of the same shaft with the main hoist.

For economy in equipment and operation the main hoist is operated "in balance"; *i.e.*, an empty cage or skip is lowered at the same time that a loaded cage is raised. For auxiliary hoists, because of intermittent use and the likelihood of serving several levels, unbalanced hoisting is frequently preferred. Another method is simply to counterweight for the hoist cage, making the counterweight approximately equal to the weight of the normal load.

Hoist conveyances are of two general forms: the platform cage and the self-dumping skip. The platform cage consists simply of a suspension member carrying a horizontal platform. If cars are to be handled, the platform is equipped with rails. This requires the replacement of loaded and empty cars between hoisting trips, with accompanying delays. When a self-dumping skip is employed, coal or ore is loaded directly into the top of the skip from a loading bin or underground haulage system. Upon reaching the surface the material is discharged either by inverting the skip or opening a side gate. This system saves time and labour and permits hoists to make as many as three trips per minute in shafts of about 500 ft. in depth.

The usual practice is to hoist in skips of capacities up to 15 tons and operate at maximum rope speeds up to 3,000 or 4,000 ft. per minute. The hoisting machine consists of a drum mounted on a shaft which is connected to a gear reduction unit and driven by an electric motor. Steel hoist cables are wound on the drum which is generally cylindrical but for certain hoists may be conical in whole or in part. A single-drum type has but one cylindrical drum and the hoisting rope winds onto the drum while the counterweight rope unwinds. Where the depth of the shaft does not permit the use of a single drum for counterweighting, two drums are used mounted on the single shaft. A drum diameter of nine feet is not uncommon.

Another hoist design provides a cylindrical drum over which the cables pass but are not fastened. This is identified as a friction- or Koepe-type hoist. This type is used for both production and service hoists and is very popular in Europe.

Either A.C. power, using a wound rotor induction type A.C. motor, or D.C. power, with generator voltage control, is normally used. However, the largest hoist equipment employs the D.C. adjustable voltage drive system. The motor requirements may be as high as 8,000 h.p. (H. C. Hk.)

ELEVATOR, GRAIN: *see* GRANARY AND GRAIN ELEVATOR.

ELF, a diminutive supernatural being of Teutonic mythology, usually mischievous, causing diseases and evil dreams, stealing children and substituting changelings; it differs from the Romanic, less sinister fairy (O.E. *aelf*; *cf.* Ger. *Eloe* and *Alp*, "nightmare"). Prehistoric flint implements, in England known as elf bolts or elf arrows, were looked on as weapons of the elves, with which they injured cattle. A tangle in the hair was called an elflock.

ELGAR, SIR EDWARD WILLIAM (1857–1934), the foremost English composer of the beginning of the 20th century, was born at Broadheath, near Worcester, on June 2, 1857. His father, a Roman Catholic, was a music dealer and organist. Leaving school at the age of 15, Elgar worked for a few months in a lawyer's office, but his talent, developed through studying works in his father's shop and through taking part in local activities, clearly lay in music. He was an excellent violinist and a member of several orchestras; he also played the bassoon and organ and made an early start as a composer. Local conducting appointments and his position as organist of St. George's Roman Catholic church afforded him opportunities of hearing his own works. An indefatigable student, Elgar invested his early savings in a visit to Leipzig (1883) and in violin lessons in London. In 1883 his *Intermezzo Mauresque* was well received in Birmingham.

Since he held no important post and worked in the provinces Elgar's progress as a composer was slow. After his marriage to Caroline Alice Roberts in 1889 he moved to London in the hope of achieving more. But, although he published minor pieces such as *Salut d'Amour* (*Liebesgrüss*), the move was unproductive, and in 1891 the Elgars returned to Worcestershire. Almost immediately larger works began to make an impression, at least in the Midlands, where choral societies were eager for new and exciting scores. Among these works were *The Black Knight* (Worcester, 1893), *Lux Christi* (*The Light of Life*, Worcester, 1896), and *King Olaf* (Hanley, 1896). August Manns, who had earlier befriended Elgar, conducted *King Olaf* at the Crystal Palace in 1897, and in the same year the *Imperial March* gave further notice that a new and distinctive talent had arrived. In 1898 *Caractacus*, dedicated to Queen Victoria, was warmly received and Elgar com-

mented that an article in *The Musical Standard* had given "me the place I've fought for."

Caractacus was followed by the *Enigma Variations* (1899), of which the quality was immediately perceived by Hans Richter, who continued to support Elgar's genius; *Sea Pictures* (1899); *The Dream of Gerontius* (1900), the fulfilment of an early ambition; the overture *Cockaigne* (1901); and the first two *Pomp and Circumstance* marches (1901). In 1902 a performance in Düsseldorf of *Gerontius*—unsuccessful at its first Birmingham performance—drew praise from Richard Strauss, who said that "with that work England for the first time became one of the modern musical states." Other performances followed in Germany, and in America, where a keen protagonist for Elgar's music was S. S. Sanford of Yale. To him was dedicated the *Introduction and Allegro* for strings (begun 1901, finished 1905).

During these years Elgar became the favourite composer of King Edward VII, and acknowledgment of his national standing came in the form of honorary degrees and in 1904 a knighthood. Among other notable compositions of this period were the oratorio *The Apostles* (1903), the overture *In the South* (1904) and the third *Pomp and Circumstance* march (1905). But in spite of the success of his works, which he conducted both in England and abroad, and the prolificacy of his pen Elgar was relatively poor. Thus, against his inclinations, he accepted the new chair of music in Birmingham university in 1904. The professorship brought many difficulties and Elgar, whose health was rarely good, was glad to hand it over to Granville Bantock in 1908. Honours continued to accrue and included the Freedom of Worcester (1905) and honorary degrees of Mus.D. (Yale) and M.A. (Birmingham).

In 1906 *The Kingdom*, sequel to *The Apostles* and the second part of an unfinished trilogy, was performed in Birmingham; in 1908 the Symphony in A flat was given in Manchester, and in 1910 the masterly Violin Concerto appeared. In the next year the second symphony (based, as was much of Elgar's music, on earlier sketches) was heard. This was a tribute to Edward VII, for whose successor Elgar wrote more coronation music. At this point he became a member of the Order of Merit and went to live in London. Before the outbreak of World War I the symphonic study *Falstaff* (reflecting Elgar's love of Shakespeare) was produced at Leeds (1913).

During the war Elgar wrote some occasional pieces, of which *The Spirit of England* (1916–17) contains the best music. Immediately after the war came a violin sonata and string quartet (1918) and a piano quintet and violoncello concerto (1919), all remarkable for their reflective quality. The death of Lady Elgar in 1920 was a blow from which Sir Edward never fully recovered. As Master of the King's Musick (1924) he wrote further occasional pieces, but he preferred a quiet, semiretired life in Worcestershire. Friendship with Bernard Shaw, however, stimulated in Elgar a late desire to compose more, and at his death, at Worcester on Feb. 23, 1934, he left unfinished a third symphony, a piano concerto, and an opera based on Ben Jonson's *The Devil Is an Ass*. A memorial window in Worcester cathedral recalls Elgar's long association with the city and the Three Choirs Festival.

One of the last of the romantic composers, Elgar is also one of the greatest. His fame rests on bold tunes—some have the force of folk music—striking colour effects, and a massive sense of structure. But, like Schumann, whom he admired, he reveals an acute sensitivity. Such qualities may also be found in his prose writings.

Bibliography.—P. M. Young, *Elgar O.M.* (1955), *Letters of Edward Elgar* (1956); D. M. McVeagh, *Edward Elgar* (1955).

(P. M. Yo.)

ELGIN, a city, royal and small burgh and the county town of Moray, Scot., lies on the river Lossie 5 mi. S. of Lossiemouth and 38 mi. E.N.E. of Inverness by road. Pop. (1961) 11,971.

Elgin was created a royal burgh by Alexander I and received its charter from Alexander II in 1234. The once splendid cathedral of Moray, formerly called the "lantern of the north" and now a ruin, was founded in 1224 when the church of the Holy Trinity was converted into a cathedral. It was partially burned in 1270 and almost destroyed in 1390 by Alexander Stewart, the "wolf of Badenoch," earl of Buchan and natural son of Robert II. Edward I (*q.v.*) of England stayed at the castle in 1296, and to blot out the memory of his visit the building was destroyed immediately after national independence had been reasserted. The hill on which the castle stood was renamed the Lady hill (after St. Mary's chapel, belonging to the castle) and now bears the Gordon monument. In 1402 Alexander Macdonald, son of the lord of the Isles, set fire to the town, but spared the cathedral. In 1452 half the town was burned by Sir Alexander Seton, lord of Gordon and first earl of Huntly, who, when he died in 1470, was buried in the cathedral—then the burial place of the Gordon family. Elgin was plundered by Montrose's army in 1645. In 1746 Prince Charles Edward spent a few days at Thunderton house, built in the early 14th century and now a hotel. Samuel Johnson and James Boswell had a "vile dinner" at the Red Lion inn in 1773.

Elgin is a centre of education, with Elgin academy, formerly the grammar school and now serving part of the county, and several schools. There is a literary and scientific society, and a museum, founded in 1842 by the society, which contains fossils from the Old Red Sandstone and two 8th-century incised figures of bulls from Burghead. Maj. Gen. Andrew Anderson (1746–1824) left a charitable and educational endowment called the Elgin institution, later renamed Anderson's institution and now Anderson's, which maintains a home for old people.

Near St. Giles's church (1828) in the broad High street is the Muckle cross (erected *c.* 1630, destroyed by vandals in 1791, rebuilt in 1888). At the eastern end of the High street is the 17th-century (restored) Little cross and near the western end rises Lady hill from the top of which the view commands the links (undulating ground along the seashore) of the Lossie. South of the Little cross is a convent incorporating the remains of a 15th-century Franciscan friary, the chapel of which, Greyfriars chapel, has been well restored. To the north, between High street and the Lossie, is a 40-ac. park presented to his native town by Sir George Cooper in 1903 and containing Grant lodge, a Georgian mansion which houses the public library.

The chief industries are farming, distilling and wool milling and the making of nets. There is a cattle show in July.

After its destruction in 1390 (*see* above) the cathedral was practically rebuilt on a scale that made it one of the finest churches in Scotland. Its design was that of a Jerusalem cross, with two flanking towers at the east end, two at the west end, and one in the centre at the intersection of the roofs of the nave and transepts. The central tower fell in 1506 but was rebuilt in 1538 at which date the building was completed. In 1554 the Dunbar and Innes families fought a bloody skirmish inside the cathedral. The central tower again collapsed in 1711 after which the whole building, which had already suffered considerable damage, was allowed to go to ruin. It lay neglected by all but John Shanks, a cobbler, who tended the ruins for many years before the government appointed him keeper in 1825. By the time of his death in 1841 he had cleared away all the rubbish, disclosed the original plan and collected a quantity of fragments. On the boundary wall a tablet with an epitaph records Shank's services to the cathedral. The ruins are in the care of the ministry of works. Of the cathedral buildings the 15th-century chapter house is the best preserved, and a wing (dated 1406) of the Bishop's palace remains.

The church of Birnie, 3 mi. S., was the seat of the see of Moray during the rule of the first four bishops; the fourth bishop, Simon de Toeny, an Englishman, was buried in its precincts in 1184. In the church, whose chancel and nave are the original Norman, is preserved a Celtic altar bell known as the "Ronnell bell."

(H. G. T.)

Gordonstoun School.—Six miles north of Elgin the 17th-century mansion of Gordonstoun, with its historic steading (farm buildings) known as the Round square, houses Gordonstoun school. The school was founded in 1934 by Kurt Hahn, who was already well-known in Great Britain as headmaster of Salem school in south Germany and who had been imprisoned for his outspoken criticism of Hitler. Gordonstoun, while preparing boys for the usual examinations and careers, has always emphasized the development of character and of a sense of responsibility, and it is well-

known for the active rescue services run by the boys: coastguards, fire service, mountain and beach rescue. In 1962 it was announced that Charles, prince of Wales, would attend the school; his father Philip, duke of Edinburgh, is a former pupil.

ELGIN, a city of Kane county, in northeastern Illinois, U.S., on the Fox river, 36 mi. N.W. of Chicago and a part of the Chicago standard metropolitan statistical area. It is the home of Elgin National Watch company, which also makes missile and aircraft devices. Other industries employing skilled labour include the manufacture of gaskets, auto parts and electrical appliances. The community's name derives from the old Scottish hymn "The Song of Elgin." Its first white settlers came at the close of the Black Hawk War (1832), when Chicago was yet a small community. Abundant water power where the Fox river falls 100 ft. in a 20-mi. stretch, quality limestone for construction and hardwood forests were quickly appropriated by the newcomers. A dam and millrace were built in 1837. It was chartered as a city in 1854 and rechartered in 1880. The surrounding morainic knolls, covered with grasses, made excellent farm lands, and by the early 1900s Elgin manufactured many dairy products. The city has a community junior college. Elgin is the headquarters for an independent Protestant publishing house and Church of the Brethren publications. For comparative population figures *see* table in ILLINOIS: *Population.* (H. L. SM.)

ELGIN AND KINCARDINE, EARLS OF. The Scottish earldoms of Elgin and Kincardine have been held jointly since 1747, when CHARLES BRUCE (1682–1747), 4th earl of Elgin, died without male issue and was succeeded by his cousin CHARLES BRUCE (1732–71), 9th earl of Kincardine. The earldom of Kincardine was created in 1647 for SIR EDWARD BRUCE (d. 1662); that of Elgin was created in 1633 for THOMAS BRUCE (1599–1663), 2nd Baron Kinloss.

Charles Bruce's third son THOMAS (1766–1841), 7th earl of Elgin and 11th earl of Kincardine, is famous for his acquisition of the Greek sculptures now known as the "Elgin Marbles." He was born on July 20, 1766, and succeeded to his brother's titles at the age of five. Entering the army in 1785 and rising later to the rank of major general, Elgin began his diplomatic career in 1790. Envoy at Brussels in 1792 and at Berlin in 1795 during the first phase of the war against revolutionary France, he was appointed envoy extraordinary at Constantinople in 1799, retaining the post until 1803. Detained in France on his way home through the rupture of the treaty of Amiens, Elgin did not reach England until 1806, and found his reputation under heavy attack. Though serving as a Scottish representative peer between 1790 and 1840, he took little further part in public life, and died in Paris on Nov. 14, 1841.

Keenly interested in classical art, Elgin secured permission from the Turks after his arrival in Constantinople first to record and then to remove Greek antiquities. Between 1803 and 1812 his great collection of sculptures, taken chiefly from the Parthenon at Athens, was brought to England, and became the subject of violent controversy. Elgin was denounced as a dishonest and rapacious vandal, notably by Lord Byron, while the quality of his acquisitions, later regarded as exceptional, was questioned. In 1810 he published a *Memorandum* defending his actions and judgment. On the recommendation of a parliamentary committee, which also vindicated Elgin's conduct, the "Marbles" were bought by the nation in 1816 for £35,000, considerably below their cost to Elgin, and deposited in the British museum, where they remain on view.

The 7th earl was succeeded by his third son JAMES (1811–63), 8th earl of Elgin, who became an outstanding colonial administrator. He was born in London on July 20, 1811, and educated at Eton and Christ Church, Oxford, where his academic career was distinguished. He was a prominent member of that group of brilliant contemporaries, including W. E. Gladstone, who first attained fame as admirers of Sir Robert Peel. Elgin was appointed governor of Jamaica by Peel in 1842. Faced by the complex economic and social situation created by the abolition of slavery, he showed much skill in conciliating the planters, encouraging new methods of cultivation and improving the condition of the liberated Negroes. In 1846 Elgin became governor general of Canada, where even greater difficulties confronted him. Racial animosity between the French and British settlers was intense; the colonists were insisting upon more freedom from imperial control; economic problems, especially after repeal of the corn laws, were acute, and relations with the United States most unsatisfactory. Son-in-law of Lord Durham, with whose views on Canadian affairs, expressed in the report of 1839, he sympathized, Elgin effectively established the pattern of responsible government, notably by his actions in 1847–49 and 1854; eased racial tension and promoted efficient administration; and in the reciprocity treaty of 1854 reached an understanding with the United States. His services were recognized by his creation as Baron Elgin in the British peerage in 1849. He returned to England in 1854.

On the formation of Lord Palmerston's first government in 1855, Elgin declined the offer of a seat in the cabinet. Two years later, upon the outbreak of war after the "Arrow" incident, he accompanied the expedition to China as special envoy. En route, Elgin was diverted to India by the appeals of his friend Lord Canning for assistance in suppressing the Mutiny, and left his troops behind for this purpose upon resuming his eastward journey after a brief delay. In 1858 he negotiated the treaty of Tientsin with the Chinese, successfully securing British demands, and completed his mission by obtaining Britain's first concession from Japan through the treaty of Yeddo. After an interlude as postmaster general in Palmerston's ministry in 1859, Elgin returned to China to enforce fulfillment of his treaty. He was supported by a punitive expedition, which sacked the summer palace at Peking. Late in 1860, terms were imposed upon the Chinese. Soon after he reached England, Elgin was appointed viceroy and governor general of India, and arrived to take up his post in 1862. As the first viceroy directly selected by the crown and subordinated to the secretary of state for India under the act of 1858, which brought to an end the rule of the East India company, Elgin faced many difficulties, and concentrated upon establishing the new system during his short period in office. However, his health had been undermined, and he died at Dharmsala in the Punjab on Nov. 20, 1863.

Universally respected, especially after his Canadian achievements, Elgin was a classic example of the great British proconsuls of the 19th century. Self-reliant and decisive, adaptable and indefatigable, he proved capable of understanding and handling widely varying and often highly complicated problems in imperial government. Elgin combined exceptional constructive qualities as a reforming administrator with force and daring as a diplomat. However his violence in the latter role, notably in the far east, though common at the time, contrasts sharply with the sympathy and liberality he displayed elsewhere.

The 8th earl's son and successor VICTOR ALEXANDER (1849–1917), 9th earl of Elgin, was born near Montreal, Que., on May 16, 1849, and educated at Eton and Balliol college, Oxford. In politics a Liberal of right-wing tendencies, Elgin was first commissioner of works under W. E. Gladstone in 1886. Emulating his father, whom he succeeded in 1863, he became viceroy of India in 1894. His viceroyalty was a period of economic stringency and Indian unrest, further complicated by frontier wars. Upon relinquishing his appointment and returning to England in 1899, Elgin was made knight of the Garter. During 1902–03 he acted as chairman of the royal commission which investigated the conduct of the South African War. From 1905 until 1908, when he retired from public life, Elgin served as secretary of state for the colonies in Sir Henry Campbell-Bannerman's administration.

Elgin died at Dunfermline, Fifeshire, on Jan. 18, 1917.

Modest and retiring, disliking the requisite pomp, Elgin was not regarded as a successful or effective viceroy of India. His tenure of the colonial office saw the generous and much-praised settlement with the Boers in South Africa. However, this was primarily the work of the prime minister, Sir Henry Campbell-Bannerman; and Elgin, by temperament a Whig, had little sympathy for his colleagues' more radical innovations in imperial government, being particularly opposed to Lord Morley's proposals for Indian reform.

He was succeeded by his eldest son, EDWARD JAMES (1881–), 10th earl of Elgin. (A. F. T.)

ELGINSHIRE: *see* MORAY.

ELGON, MOUNT, is a Late Tertiary volcano situated astride the Kenya-Uganda boundary in approximately 1° N. and 34° 30′ E. Its extrusions cover about 1,250 sq.mi. and consist largely of fragmental rocks (tuffs and agglomerates) and only to a small extent of lavas. Fossil vegetation within the volcanics and in the underlying Bugisu sedimentaries helps in dating the mountain. The general slope of Elgon is gentle and its outline unimpressive. On the east and southeast at about 6,200 ft. its relief merges with that of the Uasin Gishu plateau, but in the west and northwest spectacular cliffs dominate the 3,600-ft. plains of eastern Uganda. The *caldera* rim, about 5 mi. in diameter, contains several peaks of which Wagagai (14,178 ft.) is the highest. In the summit zone there is ample evidence of former glaciation: moraines reach down the outer slopes as far as 11,000 ft., and there are cirques and rock tarns around the *caldera* rim. On the uneven floor of the *caldera* are considerable areas of swamp, tapped by the Suam river which flows through a deep cleft in the northeastern part of the rim and as the Turkwell ultimately reaches Lake Rudolf. The other numerous streams which furrow the slopes drain either via the Nzoia river to Lake Victoria or westward into the Lake Kyoga system. Caves have been formed by solution in ash bands between more resistant agglomerate beds.

The moorland zone, containing tree heaths and giant groundsels and lobelias, extends down to 10,000 ft., where it is succeeded by bamboo forest. Below 8,300 ft. is a temperate deciduous forest, especially rich in the west and south. The Bantu-speaking Bagisu, cultivators of coffee, bananas, millet and maize are in intensive occupation of the western slopes below the forest reserve. The Nilo-Hamitic Sebei are confined to the northern side; the kindred Elgonyi (Kony) formerly lived on the southern slopes. Elgonyi was the Masai name for the mountain, but the Lugisu name, Masaba, remains current in the Bantu languages. Joseph Thomson visited the southern side of Elgon in 1883; in 1890 F. J. Jackson and Ernest Gedge made a north-south traverse of the *caldera*.

See E. Nilsson, "Quaternary Glaciations and Pluvial Lakes in British East Africa," *Geogr. Ann., Stockh.*, vol. xiii (1931); K. A. Davies, *The Building of Mt. Elgon* in "Uganda, Geological Survey" (1952). (S. J. K. B.)

EL GRECO: *see* GRECO, EL.

ELI (HELI in the Douai version of the Bible), a member of the ancient priesthood founded in Egypt (I Sam. ii, 27), priest of the temple of Shiloh, the sanctuary of the ark, and also judge over Israel. In the history preserved to us he appears in the weakness of extreme old age, unable to control the petulance and rapacity of his sons. While the central authority was thus weakened, the Philistines advanced against Israel and gained a complete victory in the great battle of Ebenezer, where the ark was taken. On hearing the news Eli fell from his seat and died. After these events the sanctuary of Shiloh appears to have been destroyed and the house of Eli settled at Nob. In the fall of his house later writers saw the fulfillment of the prophecies of judgment uttered in the days of Eli against his corrupt house (I Sam. ii, 27 *et seq.*; iii, 12 *et seq.*).

ELIAS OF CORTONA (*c.* 1180–1253), disciple of St. Francis of Assisi (*q.v.*) and general of the Franciscan order, was born near Assisi (he is called "of Cortona" from the place of his death). In 1217 he headed the new Franciscan mission to the Holy Land, being designated first minister provincial of Syria. After visiting the holy places with St. Francis, he returned with him to Italy in 1220. Elias played a leading role in early Franciscan history. Francis' vicar from 1221, he governed the order from Francis' death (1226) until the general chapter of 1227. Later he was commissioned by Gregory IX to build the basilica of Assisi. Elected general of the order in 1232, he did much good, but in 1239 he was deposed because of his despotic rule. He withdrew to Assisi and then to Cortona, where in 1245 he built a church and cloister in honour of St. Francis. Supporting the antipapal policy of the emperor Frederick II, he was excommunicated in 1240 and again in 1244, when Frederick made him ambassador to Nicaea and Constantinople. Toward Easter 1253 he fell sick and was induced to make his peace with the order and the church before his death on April 22.

BIBLIOGRAPHY.—E. Lempp, *Frère Elie de Cortone* (1901); D. Sparacio, *Frate Elia* (1923); S. Attal, *Frate Elia, Compagno di S. Francesco* (1936). (I. C. BY.)

ELÍAS PIÑA (until 1961, SAN RAFAEL), a province in the extreme western part of the Dominican Republic. Area 690 sq.mi. It is an upland plain drained by the Artibonito river, which flows westward into Haiti, and is bounded on the north by the Cordillera Central and on the south by the Sierra de Neiba. Its main crops are peanuts, corn, bananas and coffee. Livestock raising and lumbering are important. The province was created in 1942 from the western part of Benefactor province under the plan to develop the frontier regions on the Haitian border. Pop. (1960) 43,266. The capital, Elías Piña (pop. [1960] 2,890), was called Comendador until 1930. (D. R. D.)

ELIE AND EARLSFERRY, a royal burgh of Fife, Scot., on the Firth of Forth, with a quiet harbour. Pop. (1961) 1,128. It is 12 mi. S. of St. Andrews by road but 20 mi. distant by the railway, which follows the coast. Though it retains a ruined castle and some old houses, and the parish church dates from 1639, Elie is mostly modern. It is a holiday resort, with a fine golf course and yellow sands. Elie harbour, then important, was granted a charter as a port in 1601. The royal burgh of Earlsferry was amalgamated with it in 1929. Its charter, granted by Malcolm Canmore, was burned; it was afterward renewed by James VI. The place derived its name from its use by the earls of Fife as a ferry to the opposite shore of East Lothian (Haddingtonshire), 8 mi. distant. Macduff's cave near Kincraig point is traditionally that in which the thegn took refuge from Macbeth.

ÉLIE DE BEAUMONT: *see* BEAUMONT, (JEAN BAPTISTE ARMAND LOUIS LÉONCE) ÉLIE DE.

ELIJAH (Greek ELIAS), a Hebrew prophet from Tishbe in east-Jordanian Gilead, who, as recounted in the Books of Kings, saved the religion of Yahweh from being corrupted by the nature-worship of Baal. His decisive intervention ranked him with Moses, the founder of Israel as a nation and religious group, and their two names appear together in Old Testament prophecy, which promised Elijah's return as precursor of "the day of Yahweh" (Mal. iv, 4–6). Later Jewish tradition saw him as the great helper in need and the herald or partner of the Messiah. Thus in the New Testament John the Baptist, and even Jesus were held by some to be Elijah in person. Early Christian tradition in general held that his function was fulfilled only by John the Baptist, though Elijah himself appears in glory with Moses at the Transfiguration, to testify to Jesus' mission (Mark ix, 4, etc.).

His name, also transliterated Eliyyahu or Eliyyah, means "Yahweh is (my) God"; the Greek form, Elias, was taken over by the Vulgate. In both the Eastern and the Western Church Elijah is commemorated on July 20. Islam also recognizes him as a prophet.

Sources.—The brilliant and dramatic narratives concerning Elijah were probably composed before 800 B.C., as there is no trace of the polemic of later prophets against the sanctuaries of Dan and Bethel. The tradition may well come from the prophetic guilds which Elijah influenced at least through his disciple Elisha (*cf.* II Kings ii). Hence there is good ground for ascribing the doctrinal content of the traditions to Elijah himself.

Story.—These traditions relate Elijah's sudden appearance under King Ahab to proclaim a drought in punishment of the cult of Baal (either the Tyrian city-god, Melkart, or, more probably, Baal Shamem, lord of the heavens). Elijah, hiding from the king's anger, is fed by ravens and then sheltered by a Phoenician widow, whose food is miraculously multiplied and whose son is brought back from the dead by the prophet. On Mt. Carmel, a trial of strength with 450 prophets of Baal ends in the manifest victory of Yahweh, the God of Israel, and the drought ends with "a little cloud like a man's hand rising out of the sea." A pilgrimage to Mt. Horeb (Sinai) shows Elijah first disheartened by the long loneliness of his fight, then miraculously sustained and encouraged. Ahab's judicial murder of Naboth and confiscation of his vineyard

call forth Elijah as upholder of the moral law, as before he had come forward as champion of monotheism. Later Ahab's son King Azariah appeals to the Baal of Ekron, and Elijah once more upholds the exclusive rights of Yahweh by bringing down "fire from heaven." Finally he is taken up to heaven in a whirlwind. This last episode properly belongs to another cycle of tradition, that of Elisha (II Kings ii–xiii), on whom fell "the mantle of Elijah."

Historical Setting.—The Israelite king Omri allied himself with the Phoenician cities of the coast. His son Ahab was married to Jezebel, daughter of Ethbaal, king of Tyre and Sidon. Jezebel, with her Tyrian suite and a large cult personnel (I Kings xviii, 19, II Kings x, 19), cherished her native religion in a sanctuary built for Baal in the royal city of Samaria (I Kings xvi, 31–32). This meant not that Baalism became the official cult of the state but that the people accepted Baal as well as Yahweh, thus putting Yahweh on a par with a nature-god whose supreme manifestations were the elements and biological fertility, celebrated often in an orgiastic cult. Jezebel was ruthless in repressing opposition to Baal, and Elijah became the heroic embodiment of the Israelite tradition.

Theological Significance.—Epoch-making in the history of monotheism was the climax of Elijah's struggle with Baalism. The memorable words, "If Yahweh is God, follow him, but if Baal, then follow him"—especially when taken with the prayer "Hear me, Yahweh, that this people may know that you, Yahweh, are God"—show that more is at stake than allotting to divinities their particular spheres of influence (I Kings xviii, 21, 37). The incisive and momentous question is whether Yahweh or Baal is God, simply and universally. Elijah's words proclaim that there is no reality except the God of Israel, there are no other beings entitled to the name of divinity. The ridicule he heaps on Baal reduces the god to less than a demon; he is impotent because nonexistent. The acclamation of the people, "Yahweh, he is God" (I Kings xviii, 39), expresses a fully conscious monotheism, never before perhaps brought home to them so clearly.

The deepest prophetic experience of Elijah took place on his pilgrimage to Horeb, where he learns that God is not in the storm, the earthquake or the lightning. Nature, so far from being God's embodiment, is not even an adequate symbol. God is invisible and spiritual, best known in the intellectual word of revelation, "the still, small voice" (I Kings xix, 12). The transcendance of God receives here one of its earliest expressions. In the story of Elijah is also expressed for the first time a thought that was to dominate Hebrew prophecy. In contrast to the bland hopes of the people, salvation is bestowed only on a "remnant," those purified by God's judgment, "all the knees who have not bowed to Baal" (I Kings xix, 18). Elijah was also the first to see the nations united to God, if only as the instrument of his justice against Israel (I Kings xix, 15). The theme of the later prophets, that morality must be at the heart of ritual worship, is also the teaching of Elijah, who upholds the unity of law and religion against the despotic cruelty of a king influenced by a pagan wife (I Kings xxi). His sudden, mysterious appearances remained a symbol of the unfettered initiative of God in history. The remote austerity of his life was an impressive testimony to other-worldly values, just as his being taken to heaven showed Israel the possibility of a life beyond this. His work may also be regarded as a protest against every effort to find religious experience in self-induced ecstasy and sensual frenzy rather than in a faith linked with reason and morality.

BIBLIOGRAPHY.—*See* commentaries cited in bibliography to KINGS, BOOKS OF. *See* also H. L. Strack and P. Billerbeck, *Kommentar zum Neuen Testament aus Talmud und Midrasch*, vol. iv, pp. 764–798 (1928); J. Jeremias, *s.v.* Elias, in G. Kittel (ed.), *Theologisches Wörterbuch zum Neuen Testament*, vol. ii, pp. 930–943 (1935); *Élie* (Études Carmélitaines, 1956); G. Fohrer, *Elia* (1957); H. H. Rowley, *Elijah on Mount Carmel* (1960); S. M. Segal, *Elijah, a study in Jewish Folklore* (1935). (KN. S.)

ELIJAH BEN SOLOMON (1720–1797), known as ELIJAH VILNA, HAGRA or the VILNA GAON, who was born near Brzesc (or perhaps at Vilna), in Lithuania, on April 23, 1720, left a lasting impress upon many phases of Jewish learning. Saint as well as scholar, he realized in his personality the ideals distinctive of Lithuanian Jewry—love of learning, extreme piety and meticulous scrupulousness in every sphere of conduct. As a mark of nearly universal reverence, the title gaon, borne by the heads of the Babylonian academies and virtually extinct for many centuries, was bestowed upon him by the people. His scholarship embraced mastery of every field of study in the millennial Jewish literature up to his own time—mathematics and astronomy, philosophy and Cabala, law and textual criticism.

Elijah's most important contribution was his synoptic view of Jewish learning and his critical methods of study. In an age of narrow, puritanical piety, he broadened the conception of Torah learning to include secular sciences, on the ground that a complete understanding of Torah necessitated the study of mathematics, astronomy, history and Hebrew grammar. He wrote treatises in Hebrew on trigonometry, astronomy and Hebrew grammar, and encouraged translations of works on these subjects into Hebrew. He introduced the methods of textual criticism in the study of Jewish law and mysticism. In general, his influence was felt in the direction of an increased emphasis on rationalism and synthesis. The hints contained in his cryptic remarks and comment on the classic texts of Judaism were studied assiduously by many generations of scholars. However, Elijah's rationalism was associated with an extremely zealous reverence for tradition and a disdain of popular needs. Elijah led the battle against Hasidism, which he suspected of pantheistic and pseudomessianic leanings. His authority cemented the opposition of the Mithnagdim (opponents) to Hasidism.

Elijah died at Vilna on Oct. 9, 1797. His philosophy of Judaism was continued by his disciple, Hayim ben Isaac of Volozhin, who founded the great Volozhin (Lith.) academy, an institution which trained several generations of scholars, rabbis and leaders.

See S. Schechter, *Studies in Judaism* (1924); M. Waxman, *A History of Jewish Literature*, vol. iii (1936).

ELIOT, CHARLES WILLIAM (1834–1926), U.S. educator and leader in public affairs, president of Harvard university for 40 years and editor of the 50-volume *Harvard Classics* (1909–10), the famous "five-foot shelf of books," was born in Boston, Mass., on March 20, 1834. A graduate of Harvard (1853), he taught mathematics and science there from 1854 to 1863. In 1865, following two years of study and travel in Europe, he became professor of analytical chemistry at the Massachusetts Institute of Technology. There he served until 1869, when he was appointed president of Harvard, which, when he retired in 1909, he had successfully elevated into an institution of world renown.

Contending that higher learning in the United States needed to be "broadened, deepened and invigorated," Eliot demanded a place for the sciences along with the humanities in any sound program of liberal education. To counter the rigidity of Harvard's curriculum—which, following the general practice, was almost totally prescribed—Eliot successfully championed the elimination of all required courses. Though such free election was opposed, under Eliot it was maintained, and it was not until the regime of his successor, A. Lawrence Lowell, that a compromise between it and the old prescriptive system was inaugurated.

Eliot's influence reached into secondary education. During his presidency Harvard raised its entrance requirements, and some other major colleges did likewise. This in turn effected a corresponding rise in secondary school standards. Eliot supported correlation of certain high school subjects, greater thoroughness in the teaching of written English and recognition of the particular needs of students not preparing for college. In the report of a national committee on secondary education (1892) he urged the reorganization of American public education to introduce the study of foreign languages and mathematics during the student's seventh school year. The idea was embodied later (1910) in the junior high school.

After retirement Eliot maintained his broad interest in public and educational matters. He supported strong participation in the League of Nations, some sort of military training, profit sharing between labour and management and national prohibition. He died at Northeast Harbor, Me., on Aug. 22, 1926.

The "unique place in American public life" occupied by Eliot was described by William Allan Neilson: "For 40 years the head of a great university, he exercised on all classes of his countrymen an influence far beyond that of the usual academic dignitary. During the latter part of his presidency, and still more after he became emeritus, he was looked up to by hundreds of thousands of his fellow-citizens as a guide not merely in educational matters, but in all the great questions that have agitated the public mind, political, industrial, social and moral. Other leaders of opinion have come and gone, and some for a time have been more conspicuous; but it is impossible to name a figure who so continuously dominated our intellectual horizon for so long a period."

Eliot's writings include *A Compendious Manual of Qualitative Chemical Analysis* with F. H. Storer (1869), many times reissued and revised; *The Happy Life* (1896); *American Contributions to Civilization, and Other Essays and Addresses* (1897); *Educational Reform, Essays and Addresses 1869–1897* (1898); *More Money for the Public Schools* (1903); *University Administration* (1908); *The Durable Satisfactions of Life* (1910); *The Conflict Between Individualism and Collectivism in a Democracy* (1910); and *A Late Harvest* (1924). His annual reports as president of Harvard were notable contributions to the literature of education.

See Henry James, *Charles W. Eliot,* 2 vol. (1930), which lists 14 pages of bibliography, including Eliot's books, articles and published speeches. (A. E. M.; X.)

ELIOT, GEORGE, the pen-name of Mary Anne Evans (1819–1880), one of the great English novelists of the Victorian age, whose novels developed the method of psychological analysis characteristic of modern fiction. She was born at Chilvers Coton, Warwickshire, Nov. 22, 1819, on the Arbury estate of Francis Newdigate, for whom her father was agent. Her command of the Midland dialects is not strange, for she heard little else until she went as a boarder to Mrs. Wallington's school at Nuneaton (1828–32). There she came under the influence of the principal governess Maria Lewis, who along with a formal Johnsonian diction inculcated a strong evangelical piety in the impressible young girl. At her last school (1832–35), conducted by the daughters of the Rev. Francis Franklin, the Baptist minister at Coventry, her religious ardour increased. She dressed severely and engaged earnestly in good works. The school gave her a reading knowledge of French and Italian, and after her mother's death had compelled her to return home to keep house for her father, he let her have lessons in Latin and German. When he turned his business over to his son Isaac and left Griff house, she moved with him in March 1841 to Foleshill road, Coventry.

There she became acquainted with a prosperous ribbon manufacturer Charles Bray (1811–84), a self-taught freethinker, who campaigned boldly for secular education, trades unions, extension of the franchise and other radical causes. Mrs. Bray came of an earnest Unitarian family. Her sister Sara Hennell was immersed in theological and philosophical speculations and her brother Charles Hennell was the author of *An Inquiry concerning the Origin of Christianity* (1838, 2nd ed. 1841), in which, independently of D. F. Strauss, he reached the rationalist conclusions of the higher criticism. Miss Evans' reading of the book merely precipitated a break from orthodoxy which had been long in preparation. Her favourite Scott's novels had early shown her that papists might have noble characters; and she knew many dissenters like the Franklins, the Hennells, and her aunt Mrs. Samuel Evans, a Methodist preacher, who lived better lives than many churchmen. The works of Isaac Taylor (1787–1865) and various books on the relation between the Bible and science had instilled in Miss Evans' keen mind the very doubts they were written to dispel. In 1842 she told her father that she could no longer go to church. The ensuing storm raged for several months before they reached a compromise leaving her free to think what she pleased so long as she appeared respectably at church, and she lived with him until his death in 1849.

The Brays and the Hennells quickly drew her from extreme provincialism, introducing her to many ideas in violent antithesis with her father's Tory principles. When Charles Hennell married in 1843, she took over from his wife the translating of Strauss's *Das Leben Jesu.* Published anonymously as *The Life of Jesus* (3 vol., 1846), it had a profound influence on English rationalism. After the wedding Mrs. Hennell's father R. H. Brabant invited Miss Evans to visit at Devizes. A rather silly man, he had worked for years on a book (never completed) that was to dispose forever of the supernatural elements in religion. They read German and Greek together and discussed theology on long walks; soon Mrs. Brabant became jealous of their intimacy and, before the term of her visit, Miss Evans was forced to leave. Mrs. Hennell felt that her father had acted ungenerously. Out of the humiliation of this episode George Eliot was to draw the horrible vividness of Mr. Casaubon in *Middlemarch.*

She spent the winter of 1849–50 at Geneva, reading widely, living in the family of Francois D'Albert-Durade, who painted a portrait of her. Like those by Mrs. Bray (1842) and Sir Frederic Burton (1865), all in the National Portrait gallery, it shows her with light brown hair, gray-blue eyes, and a very fair complexion. The common misconception of George Eliot as dark-haired and sallow derives from Rajon's etching of the Burton portrait. Returning to Coventry, she spent the rest of 1850 with the Brays, pondering how to live on the £100 a year left by her father. After John Chapman, the publisher of *The Life of Jesus,* got her a chance to review R. W. Mackay's *The Progress of the Intellect* in the *Westminster Review* (Jan. 1851), she decided to settle in London as a free-lance writer, and in Jan. 1851 she went to board with the Chapmans at 142, Strand.

Soon after her arrival Mrs. Chapman and the children's governess, who was Chapman's mistress, became jealous of Marian, as she now signed her name; and after ten weeks she returned to Coventry in tears. Doubtless her feelings were strongly attracted to the magnetic Chapman, whose diary supplies these details, but there is no evidence that she was ever his mistress. A few months later he bought the *Westminster Review,* and when he had explained his domestic complications, Marian, contrite at the pain she had unwittingly caused, returned to London. For three years, until 1854, she served as subeditor of the *Westminster,* which under her influence enjoyed its most brilliant run since the days of John Stuart Mill. At the Chapmans' evening parties she met many notable literary figures in an atmosphere of political and religious radicalism. Across the Strand lived the subeditor of the *Economist,* Herbert Spencer (*q.v.*), whose *Social Statics* (1851) Chapman had just published. Marian shared many of Spencer's interests and saw so much of him that it was soon rumoured that they were engaged. Though he did not become her husband, he introduced her to the two men who did.

George Henry Lewes (*q.v.*) was the most versatile of Victorian journalists. In 1841 he had married Agnes Jervis, by whom he had three sons. In 1850 Lewes and his friend Thornton Leigh Hunt founded a radical weekly called the *Leader,* for which he wrote the literary and theatrical sections. In April 1850, two weeks after the first number appeared, Agnes Lewes gave birth to a son whose father was Thornton Hunt. Lewes, being a man of liberal views, had the child registered as Edmund Lewes and remained on friendly terms with his wife and Hunt. But after she bore Hunt a second child in Oct. 1851, Lewes ceased to regard her as his wife, though, having condoned the adultery, he was precluded from suing for divorce. At this moment of dejection, his home hopelessly broken up, he met Marian Evans. They consulted about articles and went to plays and operas that Lewes reviewed for the *Leader.* Convinced that his break with Agnes was irrevocable, Marian determined to live openly with Lewes as his wife. In July 1854, after the publication of her translation of Feuerbach's *Essence of Christianity,* they went to Germany together. In all but the legal form it was a marriage, which continued happily until Lewes' death in 1878. "Women who are content with light and easily broken ties," she told Mrs. Bray, "do *not* act as I have done. They obtain what they desire and are still invited to dinner."

At Weimar and Berlin she wrote some of her best essays for the *Westminster* and translated Spinoza's *Ethics* (still unpublished), while Lewes worked on his pioneer life of Goethe. By his pen alone he had to support his three sons at school in Switzerland as well as Agnes, whom he gave £100 a year, which was continued

until her death in 1902. She had four children by Hunt, the last born in 1857, all registered under Lewes' name. The few friends who knew the facts agreed that toward Agnes his conduct was more than generous, but there was a good deal of malicious gossip about the "strong-minded woman" who had "run off with" her husband. Marian's deepest regret was that her act isolated her from her family in Warwickshire. She turned to early memories, and, encouraged by Lewes, wrote a story about an episode in Chilvers Coton parish during her childhood. Published in *Blackwood's Magazine* (1857) as "The Sad Fortunes of the Reverend Amos Barton," it made an instant success. Two more tales, "Mr. Gilfil's Love-Story" and "Janet's Repentance," also based on local events, appeared serially in the same year and Blackwood republished all three as *Scenes of Clerical Life* (2 vol., 1858) under the pseudonym George Eliot.

Adam Bede (3 vol., 1859), her first long novel, she described as "a country story—full of the breath of cows and the scent of hay." Its masterly realism—"the faithful representing of commonplace things"—brought to English fiction the same truthful observation of minute detail that Ruskin was commending in the Pre-Raphaelites. The book is rich in humour. The germ of the plot was an anecdote her Methodist aunt told of visiting a girl condemned for child murder. The dialect of the Bedes she had heard her Derbyshire uncles speaking with her father, some of whose early experiences she assigned to Adam. But what was new in English fiction was the combination of deep human sympathy and rigorous moral judgment. *Adam Bede* went through eight printings within a year, and Blackwood doubled the £800 paid for it and returned the copyright.

In *The Mill on the Floss* (3 vol., 1860) she returned again to the scenes of her early life. Maggie Tulliver's devotion to her brother Tom must have made poignant reading for Isaac Evans. The first half of the book with its remarkable portrayal of childhood is irresistibly appealing and the study of Maggie's mingled motives reaches a new level of psychological subtlety. While the tragic ending in the flood was the first thing fixed on for the story, it does not strike every reader as inevitable.

At this time historical novels were in vogue and when Lewes during their visit to Florence in 1860 suggested Savonarola as a good subject, George Eliot grasped it enthusiastically and began to plan *Romola* (1861–62). First, however, she wrote *Silas Marner* (1861), which had thrust itself between her and the Italian material. Its brevity and perfection of form made this story of the weaver whose lost gold is replaced by a strayed child the best-known of her books, though it has suffered unfairly from being forced on generations of school children. *Romola* was planned as a serial for *Blackwood's* until an offer of £10,000 from the *Cornhill Magazine* induced George Eliot to desert her old publisher; but rather than divide the book into the 16 installments the editor wanted, she accepted £3,000 less, an evidence of artistic integrity few writers would have shown. Details of Florentine history, setting, costume and dialogue were scrupulously studied at the British museum and during a second trip to Italy in 1861. It was published in 14 parts between July 1862 and Aug. 1863. Against the sneer that the book smells of the lamp one may place the admiration of Robert Browning, who knew Florence well, and Bernard Berenson, who considered it the best of George Eliot's works. The threads of Tito's egoism are dissected with particular skill. Though the book lacks the spontaneity of the English stories, it has been unduly disparaged.

George Eliot's next two novels are laid in England at the time of the Reform bill. In *Felix Holt the Radical* (3 vol., 1866) she drew the election riot from recollection of one she saw at Nuneaton in Dec. 1832. The initial impulse of the book was not the political theme but the tragic character of Mrs. Transome, one of her greatest triumphs. The intricate plot popular taste then demanded now tells against the novel. *Middlemarch* (8 parts, 1871–72) is by general consent George Eliot's masterpiece. Under her hand the novel had developed from a mere entertainment into a highly intellectual form of art. Every class of Middlemarch society is depicted from the landed gentry and clergy to the manufacturers and professional men, the shopkeepers, publicans, farmers and labourers. Several strands of plot are interwoven to reinforce each other by contrast and parallel. Yet the story depends not on close-knit intrigue but on showing the incalculably diffusive effect of the unhistoric acts of those who "lived faithfully a hidden life and rest in unvisited tombs."

Daniel Deronda (8 parts, 1876), in which George Eliot comes nearest the contemporary scene, is built on the contrast between Mirah Cohen, a poor Jewish girl, and the upper-class Gwendolen Harleth, who marries for money and regrets it. The less convincingly realized hero Daniel, after discovering that he is Jewish, marries Mirah and departs for Palestine to establish a home for his nation. The picture of the Cohen family evoked grateful praise from Jewish readers. But the best part of *Daniel Deronda* is the keen analysis of Gwendolen's character, which seems to many critics the peak of George Eliot's achievement.

Her other works include *The Spanish Gypsy* (1868), a drama in blank verse; *The Legend of Jubal and Other Poems* (1874), which ends with her best-known lines, "Oh, May I Join the Choir Invisible"; *Impressions of Theophrastus Such* (1879), a collection of moral essays; and a posthumous volume *Essays and Leaves from a Note-Book* (1884), mostly reprinted from journals.

In 1863 the Leweses bought the Priory, 21, North Bank, Regent's park, where their Sunday afternoons became a brilliant feature of Victorian life. There on Nov. 30, 1878, Lewes died. For nearly 25 years he had fostered her genius and managed all the practical details of life, which now fell upon her. Most of all she missed the encouragement that alone made it possible for her to write. For months she saw no one but his son Charles Lee Lewes; she devoted herself to completing the last volume of his *Problems of Life and Mind* (1873–79) and founded the George Henry Lewes Studentship in Physiology at Cambridge. For some years her investments had been in the hands of John Walter Cross (1840–1924), a banker to whose family Spencer had introduced the Leweses. Cross's mother had died a week after Lewes. Drawn by sympathy and the need for advice, George Eliot soon began to lean on him for affection too. On May 6, 1880, they were married at St. George's, Hanover square. Cross was 40; she was in her 61st year. After a wedding trip in Italy they returned to her country house at Witley before moving to 4, Cheyne walk, Chelsea, where on Dec. 22, 1880, George Eliot died. She was buried at Highgate cemetery.

BIBLIOGRAPHY.—J. W. Cross, *George Eliot's Life as Related in Her Letters and Journals*, 3 vol. (1885; rev. ed., 1 vol., 1887); *The George Eliot Letters*, ed. by Gordon S. Haight, 7 vol. (1954–55); Leslie Stephen, *George Eliot* (1902); Gordon S. Haight, *George Eliot and John Chapman* (1940); Joan Bennett, *George Eliot: Her Mind and Art* (1948); F. R. Leavis, *The Great Tradition* (1948); Barbara Hardy, *The Novels of George Eliot* (1959); James D. Barry, "The Literary Reputation of George Eliot's Fiction. A Supplementary Bibliography," *Bulletin of Bibliography*, xx, pp. 176–182 (1959). (G. S. H.)

ELIOT, SIR JOHN (1592–1632), English parliamentarian and orator, took a leading part in the early conflicts between Charles I and the house of commons, and died in the Tower of London where he was imprisoned for his defense of parliamentary liberties. He was born at St. Germans, Cornwall, on April 11, 1592, the only child of Richard Eliot, a wealthy west-country landowner. John was probably educated at Blundell's grammar school at Tiverton and he studied at Exeter college, Oxford, though he did not take a degree. At the age of 17 he married Radigund Gedy, sole heiress of a rising Cornish gentleman. Two months after the marriage his father died.

Before settling down to manage his estates Eliot traveled on the continent where he met George Villiers, later 1st duke of Buckingham (*q.v.*). Eliot became member of parliament for his native borough of St. Germans in April 1614 and he was knighted through Buckingham's influence in May 1618. He was not in the parliament of 1621, but in Dec. 1622 Buckingham, then lord admiral, appointed him vice-admiral of Devon. The outset of Eliot's career was clouded, for when in June 1623 he succeeded in arresting a well-known pirate, Capt. John Nutt, the latter turned the tables on him by accusing him of bribery and corruption and for nearly six months Eliot found himself in prison until released by Buckingham's intervention. Shortly after his release Eliot was elected

member of parliament for Newport in Cornwall. He won his reputation as an orator in the session of 1624; he was described by another member as "a spruce young man, and as spruce a speaker." In those days it was common enough for royal officials to attack government policies so long as they did not criticize the king personally, and Eliot, while he praised the king and Buckingham, also condemned the government's illegal taxes and defended freedom of speech in the house of commons.

Between the parliaments of 1624 and 1626 Eliot became alienated from his patron and superior. In the summer of 1624 he had become involved in a quarrel with Sir James Bagg, the vice-admiral of Cornwall, whose influence with Buckingham was greater than his own. Moreover Eliot was in Dec. 1625 deeply upset by the disastrous failure of the expedition to Cádiz which Buckingham had organized. In the parliament of 1626, for which he was again elected member for St. Germans, Eliot sharply attacked the government in a carefully prepared speech. Soon afterward he openly condemned Buckingham for the mistakes and extravagances of his foreign policy and became one of the managers of the impeachment proceedings against him. On May 11 Charles I ordered Eliot's arrest and sent him to the Tower of London. He was released some days later, just before the dissolution of parliament.

During the interval between parliaments the wrath of the court fell upon Eliot. A commission was appointed to examine his accounts and as a result of its findings, he was suspended from his vice-admiralty. Meanwhile in June 1627 he was one of 75 prominent gentry imprisoned for refusing to contribute to a royal loan. Released in Jan. 1628, he was now a recognized leader of the opposition to the court, a great power in Cornwall and a first-class orator. In the parliament which met in March he made a number of speeches in defense of the liberty of the subject and in favour of the preservation of religion from the high church tendencies of the archbishops. He claimed that the rights of parliament were being violated and resisted the payment of additional taxes for the king's wars until the nation's grievances had been righted. He took part in debates on the Petition of Right, and continued to attack Buckingham, but he left parliament before the end of the session because of the death of his wife.

After the assassination of Buckingham in Aug. 1628 and the reassembly of parliament in Jan. 1629, Eliot assaulted the king's government in even more radical terms. He drew up three resolutions condemning religious policy and the illegal levy of customs duties. Though the king ordered the adjournment of parliament, the speaker of the house of commons was forcibly held down in his chair whilst Eliot's resolutions were carried. Eliot and nine others were then arrested and he was made a close prisoner. In Feb. 1630 the court of king's bench fined him £2,000 and ordered him to be detained in the Tower of London during the king's pleasure unless he admitted his offense and gave security for good behaviour. Standing firm on his parliamentary privileges, Eliot refused to submit.

While in the Tower, Eliot wrote several books including *Negotium posterorum,* an account of the parliament of 1625, *An Apology for Socrates,* and *The Monarchie of Man,* an essay on politics and ethics. His close confinement in the Tower undermined his health. Charles I, who never forgave him, refused him permission to leave the Tower and when he died there on Nov. 28, 1632, would not allow his body to be taken home to Cornwall. Instead he ordered: "Let Sir John Eliot's body be buried in the church of the parish where he died." In Dec. 1667 the house of lords reversed the judgment on Eliot and declared that his imprisonment had been contrary to the freedom and privileges of parliament. Though as an orator Eliot was not always effective and as a political leader was less practical than other parliamentarians, future generations rightly came to regard him as a patriot and a parliamentary hero.

See Harold Hulme, *The Life of Sir John Eliot* (1957). Eliot's most important writings were edited by A. B. Grosart (1879–82).
(M. P. A.)

ELIOT, JOHN (1604–1690), the New England "apostle to the Indians," was born at Widford, Hertfordshire, Eng., in 1604 and educated at Jesus college, Cambridge. He emigrated to Boston in 1631 and from 1632 until his death was pastor of the church at Roxbury. With the support of his congregation and of his fellow ministers he undertook a mission to the Indians, which Cotton Mather declared to be "New England's peculiar glory." Small grants were regularly made by the general court of Massachusetts, but the chief financial aid came from England, where Eliot's work inspired the creation of the Company for Propagating the Gospel in New England and Parts Adjacent in North America (1649; rechartered 1661). Even after the Massachusetts Indian towns were extinct, Eliot's influence continued to expand as his example inspired the rising missionary movement, and his methods set the pattern of Indian missions for almost two centuries.

After linguistic preparation, Eliot began preaching at Nonantun (Newton) in 1646, shortly after the first efforts of Thomas Mayhew, Jr., on Martha's Vineyard. A town of "praying Indians" arose there, but was moved to Natick in 1651. Fourteen such villages in 1674 held 4,000 converts; but they suffered grievously during King Philip's War the following year, when both colonists and pagan Indians vented their wrath upon them. Although the original towns never fully recovered, the Indian mission as a whole expanded throughout New England.

The motivation behind Eliot's work was a desire to glorify God, honour Christ and show compassion to physically and spiritually destitute men. Civilization went hand in glove with evangelization in Eliot's view. Consequently, converts won through public preaching were gathered into Christian towns, governed by a biblical code of laws, and gradually introduced to the English manner of life. A school in each village instructed in elementary subjects and doctrine through the vernacular, taught the English language and trained in handicrafts by which the people could support themselves. After severe testing believers were covenanted into a Puritan "church-state." Native teachers and evangelists were trained and employed. The entire range of literature in the Massachusetts Algonkian language needed for the enterprise was produced by Eliot himself, beginning with the *Catechism* of 1653. This included the great achievement of the translation of the Bible (New Testament in 1661, Old Testament in 1663), the first in any language to be printed in America. His English works were *The Christian Commonwealth* (1659) and *Harmony of the Gospels* (1678). Eliot died on May 21, 1690.

BIBLIOGRAPHY.—*See* the biographical sketches in Cotton Mather's *Magnalia* (1702) and Williston Walker's *Ten New England Leaders* (1901); but for an intimate view of the mission work refer to the contemporary *Eliot Tracts,* published in London, but readily available in reprints. For the great Indian Bible *see* George Parker Winship, *The Cambridge Press, 1638–1692* (1945). (RT. P. B.)

ELIOT, THOMAS STEARNS (1888–1965), American-born British poet and critic, who exercised a strong influence on thought in the period between World Wars I and II. His experiments in diction, style and versification revitalized English poetry, and in a series of critical essays he shattered old orthodoxies and erected new. His pronouncements on social and cultural questions made him a 20th-century Matthew Arnold, and, as a director of the publishers Faber and Faber, he was a generous and discriminating patron of young poets. With the publication of *Four Quartets,* during World War II, he reached a wide public, and recognition of his standing as the greatest living English poet and man of letters came in 1948 with the award of the Order of Merit and the Nobel prize for literature.

Eliot was born in St. Louis, Mo., on Sept. 26, 1888, and was educated at Harvard, Paris and Oxford universities. He came of New England stock, his forbears having left East Coker in Somerset in the 17th century. He established himself in London in 1914 and from 1917 to 1919 he was assistant editor of the *Egoist.* His first volume of poems, *Prufrock and Other Observations,* was published in 1917. Five hundred copies were printed, to be sold at one shilling, and the edition took three years to sell out. In 1922 he founded the *Criterion* (which he edited until its demise in 1939) and in the same year he published *The Waste Land,* the poem by which he first became famous.

The Waste Land was dedicated to Ezra Pound, another U.S. expatriate, to whose criticism it owed much of its condensed force of style. It expressed with great power the disenchantment, dis-

illusionment and disgust of the postwar period. In a series of vignettes, loosely linked by the legend of the search for the Grail, it portrayed a sterile world of panic fears and barren lusts, waiting for some sign or promise of redemption. Its manner was highly allusive and the poet provided notes and references, originally, as he later owned, to fill up space. This learned apparatus distracted some readers and critics from perceiving the true originality of the poem, which lay rather in its rendering of the universal human predicament of man desiring salvation, and in its manipulation of language, than in its range of literary reference. In his earlier poems Eliot had shown himself to be a master of the poetic phrase. *The Waste Land* showed him to be, in addition, a metrist of great virtuosity, capable of astonishing modulations from the sublime to the conversational. In every way *The Waste Land* broke with the post-romantic tradition of 19th-century poetry. Its models were the French Symbolists, Dante and the English metaphysical poets.

During this period Eliot gained an equal reputation as a critic. His first collection of essays, *The Sacred Wood*, appeared in 1920, and three essays on 17th-century poetry, collected under the title *Homage to John Dryden* (1924), were also highly influential. Much of the critical writing of the next 20 years consisted of expansions of, or glosses on, the neoclassic doctrine laid down in the essay on "Tradition and the Individual Talent" in the first book, and the analysis of the merits of metaphysical poetry in the second. Eliot's distaste for the romantic poets and his attack on Milton's diction and syntax arose partly out of his own struggle to create new verse rhythms based on the rhythms of contemporary speech. They were linked too with his concern with cultural problems and his belief that since the 17th century there had been a progressive impoverishment of English culture and a coarsening of sensibility.

In 1927, the year in which he became a British citizen, Eliot was confirmed in the Church of England, and from this time on he became increasingly occupied with sociological and cultural questions. He developed his views through his editorship of the *Criterion* and in such works as *Thoughts after Lambeth* (1931), *After Strange Gods* (1934), *The Idea of a Christian Society* (1939) and *Notes Towards the Definition of Culture* (1948). In these works he appeared opposed to the main intellectual currents of the period and he was regarded by many as reactionary and obscurantist. But, just as his return to Christian orthodoxy anticipated a general movement among intellectuals during the war and immediately after, so his political thinking was more in tune with the postwar revival of conservatism than with the predominantly leftwing thought of the 1930s.

The series of "Ariel Poems," published separately in paper wrappers, and *Ash Wednesday* (1930) expressed the pangs and the strain involved in the acceptance of religious belief and religious discipline. These poems were written in a more relaxed, musical and meditative style than his earlier works in which the dramatic element had been stronger than the lyrical. Eliot's critical principles, his exaltation of objectivity and his belief in the social function of poetry, drew him toward the drama, the most objective and the most social of the arts. After a dramatic fragment, *Sweeney Agonistes* (1932), he wrote a set of choruses for a pageant *The Rock* (1934) and a modern miracle play on the martyrdom of Thomas Becket, *Murder in the Cathedral* (1935). The most striking feature of this, his most successful play, was the use of a chorus in the traditional Greek manner to make apprehensible to common humanity the meaning of the heroic action. *The Family Reunion* (1939) was less popular. It contained scenes of great poignancy and some of the finest dramatic verse since the Elizabethans; but the public found this translation of the story of Orestes into a modern domestic drama baffling, and was uneasy at the mixture of psychological realism, mythical apparitions at a drawing-room window, and a comic chorus of uncles and aunts.

World War II put a stop to theatrical experiments and Eliot turned again to poetry, producing what some critics regard as his finest single work, *Four Quartets* (1943). The first of the Quartets, *Burnt Norton*, had appeared in the *Collected Poems* of 1936. It was a subtle meditation on the nature of time and its relation to eternity. On the model of this Eliot wrote three more poems, *East Coker* (1940), *The Dry Salvages* (1941) and *Little Gidding* (1942), in which he explored through images of great beauty and haunting power his own past, the past of the human race and the meaning of human history. Each of the poems was self-subsistent; but when published together they were seen to make up a single work, in which themes and images recurred and were developed in a musical manner and brought to a final resolution. The work made a deep impression and even those who were unable to accept the poem's informing beliefs recognized the intellectual integrity with which Eliot pursued his high theme, the originality of the form, and the technical mastery of the verse.

After the war Eliot returned to the theatre with *The Cocktail Party* in 1949, *The Confidential Clerk* in 1953 and *The Elder Statesman* in 1958. In these plays Eliot accepted the conventions of the west-end play at their most conventional, subduing his style to a conversational level, eschewing the lyrical passages which gave beauty to his earlier plays, as well as such devices as the use of a chorus and of formalized, semi-antiphonal dialogue. Only *The Cocktail Party* achieved a popular success. In spite of their obvious theatrical defects and a failure to engage the sympathies of the audience for the characters, the plays succeeded in handling moral and religious issues of some complexity while entertaining the audience with farcical plots and some shrewd social satire. But it seems likely that Eliot's plays, like his polemical works, will be valued mainly for the light they throw on his poetry and his literary criticism. It is as a poet and critic that his place in English literature is secure.

Among Eliot's principal works, other than those mentioned, are: poetry—*Collected Poems, 1909–62* (1963), *Old Possum's Book of Practical Cats* (1939); essays—*Selected Essays* (1932; rev. ed., 1951), *The Use of Poetry and the Use of Criticism* (1933); *On Poetry and Poets* (1957). He died in London on Jan. 4, 1965.

BIBLIOGRAPHY.—Donald Gallup, *T. S. Eliot: a Bibliography* (1952); F. O. Matthiessen, *The Achievement of T. S. Eliot*, 3rd ed. (1958); Helen Gardner, *The Art of T. S. Eliot* (1949). (HN. G.)

ELIS (ELEA; modern ILIA) was an ancient Greek state in the northwest corner of the Peloponnese, which took its name from its chief city, Elis, situated on a ridge now called Kalaskopi, on the south bank of the Peneus river. At its greatest extent Elis included (1) Hollow Elis, the basin of the Peneus river and its tributary the Ladon, together with the coastal plain in the west and the mountainous district of Acrorea (the foothills of Mt. Erymanthus) in the east; (2) Pisatis, to the south, roughly coextensive with the plain on the north bank of the Alpheus (Alfios) river; (3) Triphylia, the hilly area between the Alpheus and the northern border of Messenia, the Neda river. Comparatively high rainfall (now 33.8 in. annually) produced good pasture and arable land in the low-lying parts, especially in Hollow Elis, and the area was noted for its cattle, horses and flax. The popularity and prestige of the Olympic games, celebrated (from 776 B.C.) in a sanctuary on the north bank of the Alpheus river (*see* OLYMPIA) also brought much incidental profit to the Eleans. Most of the district is included in the *nomos* (department) of Ilia in the modern Greek administration. Its chief town is Pyrgos (*q.v.*), just to the north of the Alpheus.

Little is known of the Neolithic and Early Bronze Age cultures. During the Mycenaean period (*c.* 1500–1200 B.C.) Hollow Elis, of which the chief towns were Elis and Buprasium, was occupied by an Aeolian tribe, the Epeioi, who also dominated the Arcadian population of Pisatis (*see* ARCADIA); Triphylia by a tribe called the Caucones, also probably of Arcadian stock, who were subject to the kings of Pylos (*q.v.*) in Messenia.

At the time of the Dorian invasion (12th century B.C.) a section of the people known to modern scholars as "northwest Greeks," traditionally led by a certain Oxylus, crossed the Gulf of Corinth from their home in Aetolia and settled in Hollow Elis, spreading later into Pisatis. Perhaps from the start the city of Elis claimed suzerainty over the other cities of Hollow Elis and over those of Pisatis, but in the case of the latter at any rate its control was uncertain and intermittent. So much can be inferred from the tradition of a struggle between Elis and the Pisatans for control

of the Olympic games. Stories survive of a certain Pantaleon and his sons Damophon and Pyrrhus, tyrants of "Pisa" in the second half of the 7th and in the early 6th century B.C. They are probably historical figures, the rulers, not of "Pisa," as there was no city of that name in Pisatis in historical times, but of some one Pisatan city which for a time dominated the others or perhaps led the resistance to Elis. By 572 B.C. at the latest, the area had been subjected.

In 479 B.C. the Eleans were unenthusiastic members of the anti-Persian alliance led by Sparta and soon afterward broke with Sparta, adopted a democratic constitution (*c.* 471 B.C.), probably under the influence of the Athenian Themistocles (*q.v.*), and with it the democratic principle of "synoecism" (union of small townships into one city). The city of Elis thus became the only administrative centre. Later inscriptions mention an assembly, a council of 500 (perhaps on the Athenian pattern; *see* BOULE), and give the titles of some magistrates; but of the working of the constitution little or nothing is known.

In the Peloponnesian War (*q.v.*) the Eleans were again allied to Sparta until 420 when they temporarily joined the alliance of Athens, Argos and Mantinea. This break led Sparta to support the independence of Triphylia which, with its chief cities Lepreum and Mecistus, had been continually threatened by Elis and twice after 471 B.C. almost subdued. Thereafter Elis had to face intermittent Spartan hostility until 365 when Triphylia's cause was taken up by the Arcadians. Later again the Achaean league (*q.v.*) became Elis' chief opponent. But by alliance with the Thebans, then with the Macedonians and later with the Aetolians, and emphasizing throughout the sanctity of the Olympic games and the neutrality of Elis as their president, the Eleans were able to retain their territory and in some sense their independence, even during the Roman occupation of Greece (from 146 B.C.), to disintegrate only with the collapse of the Roman empire. The area was later occupied by the Franks, the Venetians and the Turks (*see* GREECE: *History*). As a *nomos* of modern Greece it retains its agricultural prosperity and has recaptured the position it held in classical times as a tourist centre.

The area contains one of the finest archaeological sites in Greece, Olympia, where the superb excavations of the German institute in Athens are in progress (*see* OLYMPIA). Limited excavations have also taken place at the site of the city of Elis and there have been random finds and some digging at other scattered points, perhaps most notably at the Mycenaean tombs of Kakovatos in Triphylia, formerly Elean Pylos.

BIBLIOGRAPHY.—Pausanias, books v, 1–6 and vi, 23–26; Strabo, *Geography*, pp. 336–359. *See* also A. Philippson, *Die griechischen Landschaften*, vol. iii, ch. 7 (1959). (W. G. F.)

ELISABETHVILLE, a city in the Democratic Republic of the Congo and capital of East Katanga province, lies in the most southerly part of the country 4,032 ft. above sea level. Pop. (1958 est.) 182,638, the majority of whom are African. Although it is near the equator, the altitude gives it a temperate climate. Elisabethville (named after the Belgian queen) was founded in 1910 as an administrative centre for the copper-mining area and has grown also into a commercial and industrial town. It is built on a grid pattern with wide streets, trees and a public park. There are a multiracial university (1955), an academy of fine art, a school of music and the Leopold II museum rich in its geological and ethnological collections. Besides a Roman Catholic cathedral there are Orthodox and Protestant churches and a synagogue. The town is connected by rail with the rest of the colony and with the principal ports of central and southern Africa. The Trans-African roadway between Algiers and the Cape passes through it and roads also connect it with the Atlantic and the Indian oceans. There is an international airport, 4 mi. N., from which regular flights are maintained to Brussels and to other European capitals, and also to some African towns.

The most important industry is that of copper which is mined at Kipushi and smelted at the Lubumbashi works of the Union Minière du Haut-Katanga, the largest mining concern in the republic. Other industries include printing, brewing and the manufacture of flour, confectionery, cigarettes, bricks, soap and hosiery. The farms of the neighbourhood supply Elisabethville and the copper mining areas. *See* also CONGO, DEMOCRATIC REPUBLIC OF THE.

ELISHA (Gr. ELISAEUS or ELISEUS), the Hebrew prophet who succeeded Elijah (*q.v.*) in the struggle to maintain the worship of Yahweh under the kings Joram and Jehu in the face of the invading cult of the Phoenician god Baal. Called by Elijah at Yahweh's command, Elisha received at his master's ascension a double portion of his spirit (II Kings ii, 9–15), being thus approved as his spiritual first-born. The words "the chariots of Israel and its horsemen," in which Elisha summed up the greatness of Elijah, were applied by King Joash to the dying Elisha to testify that he had been equal to his inherited task (II Kings ii, 12; xiii, 14). The human contact maintained by Elisha with the prophetic guilds and the popular circles which were everywhere attached to them procured him a strong support against his adversaries and enabled him to realize the aim not achieved by the greater but more solitary Elijah: the extermination of Baal from Israel. The name Elisha ("God has saved") was transliterated into Greek in the Septuagint as Elisaios, which became Elisaeus in the Vulgate. His feast day is June 14 in both the Eastern and Western Church.

The Elisha cycle (II Kings ii, 1–x, 30; xiii, 10–21) was composed by a scribe, still under the fresh impression of Elisha's personality, who was closely attached to the prophetical guilds. Their traditions he gathered and worked into a loosely connected description of Elisha's activity. His work was later inserted, not always in its original sequence, in the biblical second book of the Kings. Critics distinguish in it two groups of stories.

The first group comprises historical tales which show Elisha on the political scene: as saving the Israelite army from dying of thirst (II Kings iii, 1–20); helping his people against hostile incursions and during the siege of its capital (vi, 8–viii, 20); or overthrowing and anointing kings—Benhadad and Hazael, kings of Syria (viii, 7–15), and Joram and Jehu, kings of Israel (ix, 1–x, 30).

The second group of stories includes so-called prophetical anecdotes in which Elisha, endowed with miraculous science and power, appears: once as cursing the mocking boys of Bethel (II Kings ii, 23–25), but ordinarily as a kindly, understanding helper in need, as in the house of the Shunammite woman whose dead son he brought back to life (iv, 8–37); among the prophetical guilds and their adherents for whom he miraculously multiplied food (iv, 1–7, 38–44; vi, 1–4); and even for Naaman, the captain of the Syrian host, who was cured of his leprosy by obeying Elisha's command to wash seven times in the Jordan (v, 1–27).

Many critics divide the second group strictly from the first, regarding its stories as merely legendary, but this procedure is unwarranted as the literary talent of the same narrator appears in both groups. On the contrary, the miraculous stories are of real importance for the historical view of Elisha. Not only do they paint in fresh and true colours the popular environment wherein he lived and acted, but they reveal an aspect of his character which completes that of the historical tales and contributes to explain his success.

See H. Gunkel, *Geschichten von Elisa* (1925). (H. JU.)

ELISHA BEN ABUYAH (fl. *c.* A.D. 100), a unique figure among the Palestinian Jews of his day, was born before the destruction of the Temple (A.D. 70). He refused to accept the current rabbinical views, though the Talmud cites his legal decisions. Most authorities believe that he was a Gnostic, but it is possible that he was simply a Sadducee, and thus an opponent not of Judaism but of Pharisaism. The efforts of his disciple, the famous Pharisee Meir, to reclaim his former master are among the most pathetic incidents in the Talmud. In later ages Elisha was regarded as the type of a heretic whose pride of intellect betrayed him into infidelity to law and morals.

ELIXIR, in alchemy, the medium which would effect the transmutation of base metals into gold; it probably had a wider meaning than "philosopher's stone." The same term, more fully *elixir vitae*, "elixir of life," was given to the substance which would indefinitely prolong life; it was considered to be closely related to, or even identical with, the substance for transmuting metals.

(*See* ALCHEMY.) In pharmacy an elixir is usually defined as a sweetened hydroalcoholic solution containing flavouring materials and usually medicinal substances. (F. L. A.)

ELIZABETH, SAINT (*c.* 1207–1231), the daughter of Andrew II, king of Hungary, was famed for her charity and love of the poor. Born probably in Pressburg (Bratislava), she was betrothed in infancy to Louis, son of the landgrave Hermann I of Thuringia, at whose court she was brought up. When Louis IV succeeded his father in 1221 the marriage took place. It proved to be an ideally happy one. Six years later, however, Louis died at Otranto on his way to the crusade, and Elizabeth had to leave her home when her brother-in-law assumed the regency. Having no further use for position or wealth, she became a member of the Third order of St. Francis. At Marburg she built a hospice for the poor and afflicted, to whose service she devoted the rest of her life. She put herself under the spiritual direction of Konrad von Marburg, an ascetic of incredible harshness and severity, who belonged to no specific order and whose method of helping her along the road to spiritual perfection was certainly not Franciscan. She died in Marburg on Nov. 17, 1231, and was canonized in 1235. Her feast day is Nov. 19.

Many stories are told of St. Elizabeth's devotion to the poor. The best known is that often depicted in art showing her meeting her husband unexpectedly on one of her errands of charity, when the loaves she was carrying were miraculously changed into roses. This legend is misleading, however, inasmuch as Louis, far from restricting her bounty to the poor, was in fact a sympathetic and loving collaborator.

See A. Huyskens, *Quellenstudien zur Geschichte der hl. Elizabeth* (1908); H. Thurston and D. Attwater (eds.) *Butler's Lives of the Saints*, vol. iv, pp. 386–391 (1956). (AE. C.)

ELIZABETH (ELIZAVETA PETROVNA) (1709–1762), empress of Russia from 1741 to 1762, was born at Kolomenskoe, near Moscow, on Dec. 18 (new style; Dec. 7 old style), 1709, the daughter of Peter the Great and Marta Skowronska. From her earliest years she delighted everyone by her beauty and vivacity, and she developed a keen political and diplomatic sense. Various proposals for her marriage failed, so that on her mother's death (May 1727) and the departure to Holstein of her sister Anna, her only remaining near relation, the 18-year-old princess found herself practically her own mistress. So long as Prince A. D. Menshikov remained in power, she was well treated by the government of Peter II, but the Dolgorukis, who supplanted Menshikov and hated the memory of Peter the Great, practically banished Peter's daughter from court. While still in her teens Elizabeth made a lover of Aleksei Shubin, a sergeant in the Semenovski guards, and after his banishment to Siberia by order of her cousin the empress Anna Ivanovna she consoled herself with a young illiterate Cossack, Aleksei Razumovski, whom she is said to have married. During Anna Ivanovna's reign (1730–40), Elizabeth effaced herself as much as possible; but under the regency of Anna Leopoldovna for Ivan VI the course of events drove her to overthrow the government. The French ambassador, J. J. Trotti, marquis de La Chétardie, who was plotting to destroy the Austrian influence then dominant at the Russian court, seems to have urged her on and lent her a small sum of money, but he took no part in the coup when it came. Elizabeth acted because she had reason to fear imprisonment in a convent for life. On the night of Dec. 5–6 (N.S.), 1741, she assembled her friends and her household, drove to the barracks of the Preobrazhenski guards, enlisted their sympathies by a stirring speech and led them to the winter palace, arresting the ministers on the way. She seized the regent and her children in their beds, summoned all the notables, civil and ecclesiastical, to her presence and was proclaimed empress without further ado.

Fortunately Elizabeth had inherited some of her father's genius for government. After abolishing the cabinet council system in favour during the rule of the two Annas and reconstituting the senate as it had been under Peter the Great with the chiefs of the departments of state, all of them now Russians again, as ex officio members under the presidency of the sovereign—the first care of the new empress was to compose the quarrel with Sweden. In Jan. 1743 direct negotiations between the two powers were opened at Åbo, and in August Sweden ceded to Russia all the southern part of Finland east of the Kymmene river, including the fortresses of Villmanstrand and Fredrikshamn. Much of this success in foreign policy was due to the new vice-chancellor, A. P. Bestuzhev-Ryumin (*q.v.*). He represented the group in the council opposed to France and Prussia and wanted to bring about an Anglo-Austro-Russian alliance. Hence the reiterated attempts of Frederick the Great and Louis XV to get rid of him, which made the Russian court during the earlier years of Elizabeth's reign the centre of a tangle of intrigue. When the peace of Aix-la-Chapelle ended the War of the Austrian Succession (*q.v.*) in 1748, Bestuzhev had extricated his country from the Swedish imbroglio; reconciled Elizabeth with the courts of Vienna and London, her natural allies; restored Russia's influence in Poland, Turkey and Sweden; and isolated the king of Prussia by environing him with hostile alliances. All this would have been impossible but for the steady support of Elizabeth, who trusted Bestuzhev implicitly, despite the insinuations of his innumerable enemies, most of whom were her personal friends.

The great event of Elizabeth's later years was the Seven Years' War (*q.v.*). Elizabeth regarded the treaty of Westminster of 1756 (whereby Great Britain and Prussia agreed to unite their forces to oppose the entry into, or the passage through, Germany of the troops of every foreign power) as utterly subversive of the previous conventions between Great Britain and Russia. Her fear of Frederick the Great, who was "to be reduced within proper limits" so that "he might be no longer a danger to the empire," induced Elizabeth to accede to the Franco-Austrian treaty of Versailles against Prussia: and in May 1757 the Russian army, 85,000 strong, advanced against Königsberg. Neither the serious illness of the empress, which began with a fainting-fit at Tsarskoe Selo (Sept. 1757), nor the fall of Bestuzhev (Feb. 1758), nor the cabals and intrigues of the various foreign powers at St. Petersburg, interfered with the progress of the war, and the crushing defeat of the Prussians at Kunersdorf (Aug. 1759) at last brought Frederick to the verge of ruin. From the end of 1759 to the end of 1761, Elizabeth's unshakable firmness was the one political force that held together the heterogeneous jarring elements of the anti-Prussian combination.

In March 1760 a fresh convention was signed between Russia and Austria, a secret clause of which, never communicated to the French, guaranteed East Prussia to Russia as an indemnity for war expenses. The failure of the campaign of 1760, so far as Russia and France were concerned, induced the French to inform St. Petersburg that they desired peace. Elizabeth's answer was to decline any pacific overtures until the original object of the league had been accomplished. Simultaneously, she proposed in a secret letter to Louis XV a new treaty of closer alliance without Austria's knowledge. Her object seems to have been to reconcile France and Great Britain, in return for which service the French were to throw all their forces into the German war. This bold project foundered upon Louis XV's jealousy of Russian influence in eastern Europe and his fear of offending Turkey. It was finally arranged by the allies that their envoys at Paris should fix the date for a peace congress and that, in the meantime, the war against Prussia should be vigorously prosecuted. Frederick's situation had become desperate when his difficulties were dispersed by the death of the Russian empress, in St. Petersburg, on Jan. 5, 1762 (N.S.; Dec. 25, 1761, O.S.).

See R. N. Bain, *The Daughter of Peter the Great* (1899); K. Waliszewski, *La Dernière des Romanoff: Elisabeth Ire, impératrice de Russie* (1902).

ELIZABETH I (1533–1603), queen of England and Ireland from 1558 to 1603, was born at Greenwich palace on Sunday, Sept. 7, 1533, the only child of Henry VIII by his second wife, Anne Boleyn. Henry had defied the pope's authority and carried England into schism in order to divorce his first wife, Catherine of Aragon, and marry Anne. However, his second marriage, like his first, failed to produce the male heir regarded as essential to the future tranquillity of the realm, and this soon helped to turn him against Anne. Before Elizabeth was three years old, her father had her mother beheaded for alleged adultery and the mar-

riage declared, on other grounds and somewhat inconsistently, invalid from the beginning. The child was perhaps too young and, living from the age of three months in her separate household at Hatfield, certainly too little acquainted with her mother to suffer any deep or lasting impression from her loss. By Henry's act, however, she became illegitimate and lost her position as heir presumptive and her brief precedence over Catherine's daughter Mary. Her legitimacy was never legally re-established, though a statute of 1544 placed her next in the order of succession after Henry's other children, Edward (b. 1537) and Mary, and their issue, an arrangement reaffirmed in the will which Henry was empowered by statute to make. Her childhood, however, was not unhappy. In Henry's last wife, Catherine Parr, she found a second mother; a growing affection bound her to the young Edward, with whom she was brought up; and Mary, 17½ years her senior, grew more friendly once their common illegitimacy banished all questions of precedence. Under tutors chosen from among the leading English exponents of the new learning—William Grindall and, after his death (1548), Roger Ascham—she read widely and zealously in the Greek and Latin classics, became proficient in French, Italian and (later) Spanish and developed a fair Italian handwriting.

Position Under Edward VI and Mary.—Soon after Edward's accession she was brought into some danger by the intrigues of Thomas, Lord Seymour, brother of Edward Seymour, the protector Somerset, and recently married to Henry's widow, Catherine Parr, in whose household she was living. Seymour paid her somewhat indelicate attentions even in his wife's lifetime and, after Catherine's death (Sept. 1548), he was accused of plotting to marry her and overthrow his brother's government. Thus, when he was arrested and executed, Elizabeth and her servants were also subjected to a searching inquiry. The examinations, however, proved only "that she hath a very good wit and nothing is to be gotten of her but by great policy." Having survived this ordeal, she lived unscathed through the rest of the reign. She avoided the enmity and the still more dangerous favour of the duke of Northumberland and rode by Mary's side in triumph into London (Aug. 3, 1553) after the failure of his plot to oust them both by putting his daughter-in-law, Lady Jane Grey, on the throne.

Once again, however, Elizabeth's position grew perilous now that her zealously Roman Catholic sister was queen. Reluctantly she went to Mass, but her very birth identified her with the anti-papal cause and few believed in the sincerity of her conversion. Hence, so long as Mary remained childless, Elizabeth's mere existence cast doubt upon the permanence of England's return to the Roman fold, for neither council nor parliament was ever willing to deprive her of her statutory place as next in the order of succession. Moreover, as soon as Mary's first popularity wore thin and opposition mounted against her project to marry Philip of Spain, her opponents inevitably turned their thoughts and hopes to Elizabeth. She became a present as well as a future menace. It is not known what knowledge she had of Sir Thomas Wyat's plans, but his abortive rebellion brought her to the Tower of London in March 1554 and the Spanish ambassador pressed hard for her execution as a guarantee for Philip's safety. Again, however, searching examinations yielded no conclusive proof of complicity and Mary dared not execute one of Tudor blood upon mere suspicion. Elizabeth's imprisonment was gradually relaxed. After two months in the Tower she was sent to Woodstock in Sir Henry Bedingfield's strict custody. Later she was allowed back at court and by the autumn of 1555 was free to go down to Hatfield. By now Philip of Spain had become her protector for, with Mary ailing and still childless, he preferred even Elizabeth to the strongest Catholic claimant, Mary Stuart, wife of the French dauphin. Elizabeth stayed at Hatfield, not above suspicion but out of serious danger and steadily gaining adherents, while the disastrous French war, the loss of Calais and the burnings of Protestants tore away the last shreds of popularity from Mary, from Spain and from Rome.

Accession.—On Mary's death (Nov. 17, 1558), therefore, Elizabeth's accession was welcomed with almost universal rejoicing. She was an impressive figure, tall, with reddish-gold hair and a somewhat olive complexion. The more critical of contemporary observers thought her comely rather than handsome, but her regal and gracious bearing quickly swept such ungenerous distinctions from her people's minds. From the very beginning of her reign, at her entry into London and again during the coronation ceremonies, she manifested a remarkable mastery in the arts of stimulating popular enthusiasm. She manifested no less promptly her grasp of the art of government. She was never an easy mistress. Her tongue could be rough as well as honeyed; her ways were unpredictable and her intentions often inscrutable. She exasperated her advisers by her neglect of the opportunities and her indifference to the perils which they pointed out. Like her father and grandfather, she was an uncannily sure judge of a risk, especially of the risks attendant upon inaction. Like them, too, she kept the decision of policy in her own hands and within a few weeks of her accession the Spanish ambassador remarked that "she gives her orders and has her way as absolutely as her father did." Yet she was not afraid to choose the ablest men to serve her. She kept the number of her councilors small, but selected them with such care that she hardly ever found it necessary to dismiss them. She was careful, too, not to let their feeling of masculine superiority gain an overpowering strength by too much unity in action. Hence she tolerated, if she did not actually encourage, a certain rivalry among them and among her favourites. Also she preferred to consult them individually rather than as a body and to work through different councilors for different purposes—even the ablest and most trusted of them, William Cecil (later Lord Burghley), could not always make his advice prevail. Further, she was quick to resent any attempt by even the greatest of her favourites to dictate her policy.

Problems of Marriage and Succession.—This determination not to give herself a master probably accounts for her never marrying. Not much weight need be attached to the contemporary gossip that she was debarred from matrimony by some physical defect—that was probably no more than the expression of contemporary astonishment at a young queen deliberately rejecting the normal lot of 16th-century women. She was never the slave of her emotions and was intelligent enough to see that a husband would create more problems than he solved. Among her own subjects, she seems to have felt a real affection for Robert Dudley (later the earl of Leicester). However, the jealousy of his fellows and the all-too-opportune death of his wife, Amy Robsart—and perhaps Elizabeth's remembrance that, in her own words, the Dudleys had been traitors for two generations—ruled Leicester out. Such a match might too easily have had the same outcome as Mary Stuart's marriage to the earl of Bothwell. Among foreign suitors, it was difficult to find one who would not, as a husband, create even greater problems. The sanity of the Swedish Eric XIV and the Scottish James Hamilton, earl of Arran, was dubious. Philip II of Spain, whose condescending proposal Elizabeth politely declined in the spring of 1559, was unlikely to tempt a second English queen. His cousin, the Austrian archduke Charles, was useful as a means of mollifying Spain so long as there was any likelihood of war with France or serious danger from Mary in Scotland, but negotiations always reached a deadlock over Elizabeth's refusal either to marry him unseen or to allow him to worship as a Catholic and so become a potential focus for all Catholic discontents. The risks of French and Scottish hostility never, in Elizabeth's eyes, grew great enough to outweigh the domestic dangers of such an antidote. When, from 1568 onward, Spanish hostility was more to be feared than French or Scottish, the same obstacles were repeatedly placed in the way of her French suitors, Henry of Anjou and, after him, Francis of Alençon. Nevertheless, in the later years of the Alençon courtship Elizabeth—now in her middle forties—did carry flirtation to lengths which mere reasons of state hardly necessitated and which suggest a reluctance to throw away this last opportunity of marriage.

The chief disadvantage for Elizabeth of remaining single was, of course, that she left herself without an obvious successor. Indeed, it left, as the strongest claimant by descent, Mary Stuart, whom English Protestants in council, parliament and country would have none of, as they showed clearly during Elizabeth's

illness in 1562. Moreover, Mary Stuart was as natural a focus of Catholic hopes under Elizabeth as Elizabeth had been of Protestant hopes under Mary Tudor. To recognize her claim would have been for Elizabeth, in her own words, to spread a winding sheet before her eyes. Nor did the danger disappear when Mary was driven from Scotland (1568). Her mere presence in England, half guest, half prisoner, did much to provoke the revolt of the north in 1569 and to encourage a long series of assassination plots afterward. Only Elizabeth's distaste for executing an anointed queen kept Mary alive until 1587; if council and parliament could have had their way, she would have gone to the block with the duke of Norfolk in 1572. By 1587, however, a more acceptable successor had appeared in Mary's Protestant son, James VI of Scotland, whose accession to the English throne would solve a problem that had for long baffled statesmen on both sides of the border. Elizabeth certainly ran great risks by leaving the succession in doubt up to her dying day, but the risks were well judged. She avoided the challenge to her own supremacy which a husband, in the circumstances of the times, might have brought, and her reward—the peaceful and permanent union of the English and Scottish crowns—was great.

Religious Questions.—In ecclesiastical affairs, too, Elizabeth's chief concern was to maintain her own supremacy. As Anne Boleyn's daughter, she could have little affection for the Roman jurisdiction which made her a bastard; as queen, she would not allow her clergy to acknowledge anyone but herself as their supreme governor. Nor did she find Calvinism, with its emphasis upon an independent self-governing church, any more to her taste. Circumstances at her accession—the Marian bishops' determination to stand by the papal authority and the anti-Roman zeal of her first house of commons—drove her to rely from the start upon clergy of Protestant leanings and to accept a religious settlement which, broadly, re-enacted the Protestant measures of 1552. Yet she allowed neither clergy nor parliament to dictate how that settlement should be enforced and interpreted. While believing episcopacy to be the best vehicle for the royal supremacy, she had little reverence for bishops—and none at all for their wives. Yet, as Archbishop Edmund Grindal found, she was no more anxious to see Protestant enthusiasm encouraged among the lower clergy than she was to see the Acts of Supremacy and Uniformity too rigorously enforced against the Romanists. Undogmatic and unfanatical herself, she was probably sincere in her wish not to make windows into men's souls. However, it was chiefly a shrewd appreciation of the political dangers of religious dissension which made her reluctant to goad Catholics to desperation and eager to damp down Puritan ardour.

Relations with Rome were not finally broken off until 1561 and, largely because of Spanish diplomatic support, the popes were kept from extreme measures until 1570. The bull excommunicating Elizabeth then came too late to encourage the northern rebels in the decisive crisis of the reign. For more than 11 crucial years English Catholics had received no clear and compelling lead from the papacy. Meanwhile, the old priests were dying off; there were no new priests to replace them and instruct the new generation in the ancient faith; and Roman Catholicism in England seemed quietly dying of old age. Its demise was, of course, to be averted by the missionary work of the Jesuits and the priests trained overseas in William Cardinal Allen's new seminaries. Indeed, by the 1580s the Romanist revival became a threat not merely (like the Puritan movement) to the queen's supremacy, but to her very life. Although most of the missionaries carefully observed their instructions not to meddle with politics, the revival of Catholic fervour which they provoked stimulated others to less spiritual action. After all, the pope had absolved Elizabeth's subjects from their allegiance and Allen himself denounced her as a heretical usurper. Thus between 1581 and 1587 a long string of plots against her life found their inspiration in religion and their hope in Mary Stuart. The government had to strike, not only at the plotters themselves, but also at the missionaries who, more or less innocently, stirred such murderous ardour in their less scrupulous supporters. Yet the new penal laws, whose severity seemed so inadequate to the commons, were often tempered in their enforcement, as the earlier laws had been. It was perhaps little consolation to those who were executed to know that they were condemned as traitors and not as Catholics, but at least appreciably fewer were executed in the last 28 years of Elizabeth's reign than were burned in the last 3½ years of Mary Tudor's. The ordinary law-abiding Catholic went in little danger of his life and even the crippling recusancy fines were not invariably exacted in all places and at all times. Nevertheless the persecution and, after Mary Stuart's execution, the increasing identification of the Catholic plotters with the national enemy, Spain, were sufficient to keep the Romanist revival within bounds. By the end of the reign it was clear that the great majority of English Roman Catholics were willing enough to be loyal subjects of their Protestant sovereign if only a way could be found of saving their spiritual allegiance to the pope. If the aging Elizabeth failed to profit fully by this opportunity, at least she had done much to draw the fangs of Romanism as a political force; and that with a good deal less brutality than her parliaments wished or than her sister had used so vainly against the Protestants.

The more unpopular task of restraining the Puritans was left for as long as possible to the bishops. By the early 1570s, however, some of the Puritans had turned to attack episcopacy and to demand a self-governing presbyterian organization for the church. Furthermore, some of them sought to introduce into the house of commons bills embodying their schemes. They thus threatened the royal supremacy in a double way. Elizabeth knew as well as James VI that monarchy and presbytery agreed no better than God and the Devil. Already she had had to warn the commons against meddling in matters of state beyond their understanding, such as the question of her marriage on which they had petitioned in 1559 and 1563 and again more urgently in 1566. She could not let Puritan members now introduce bills and so seize the initiative in legislation from the crown and its councilors. It was upon this point that most of her differences with the commons arose, the Puritans maintaining that parliament's privilege of free speech allowed them to bring all matters into debate; the queen limiting that privilege, in matters of state, to free discussion of government bills. The majority accepted her ruling and the Puritans' parliamentary campaign produced more friction than fruit. She met with equal firmness their efforts to challenge her supremacy from inside the church. For a time the resistance of Archbishop Edmund Grindal balked her desire to suppress the local meetings of clergy in "classes" for mutual self-improvement, which were being used to spread Puritan ideas and presbyterian organization in the Anglican Church. She showed her disapproval by suspending Grindal, but it was not until she had found in his successor, John Whitgift, an archbishop after her own heart that she was able to make her wishes fully felt. Whitgift's measures, so disliked by Burghley, had her entire approval and they seem to have scotched the presbyterian danger. Yet here again it was only the movement's political implications that bothered the queen. She never loved Puritans but she did not persecute them, and was quite willing to employ them, when their activities did not seem to her politically dangerous.

Foreign Policy and Wars.—Another field over which Elizabeth kept close personal control was foreign policy. Her principal secretary—first William Cecil, then Sir Francis Walsingham and lastly Robert Cecil—might conduct the routine diplomatic correspondence. One or two other councilors—and on very rare occasions even the privy council as a body—might be called in to assist and advise. But the queen kept decisions in her own hands and took care to keep herself well informed. Occasionally a secretary might encourage agents abroad to send reports favourable to his own policy and might also edit his news by taking advantage of her well-known aversion to lengthy letters. However, she was too shrewd and her sources of information were too varied for deception to be easy. Moreover, her knowledge of languages enabled her to deal directly with foreign envoys. In foreign as in religious affairs, the ultimate control and responsibility rested with the queen herself. Her dislike of war and of needless expense rivaled that of Henry VII and was the despair of those who wished to see her the champion of Protestantism against the powers of the Counter-Reformation. Her concern, however, was England, not

Protestantism. She did not intend to stimulate religious dissensions within her realm by pursuing abroad the vigorous Protestant policy which she so carefully avoided at home. She knew that she would hardly draw the unwarlike Lutheran states to common action and she had learned from the French Huguenots in 1562–64 to mistrust the reliability, as well as the strength, of rebel Calvinists as allies. She understood, too, the disparity in wealth and manpower between England and the great Catholic powers of the continent. While ready to use foreign Protestants as "pressure groups," Elizabeth wisely chose to rely first upon the friendship, or the mutual jealousies, of their lawful rulers. She appreciated, as Walsingham never did, the deep-seated dynastic and national antagonisms which underlay those rulers' common Catholicism.

Thus, although in 1559–60 she openly allied with the Protestant rebels of Scotland, she preferred, once Mary had come home and Franco-Scottish relations had cooled, to seek an agreement with Mary and to use the Protestant lords as a lever rather than as the foundation of her Scottish policy. In her relations with France, at least after 1564, she likewise founded her policy upon an understanding with the French kings, while using the Huguenots as a counterpoise to the Catholic and increasingly pro-Spanish influence of the Guise faction. And just as in 1559–60 she had relied upon Spanish jealousy of France, so after 1568 she relied upon French jealousy of Spain—especially of Spanish military power in the Low Countries—to give pause to Philip's anger at her intermittent aid to the Dutch rebels and at the ceaseless depredations of the English seamen. She was thus able to postpone open hostilities for about 17 years; to give English maritime enterprise its head; to keep Dutch resistance alive; to neutralize the alarming military power of Spain just across the narrow seas in the Netherlands; and in large measure to control the no less alarming ambitions of her French ally in those regions. That in the 1580s such methods did not bring a peaceful solution of Anglo-Spanish differences, as they had done in the quarrel of 1568–73, does not altogether condemn her policy. She could hardly foresee the chapter of accidents which enabled Philip to turn from defense to aggression—the childlessness of Henry III of France; the early death of his brother Francis, duke of Alençon (from 1576 duke of Anjou), in 1584, which left the Huguenot Henry of Navarre heir to the French throne; the resulting revolt of the Holy Catholic League and collapse of the French monarchy; the extinction of the Portuguese ruling line; Spain's conquest of Portugal and Philip's consequent acquisition of the nucleus of an ocean-going navy which at last began to make an invasion of England practicable strategy. All these things had to happen before Elizabeth was burned by the fires with which she had so long and so skilfully played.

In the Spanish war which darkened the last 17 years of her reign, Elizabeth showed a similar cautious grasp of the essentials of strategy. She also showed herself as an inspiring leader in the moments of crisis, notably in her speech to her troops at Tilbury on Aug. 8, 1588, when, although the Armada was already beaten, an invasion by Parma's army from the Netherlands was expected hourly. "My loving people," she said, "we have been persuaded by some that are careful of our safety, to take heed how we commit ourselves to armed multitudes, for fear of treachery. But I do assure you, I do not desire to live to distrust my faithful and loving people. Let tyrants fear. I have always so behaved myself that, under God, I have placed my chiefest strength and safeguard in the loyal hearts and good will of my subjects. . . . I know I have the body of a weak and feeble woman, but I have the heart and stomach of a king, and of a king of England too, and think foul scorn that Parma or Spain, or any prince of Europe should dare to invade the borders of my realm; to which, rather than any dishonour shall grow by me, I myself will take up arms, I myself will be your general, judge, and rewarder of every one of your virtues in the field." It is true that she did not carry quite this aggressive confidence into her general conduct of the war. She lacked the money and the forces for ambitious military enterprises. She was somewhat skeptical of her sailors' belief that they could bring Spain to terms and guard the British Isles from invasion by naval action alone, and the course of the war tended to strengthen her doubts. Her first care therefore remained, in war as in peace, to keep the "invasion ports" of France and the Netherlands out of the hands of Spain and Spain's allies. Yet she was by no means wholly lacking in the offensive spirit. When her means, and the situation in France and the Netherlands, allowed, as in 1589 or 1595–97, she was ready enough to attack. Then Elizabeth sometimes saw more clearly than her martial advisers where the key to success lay.

Last Years.—The burdens of this long war, however, produced strains at home which heralded a new and more difficult era for the English monarchy. The aging queen could not win from the younger generation quite that admiration which in her youth she had inspired in their fathers. The favours with which she sought to sweeten the paths of obedience for them only whetted their impatience. In 1601 the most notable and most favoured of them all, the young and popular earl of Essex, carried impatience to the length of open rebellion. Only a handful of malcontents were prepared to go that far with him, but his execution cast a gloom over the nation as well as over the queen's last years. Bad trade and poor harvests were already causing misery among the poorer people, while gentry and townsmen groaned at the crown's ever-increasing calls upon their pockets and their services. In particular, the government's desperate search for revenues for the costly Irish campaigns of 1598–1603 bred a dangerous spirit of criticism which found its expression in the great debate on monopolies in the 1601 parliament. Elizabeth's "golden speech," by which she then dissolved the commons' criticism into a renewed outburst of loyal enthusiasm, showed that she still retained her command of the arts of popularity. "Though God hath raised me high," she told them, "yet this I account the glory of my crown, that I have reigned with your loves. . . . That my grants shall be made grievances to my people, and oppressions be privileged under colour of our patents, our princely dignity shall not suffer. When I heard it, I could give no rest unto my thoughts until I had reformed it. . . . It is not my desire to live or reign longer than my life and reign shall be for your good. And though you have had, and may have, many mightier and wiser princes sitting in this seat, yet you never had, nor shall have any that will love you better." She retained, too, a remarkable health and vitality, but her day was almost over and her work was done. Spain and the Irish rebels were beaten; the independence of France and the Dutch was secure; the unity of England was assured; and James of Scotland was established (though not until her dying day acknowledged) as Elizabeth's almost inevitable successor. Her ministers, courtiers and subjects had been turning their thoughts, if not their loyalty, northward for some while before her death, on March 24, 1603, ended the Tudor line. *See* also references under "Elizabeth I" in the Index volume.

BIBLIOGRAPHY.—J. E. Neale, *Queen Elizabeth I* (reprinted 1952), *Elizabeth I and Her Parliaments,* 2 vol. (1953–57); M. Creighton, *Elizabeth* (1896); J. B. Black, *The Reign of Elizabeth, 1558–1603,* 2nd ed. (1959); A. L. Rowse, *The England of Elizabeth,* 2 vol. (1951–55).

(R. B. WM.)

ELIZABETH II (1926–), queen of the United Kingdom of Great Britain and Northern Ireland and of her other realms and territories, head of the Commonwealth, was born on April 21, 1926, in Bruton street, London, where her maternal grandparents, the earl and countess of Strathmore, then lived, and was christened Elizabeth Alexandra Mary. She was the elder daughter of the duke of York (later King George VI), the second son of King George V. The princess's education was supervised by her mother, who entrusted her daughters to a governess, Marion Crawford; the princess was also thoroughly grounded in history by C. H. K. Marten, afterward provost of Eton, and had instruction from visiting teachers in music and languages.

For the princess the outbreak of World War II meant parting from her mother and father. During the early months of the war she and her sister, Princess Margaret Rose, were at Balmoral and thereafter at Royal lodge, Windsor, with occasional sojourns at Windsor castle. Her first broadcast, from Windsor castle on Oct. 13, 1940, was addressed to children evacuated from their homes. Her first public appearance without her family was in the summer of 1944, at a meeting of the governors of the Queen Elizabeth Hospital for Children. In March 1945, in spite of opposition

from her family and the government, she trained as a driver in the Auxiliary Territorial service.

At the beginning of 1947 Princess Elizabeth went with the king and queen to South Africa. While they were there she reached 21 and on her birthday broadcast to the commonwealth. She spoke of the motto borne by many of her ancestors, "I serve," and promised to devote the whole of her life to "the service of our great imperial family." After her return from South Africa, the princess's betrothal was announced to her distant cousin Lieut. Philip Mountbatten, R.N., formerly Prince Philip of Greece. The marriage took place in Westminster abbey on Nov. 20, 1947. On the eve of the wedding the king conferred upon the bridegroom the rank of royal highness and the titles of duke of Edinburgh, earl of Merioneth and Baron Greenwich. A son, Prince Charles, was born on Nov. 14, 1948, at Buckingham palace. Princess Elizabeth took up residence in Clarence house, where Princess Anne was born on Aug. 15, 1950. In 1950 the princess passed some time in Malta where the duke, who had returned to active service in 1949, was stationed.

The health of King George VI gave rise to concern during the summer of 1951 and the princess represented him at the Trooping the Colour and during the official visits of the kings of Denmark and Norway. After the king's serious operation, she set out with her husband on Oct. 7 for an important and highly successful tour of Canada. With the duke she also visited the president of the United States, Harry S. Truman, in Washington, D.C. After Christmas in England, the princess and the duke of Edinburgh started, on Jan. 31, 1952, for a tour of Australia and New Zealand. At the Forest lodge in Sagana, a wedding present from the people of Kenya, news reached them of the sudden death of King George on Feb. 6.

Elizabeth, now queen, at once flew back to England and on Feb. 8 made her speech to the accession privy council with a moving reference to "this heavy task that has been laid upon me so early in my life." The first three months of her reign, the period of full mourning for her father, were passed in comparative seclusion. But in the summer, after she moved from Clarence house to Buckingham palace, she undertook the routine duties of the sovereign and carried out her first state opening of parliament on Nov. 4, 1952. She was crowned in Westminster abbey on June 2, 1953, her husband being the first person to do homage. In Nov. 1953 she and the duke began a six-month tour of New Zealand and Australia, visiting the West Indies en route and returning by Ceylon and Uganda. In 1955 they paid a state visit to Norway, the queen's first official visit to a European country since her coronation. During the summer she also visited Wales, where she was the first reigning monarch to sit in the royal stall at St. David's cathedral since Edward I, and the Isle of Man, where she presided over the Tynwald, the ancient parliament. In 1956 she and the duke made a three-week tour of Nigeria and paid a state visit to Sweden. In 1957, after state visits to Portugal, France and Denmark the queen and Prince Philip visited Canada and the United States.

In 1958 Elizabeth paid a three-day official visit to the Netherlands and, following a custom started by Queen Victoria, sailed in the royal yacht up the west coast to Scotland, discharging certain official duties on the way. On July 26 at the close of the Commonwealth games at Cardiff she announced the creation of Prince Charles as prince of Wales. In the summer of 1959 she went to North America to open, with the president of the United States, Dwight D. Eisenhower, the St. Lawrence seaway. She also visited Chicago where she was received with enthusiasm. On Aug. 7 it was announced that the queen was expecting a baby and that all her engagements, including an official visit to Ghana, would be canceled. Her son—the first child born to a reigning British sovereign for more than a century—was born on Feb. 19, 1960, and was christened Andrew in honour of Scotland and of Prince Philip's father. Her personal decision in 1960 to link the name of Mountbatten with that of Windsor for some of her descendants showed that she was not to be influenced by criticism from the popular newspapers or court traditionalists. (R. T. B. F.)

ELIZABETH (1837–1898), empress consort of Austria from 1854, was born in Munich on Dec. 24, 1837, the daughter of Maximilian Joseph, duke in Bavaria, and Louisa Wilhelmina of Bavaria. At Ischl in Aug. 1853 the Bavarian ducal family met Francis Joseph, emperor of Austria. He fell in love with the young Elizabeth, famed as the most beautiful princess in Europe, and the marriage took place in Vienna on April 24, 1854.

The empress showed a neurotic restlessness that may have been derived from her Wittelsbach ancestry. Her disregard for court convention offended Viennese society, and her predilection for Hungary, which she first visited in 1857, affronted German sentiment. She spent much time in travel or away from Vienna; *e.g.*, at Gödöllö, north of Pest, or at the Achilleion palace, built for her in Corfu. Her care for the wounded in the Seven Weeks' War, however, won her some popularity in Austria, and she was well-loved in Hungary for her charitable activities. She helped to smooth the way for the Austro-Hungarian *Ausgleich* of 1867 (on which Francis Joseph and she went to Hungary for coronation), but in general she avoided politics. The suicide of her only son, the crown prince Rudolf, in 1889, was a shock from which she never fully recovered.

In 1898 Elizabeth was staying at Geneva. At about mid-day on Sept. 10, as she was walking down to the lake intending to take the steamer on her way to Caux, she was attacked by an anarchist, Luigi Luccheni, who dealt her a wound from which she died within a few hours.

See K. Tschuppik, *The Empress Elizabeth of Austria* (1930); M. Paléologue, *Tragic Empress*, Eng. trans. (1950).

ELIZABETH (1596–1662), British princess who from 1619 was titular queen of Bohemia, was born at Falkland castle, Fifeshire, in Aug. 1596, the daughter of James VI of Scotland (later James I of Great Britain) and Anne of Denmark. In 1603 she followed her parents to England. She was placed under the guardianship of Lord Harington and his wife, lived mostly at Combe Abbey near Coventry and, it seems, received a very good education. In 1606 she came to the royal court, where her beauty and charm attracted much attention and where she soon became a favourite subject of the poets—though the famous verses by Sir Henry Wotton (*q.v.*) belong to a later period. Gustavus Adolphus of Sweden, Philip III of Spain and Frederick V, the elector Palatine, all sought her hand. Her mother actively favoured the Spanish match, but her father, hoping to strengthen his ties with the German Protestant rulers, chose Frederick. Despite Elizabeth's grief at her brother Henry's death (Nov. 1612), the wedding took place, as planned, in Feb. 1613, with great festivities: several of Shakespeare's plays were performed (among them *The Tempest*), and John Donne wrote an epithalamium. Leaving England in April, the couple reached Heidelberg in June; their first child, Frederick Henry, was born in Jan. 1614. For the next few years Elizabeth led a life of luxurious gaiety new to the Heidelberg court till in 1619 the Bohemians, in defiance of their Habsburg king Ferdinand, offered the crown of Bohemia to Frederick. In Nov. 1619 he was crowned king, but in Nov. 1620 the Bohemian forces were defeated by the Catholic league acting for Ferdinand (then Holy Roman emperor), and Elizabeth and Frederick had to flee. They ultimately found refuge at The Hague with Prince Maurice of Orange, and Elizabeth spent the ensuing 40 years in exile. She did not lose her good spirits though her life was not happy. In 1629 her eldest son died, and in 1632 her husband. When in 1648 her second son, Charles Louis, was restored to the Palatinate, he ignored her pleas that she should join him and in 1650 a pension from the house of Orange ceased. As for help from England, this stopped with the Civil War, especially after her brother, Charles I, was beheaded. In 1661 her nephew Charles II grudgingly allowed her to return to England. On Feb. 13, 1662, she died in Westminster.

Of Elizabeth's 12 children the most important were: Charles Louis (1617–80), elector Palatine; Elizabeth (1618–80), the pupil of Descartes, who dedicated his *Principia philosophiae* (1644) to her and with whom she maintained a long correspondence; Rupert (*q.v.*), distinguished in English history; and Sophia (*q.v.*), electress of Hanover, through whom the Hanoverians succeeded to the British throne.

Queen Elizabeth's letters were published in 1953, edited by L. M. Baker.

See C. Oman, *Elizabeth of Bohemia* (1938). (R. C. Jo.)

ELIZABETH (Woodville) (1437–1492), queen consort of Edward IV of England, was already a widow with two sons when Edward first met her; for her first husband, Sir John Grey of Groby, had been killed at the second battle of St. Albans in 1461. Her great beauty aroused Edward's desire, but as she refused to become his mistress, he married her on May 1, 1464. He knew that the match would be repugnant to the Yorkist nobility, since she was the daughter and widow of Lancastrians and her birth and connections were not of royal rank; hence the marriage was not made public until Michaelmas 1464. Her unpopularity was increased by her haughtiness and her grasping nature, which quickly procured a shower of royal favours for her two sons, her five brothers and her seven sisters. She was always a devoted wife, and as she bore Edward two sons and five daughters who survived infancy, the succession seemed secure. However, when Edward died unexpectedly, leaving a boy of 12 (Edward V) as his heir, the general dislike of the queen and her kinsfolk greatly facilitated Richard of Gloucester's seizure of the throne. Her weakness of character was revealed when she gave up to him her second son, Richard, although the first was already in his power, and when she later entrusted her daughters to him, although he had proclaimed them bastards. When Henry VII came to the throne and married her eldest daughter Elizabeth (Jan. 1486) he restored her dower lands; but in Feb. 1487 she exchanged them for an annuity and withdrew to the abbey of Bermondsey. She died there on June 7 or 8, 1492, and was buried beside her husband in St. George's chapel, Windsor.

See David MacGibbon, *Elizabeth Woodville, 1437–1492* (1938). (A. R. M.)

ELIZABETH (1900–), queen consort of George VI of Great Britain, was born at St. Paul's Waldenbury, Hertfordshire, on Aug. 4, 1900, the youngest daughter of Claude George Bowes-Lyon, 14th earl of Strathmore (d. 1944), whose family was descended from Robert the Bruce, king of Scotland, and Cecilia (d. 1938), daughter of the Rev. Charles W. Cavendish Bentinck, and was baptized Elizabeth Angela Marguerite. She was brought up principally at Glamis castle and her attachment to Scotland has remained a constant element in her life. She was betrothed to the duke of York in Jan. 1923, and they were married in Westminster abbey on April 26 of the same year. The royal family welcomed the marriage and her father-in-law in particular was devoted to her.

The duchess' first home was at 145 Piccadilly, and it was from there in 1927, while her elder daughter, Princess Elizabeth (b. April 21, 1926), was still a baby, that she set out with the duke on a great imperial tour of Australia. Her second daughter, Princess Margaret Rose, was born at Glamis on Aug. 21, 1930. Throughout her life as duchess of York Elizabeth sought to merge her personality in the public life of her husband, aiming at no power or popularity independent of his. In the family circle her influence, exerted always by persuasion rather than command, was strong and she personally controlled and directed the upbringing and education of her daughters. The home life of the duke and duchess was conspicuously happy; the duchess felt the sacrifice of domestic quiet when her husband unexpectedly became king in Dec. 1936. She was crowned with him in Westminster abbey on May 12, 1937.

Elizabeth recognized that it was possible for the queen consort to be less formal than the sovereign and, with her gift of putting people at ease, she was a great support to King George on all public occasions. During the last summer of his life (1951) when he was stricken with illness, she carried out the duties of the sovereign almost singlehanded.

Following the precedent initiated by Queen Mary she was present in Westminster abbey at the coronation of her husband's successor on June 2, 1953. Shortly afterward she visited Rhodesia to open the Rhodes Centenary exhibition. She went to Rhodesia again in 1957, and was installed as president of the University College of Rhodesia and Nyasaland. She went to Canada and the U.S. in 1954 and was given a rousing welcome in both countries. She also paid a highly successful visit to Australia in 1958. At home, she became chancellor of London university in 1955, identifying herself closely with its official life and its students. (R. T. B. F.)

ELIZABETH (1843–1916), queen consort of Rumania, best-known for her writings under the name of Carmen Sylva, was born on Dec. 29, 1843, at the Monrepos castle near Neuwied in the Prussian Rhineland, the daughter of Prince Hermann of Wied. On Nov. 15, 1869, she was married to Carol I (*q.v.*), who had been prince of Rumania from 1866 and was to be king from 1881. The only daughter of the marriage died in 1874, and Elizabeth never recovered fully from this loss, suffering continuous ill-health for the rest of her life. However, she took a keen interest in Rumania's cultural development, especially in Rumanian poetry and folklore. As Carmen Sylva she collected and published many Rumanian legends and also composed many literary works of her own, in prose and in verse, writing easily in English and French as well as in German and Rumanian. *Pensées d'une reine* (1882) and *The Bard of the Dimbovitza* (1891), an English version of Elena Vacarescu's collection of folk songs, are notable. Much loved by the Rumanians, she died in Bucharest on March 2, 1916.

ELIZABETH (Élisabeth de France; called Madame Élisabeth) (1764–1794), French princess notable for her courage and exemplary Christian virtues during the Revolution, was born at Versailles on May 3, 1764, the youngest daughter of the dauphin Louis (d. 1765) and Maria Josepha of Saxony. Whereas her aunt, and two of her brothers (the future Louis XVIII and Charles X) emigrated early enough, Madame Élisabeth refused to leave her brother the king, Louis XVI, and Marie Antoinette at grips with the Revolution. She was imprisoned with them in the Temple after the suspension of the monarchy on Aug. 10, 1792, and shared all the hardships that this involved. Sentenced to death by the Revolutionary tribunal, she was guillotined on May 10, 1794. The fortitude and patience with which she bore her trials won lasting respect, especially in Catholic and royalist circles.

The *Mémoires de Madame Élisabeth*, edited by F. de Barghon Fort-Rion (1858), are of doubtful authenticity; and the *Correspondance de Madame Élisabeth de France*, edited by F. Feuillet de Conches (1868), should be consulted only with caution. There are English versions of some of the letters by K. P. Wormeley, *The Life and Letters of Madame Élisabeth de France* (1902).

Bibliography.—*Louis XVI, Marie-Antoinette et Madame Élisabeth*, ed. by F. Feuillet de Conches, 6 vol. (1864–73); G. du Fresne de Beaucourt, *Étude sur Madame Élisabeth* (1864); A. de Beauchesne, *Vie de Madame Élisabeth* (1869); the Hon. Mrs. Maxwell-Scott, *Madame Élisabeth of France* (1908); E. M. du L., *Une grande chrétienne: Madame Élisabeth* (1932); Y. de La Vergne, *Madame Élisabeth de France* (1936); M. Trouncer, *Madame Élisabeth* (1955). (Je. Le.)

ELIZABETH, a town of South Australia, founded in 1955 and named after Queen Elizabeth II, is situated on the Adelaide plains 16 mi. N. of Adelaide by road or rail. The population, 23,317 in 1961, was expected to reach 50,000 within ten years. The town was planned to cover an area of 10 sq.mi. of previously open farm lands. It was designed as a series of residential neighbourhoods, each with its own schools and shops, separated by parks and gardens and grouped around the main shopping centre and public buildings. In the early 1960s residents found employment mainly at the Weapons Research establishment, Salisbury, and in Adelaide, but factories had been established for the manufacture of motor cars, sewing machines, asbestos piping, clothing, electrical switchgear, bedding, pyrotechnics, industrial gases and steel springs. (J. Mi.)

ELIZABETH, the seat of Union county, northeastern New Jersey, U.S., fronting on Staten Island sound and Newark bay, 14 mi. from New York city, and a part of the New York-northeastern New Jersey metropolitan area. The city is highly industrialized, with many national corporations maintaining plants there. Products include sewing machines, bedding, clothing, paper supplies, printing presses, electric motors, wire goods, copper, chemicals and petroleum. The city is served by several major highways, Newark airport, Port Elizabeth harbour and trunk lines of several major railroads. Connecting Elizabeth with Staten Island is a

railroad bridge with a vertical lift span of 558 ft.

Elizabeth came into being as a settlement in 1664 through the purchase of a tract of land from the Delaware Indians. The site shortly became the meeting place (1668–82) for the first colonial assembly. Named in honour of Lady Elizabeth, wife of Sir George Carteret, one of the colony's first proprietors, it was first known as Elizabethtown. During the American Revolution the city was the scene of four military engagements, the chief one being the battle of Elizabethtown in 1780. Many historic colonial buildings are still standing and include the Nathaniel Bonnell house (erected 1682), the Belcher mansion (residence of Gov. Jonathan Belcher in the 1750s) and Boxwood hall. In 1746 the College of New Jersey, now Princeton university (*q.v.*), was established in Elizabeth. In the 1870s John Philip Holland (*q.v.*) built the first successful submarine there. Pop. (1960) 107,698. (W. L. Ca.)

ELIZABETHAN STYLE, in architecture, a term loosely applied to the early Renaissance work in England up to the end of the 16th century. The term Tudor period (*q.v.*) is generally limited to the earlier Elizabethan work, in which Gothic elements are still dominant; "Jacobean style" (*q.v.*) is used of the later portion of the period, in which Renaissance ideas had finally conquered. These three terms are overlapping and there is much variety of usage in the meanings given to them by different authorities. In general, the Elizabethan style is that which resulted from the impingement upon the vivid Perpendicular Gothic tradition, first of Italian Renaissance artists, such as Pietro Torregiano, who designed the tomb of Henry VII, and later of a great number of Flemish craftsmen who settled in England. Thus, such buildings as Hampton Court palace, Sutton place in Surrey, Hatfield house and Audley End, varying as they do from the Gothic of Hampton Court to the Renaissance of Hatfield, are all spoken of as being in the Elizabethan style.

ELK, a term of unusually confusing extension, originally applied in Europe to *Alces alces,* the large hoofed mammal known in North America as the moose (*q.v.*). The name is also loosely used to indicate several large deer of the genus *Cervus,* notably the red deer of Europe, the Kashmir stag and the Himalayan shou as well as the North American deer that are more correctly called wapiti (*q.v.*). *See* also DEER. (K. R. Kn.)

ELKHART, a city of Elkhart county, Ind., U.S., on the St. Joseph river 100 mi. E. of Chicago and 15 mi. E. of South Bend. Major industries include an electronically controlled railroad classification yard; pharmaceutical products; several manufacturers of band instruments (it has been called the "band capital of the world"); and a number of manufacturers of mobile homes. Other industries produce power, aviation and automobile equipment, castings, power presses, pumps, tools, machinery, clay-coated cardboard, office furniture, sheet-metal products and television parts. Elkhart was platted in 1832, incorporated in 1858 and chartered in 1875. The name is from the translation of the Indian name for an island at the confluence of the St. Joseph and Elkhart rivers, so called because its shape was thought to resemble that of the heart of the elk. The site was popular with the Indians and was the junction of several of their trails.

For comparative population figures *see* table in INDIANA: *Population.* (J. J. D.)

ELLENBOROUGH, EDWARD LAW, 1ST BARON (1750–1818), lord chief justice of England, who was leading counsel for Warren Hastings (*q.v.*) during the latter's impeachment trial, was born at Great Salkeld, Cumberland, on Nov. 16, 1750, where his father, afterward bishop of Carlisle, was rector. He was educated at the Charterhouse and at Peterhouse, Cambridge. He was admitted at Lincoln's Inn and was called to the bar in 1780, after spending five years as a special pleader. He had a lucrative practice at the bar, primarily in mercantile matters. His most famous case, however, was that of Hastings. At the end of this case, which lasted from 1788 until 1795, he was one of the leading figures at the bar. Dismayed at the violence of the French Revolution, he abandoned the Whig party and was appointed attorney general by the Tory government in 1801 and lord chief justice in 1802 on the death of Lord Kenyon, with the title of Lord Ellenborough. In 1806 he was subjected to great criticism for combining political and judicial office when he accepted a seat in the cabinet while retaining the chief justiceship. He had a reputation for being harsh and overbearing with counsel and in political trials for showing unmistakable bias. He was a forceful speaker in the house of lords but often overstepped the bounds of parliamentary licence. He resigned his judicial office in Nov. 1818, and died Dec. 13, 1818. (A. Dm.)

ELLENBOROUGH, EDWARD LAW, EARL OF (1790–1871), English statesman, remembered for his work in India, was born on Sept. 8, 1790, the eldest son of the 1st Baron Ellenborough. Educated at Eton and St. John's college, Cambridge, he entered the house of commons in 1813, and the house of lords on his father's death in 1818 but not until 1828 did he attain office, first as lord privy seal and then as president of the board of control for India (1828–30).

Masterful and energetic, Ellenborough sought by improving communications to bring Indian affairs under closer governmental control, and pressed for a steamship service to Alexandria, Egy., and through the Red sea to Bombay. He enlarged the staff of India house and abolished the cumbersome correspondence system of general letters, substituting annotated letters each concerned with a single subject. He strove to unleash India's economic energies, by lightening the tax burden. He ordered the governor general, Lord William Bentinck, to economize, proposed to add a financial expert to his council and supported Bentinck's use of western-educated Indians in the administration. He saw that a wealthier India could be both supplier and market for British industry, and that the Indus river, if opened to steam navigation, might carry British trade and influence into central Asia and so curb Russian expansion, which he feared unduly.

After brief spells at the board of control in 1834 and 1841 Ellenborough was appointed governor general of India in 1841. He had opposed since 1839 the costly intervention in Afghanistan, and news of disaster at Kabul and Ghazni led him to continue the plans of his predecessor, Lord Auckland, for total withdrawal. He granted his distant generals, Sir William Nott and Sir George Pollock, permission to strike an avenging blow if military conditions permitted. When they advanced victoriously to Kabul he was lavish in public praise and on the heroes' return he provided a theatrical reception, ridiculed in England but not without effect in India. He decided to make the Indus his frontier, retaining only those strong-points upon it which ensured free navigation. However, rash opportunism led him to acquiesce in actions of Sir Charles James Napier (*q.v.*), his governor in Sind, which drove the amirs into war and defeat (1843). Ellenborough's grasping haste to exploit victory by the annexation of Sind antagonized the directors of the East India company and alienated British public opinion.

Ellenborough then pursued his plans to promote trade by freeing it of tolls and duties throughout Sind, Bahawalpur, the North-Western Provinces, Madras and Bombay. But a war with Gwalior in Dec. 1843, designed to keep its large army out of hostile hands, frustrated these plans and the directors, exasperated by Ellenborough's arrogant self-will, resolved, in April 1844, to recall him.

Ellenborough served under Sir Robert Peel as first lord of the admiralty in 1846 and under Lord Derby at the board of control in 1858. There he drafted the new plan for the government of India which the Indian Mutiny had rendered necessary, but by making public a caustic dispatch censuring Lord Canning's Oudh proclamation, which Ellenborough thought betrayed a desire for indiscriminate vengeance, he roused such opposition that he chose to resign. He never held office again and died in Gloucestershire on Dec. 22, 1871.

BIBLIOGRAPHY.—A. H. Imlah, *Lord Ellenborough* (1939). *See* also Lord Colchester (ed.), *History of the Indian Administration of Lord Ellenborough* (1874); Sir Algernon Law (ed.), *India Under Lord Ellenborough* (1926). (J. B. Ha.)

ELLESMERE, FRANCIS EGERTON, 1ST EARL OF (1800–1857), English nobleman, was born Lord Francis Leveson-Gower on Jan. 1, 1800, in London; he was the second son of the 1st duke of Sutherland. When his father died in 1833 he took the name of Egerton, and inherited a vast fortune from his great-uncle, the 3rd duke of Bridgewater. He sat in parliament as a

Liberal Conservative, a follower first of George Canning and then of Sir Robert Peel, from 1822 until, in 1846, he was created earl of Ellesmere. He was not a strong party man and held no higher office than that of chief secretary for Ireland (1828–30), but was a trusted negotiator behind the scenes.

He published essays, travel sketches, poems, and translations from French and German, including a version of Goethe's *Faust*. He was a munificent patron of artists, and opened a splendid collection of paintings to the public. He lived in the Lancashire colliery district of Worsley, where he spent largely to improve the conditions of working people. He was made a knight of the Garter in 1855 and died in London on Feb. 18, 1857. (M. R. D. F.)

ELLESMERE, THOMAS EGERTON, Baron (Lord Chancellor): *see* Brackley, Thomas Egerton.

ELLESMERE ISLAND, largest of the Queen Elizabeth Islands, Canada, is the northernmost portion of North America, reaching a latitude of 83° 07′ N. at Cape Columbia. Discovered in 1616 by William Baffin, it was named in 1852 after Francis Egerton, first earl of Ellesmere. Lying generally off the northwest coast of Greenland, the island has an area of 82,119 sq.mi.; it consists of a northeastern plateau, a southwestern plateau, southeastern highlands, northwestern mountains and an extensive central mountain belt. Glaciers cover large portions of the last three regions, and fiords cut the island naturally into four areas: Grant land in the north, Grinnell land and Sverdrup land in the centre and Lincoln land in the south. Elevations reach 9,500 ft. in the northwest. Sparse vegetation, found chiefly on the western coast lands and along inland passes, supports musk oxen and caribou. Small settlements are found at Alert on the northern coast, Eureka on the west, Craig Harbour in the south and Alexandra fiord in the east. There are undeveloped deposits of coal and gypsum. (H. R. Tn.)

ELLESMERE PORT, a municipal borough (1955) of Cheshire, Eng., 10 mi. S.S.E. of Liverpool. Pop. (1961) 44,714. It is an expanding industrial town with important oil refineries on the west bank of the Manchester Ship canal with docks about 3 mi. from the entrance to the canal from the Mersey estuary at Eastham locks. It is also at the junction of the Shropshire Union barge canal with the Ship canal, the former giving access via Wolverhampton to the midlands and south, and on the Birkenhead-Manchester railway line.

The port handles grain, oil and coal and has a 2,250-ft. deep-water wharf.

ELLICE ISLANDS: *see* Gilbert and Ellice Islands.

ELLIOT, JANE (also called Jean) (1727–1805), Scottish poet known for her version of the ballad "Flowers of the Forest," was born at Teviotdale, Roxburghshire, in 1727. Her lines beginning "I've heard them liltin' at our ewe-milkin'" and set to the traditional tune of "Flowers of the Forest" were probably composed in 1756. An earlier version written by Alicia Cockburn (*q.v.*) was unknown to Jane, who incorporated in her version the first line and the refrain of the traditional lament for those fallen at Flodden.

She died at Teviotdale on March 29, 1805.

ELLIOTSON, JOHN (1791–1868), English physician and advocate of the use of hypnosis (mesmerism) in medical practice, was one of the first teachers in London to appreciate the value of clinical lecturing and one of the earliest among British physicians to urge the employment of the stethoscope. He was born at Southwark, London, on Oct. 29, 1791. He studied medicine at Edinburgh and Cambridge and subsequently in London at St. Thomas's and Guy's hospitals. In 1831 he was elected professor of the principles and practice of physic in London university, and in 1834 he became physician to University college hospital. His interest in hypnosis eventually brought him into collision with the medical committee of the hospital, and he resigned his offices in 1838. In 1849 he founded a mesmeric hospital. He died in London on July 29, 1868.

Elliotson contributed numerous papers to the *Transactions* of the Medico-Chirurgical society, of which he was at one time president; he was founder and president of the Phrenological society. Thackeray's *Pendennis* was dedicated to him.

ELLIOTT, CHARLES LORING (1812–1868), U.S. portrait painter, was born at Scipio, N.Y., Oct. 12, 1812. Early in his childhood his parents moved to Syracuse, where his father established himself as an architect. As a boy, Charles Elliott showed skill in drawing and later made successful architectural drawings for his father. For a time he worked in the studio of John Trumbull, who encouraged him to become an architect but, realizing the boy's ambition, allowed him access to the casts in the American Academy and instructed him in his work.

Elliott finally left the studio and studied under John Quidor. His illustrations for Washington Irving's *Knickerbocker's History* and J. K. Paulding's *Dutchman's Fireside* won critical praise. For several years he painted portraits in central and western New York; and finally, in 1845, opened a studio in New York city, where he became extremely successful.

It is believed that Elliott painted 700 portraits. He died in Albany, N.Y., Aug. 25, 1868.

ELLIOTT, EBENEZER (1781–1849), English poet best known for his *Corn Law Rhymes*, of which "When wilt thou save the people, O God of mercy, when?" (still sung as a hymn) is typical. He was born at Masborough, Yorkshire, on March 17, 1781. When he was 16 he began work at his father's iron foundry. He described himself as self-educated, and attributed his poetical development to long country walks and a collection of books left to his father.

His first poem, *The Vernal Walk* (1801), was followed by other works in the romantic tradition, none of them successful. His father's financial losses and the final failure of the foundry he attributed to the effects of the corn laws. However, he later set up a prosperous business in Sheffield. His hatred of the corn laws made him an active Chartist and produced a complete change in the character of his verse: the *Corn Law Rhymes* (1831) are notable for their vigorous denunciation of injustice and direct, vivid descriptions of life among the poor. Later publications were *The Splendid Village, Corn Law Rhymes and Other Poems* (3 vol., 1833–35) and *More Verse and Prose by the Corn-Law Rhymer* (2 vol., 1850). Elliott died at Great Haughton, Yorkshire, on Dec. 1, 1849.

Bibliography.—T. Carlyle, "Corn Law Rhymes," *Edinburgh Review*, lv (1832); J. Watkins, *Life, Poetry and Letters of Ebenezer Elliott* (1850); J. W. King, *Ebenezer Elliott, a Sketch* (1854).

ELLIOTT, HOWARD (1860–1928), U.S. railway executive, was born in New York city, Dec. 6, 1860. Educated as a civil engineer at the Lawrence scientific school, Harvard university, he started his railroad service with the Chicago, Burlington and Quincy railroad in 1881. In 1903 he became president of the Northern Pacific railway and in 1913 president and then board chairman of the New York, New Haven and Hartford railroad. In 1920 he resumed the presidency of the Northern Pacific as well as becoming chairman of the executive committee and of the board of directors.

Elliott was three times elected an overseer of Harvard university and from 1925 to 1927 served as president of the board of overseers of Harvard.

Elliott died at Dennis, Mass., July 8, 1928. (W. H. D.)

ELLIPSE. A conic section (*q.v.*) is a curve determined by the points of intersection of a plane and a right circular cone. If it is a closed curve, that is, if the plane cuts only one nappe (or sheet) of the cone, it is an ellipse, a curve of second degree (order and class). (*See* fig. 1.) These curves were known to the Greek geometers Menaechmus (*c.* 350 B.C.) and Apollonius (*c.* 220 B.C.), who described many of their properties.

Considered as a plane curve, an ellipse could be defined in other ways. For example, (1) the locus of a point which moves so that the sum of its distances from two fixed points is constant (greater than the distance between the two points) is an ellipse; (2) the locus of a point which moves so that the ratio of its distance from a fixed point to a given line is a constant less than one is an ellipse; (3) if the ordinates of a circle $x^2 + y^2 = a^2$ are all shrunk (or stretched) in the same ratio b/a, the resulting curve is an ellipse whose equation is $x^2/a^2 + y^2/b^2 = 1$.

Associated with an ellipse are certain special points and lines.

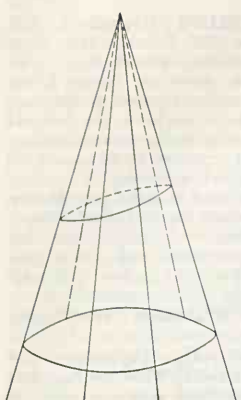


FIG. 1

Let the equation of an ellipse be $x^2/a^2 + y^2/b^2 = 1$, and its graph be the curve $ABPA'CB'$. (*See* fig. 2.) The vertical line BB' is called the minor axis and is of length $2b$. The circle about it as a diameter is called the minor circle. Similarly, the line segment AA' is called the major axis and has length $2a$; and the circle which has this line segment as a diameter is called the major circle. The circle of radius a with B as centre cuts the major axis AA' at F and F', points of special importance (called foci). Now $OF = OF' = \sqrt{a^2 - b^2}$. If one lets $OF = ae$, then $e = \sqrt{1 - b^2/a^2}$ and $1 - e^2 = b^2/a^2$. This number e is called the eccentricity and is less than one for the ellipse, equal to one for the parabola and greater than one for the hyperbola.

If P' and P are any two corresponding points on the major circle and the ellipse, and S the common foot of the ordinates $P'S$, PS, then the angle SOP' is termed the eccentric angle of P. Now $FP^2 = r^2 = (ae - x)^2 + y^2 = (ae - x^2) + (1 - e^2)\ (a^2 - x^2) = (a - ex)^2$, and $r = a - ex$. Similarly, $F'P = r' = a + ex$. Hence $FP + F'P = 2a$, and the first definition of an ellipse given in the last paragraph follows from this property. If we think of the point F (or F') as a pole, associated with this point will be a polar line QR, the locus of points of intersection of pairs of tangents at the ends of chords which pass through this point. The polar lines so determined are called the directrices of the ellipse. The second definition given in the preceding paragraph follows when the point F is used as the fixed point, the directrix as the given line, and the eccentricity as the prescribed ratio.

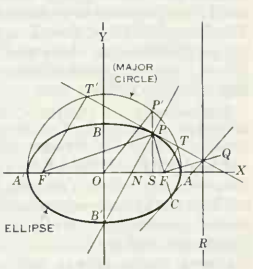


FIG. 2.—ELLIPSE WITH ASSOCIATED POINTS AND LINES

A few special properties of ellipses are here stated; others are mentioned in standard texts on plane analytic geometry.

For an ellipse, the tangent and normal at any point P (fig. 2) bisect the angles between the focal radii FP, $F'P$; hence, any ray from F is reflected at P through F' and then back to F, and all rays from either focus converge by reflection at the other, a fact which is of significance in the study of light, sound, etc.

Any perpendicular from a focus on a tangent such as (FT) meets it on the major circle (at T); the product of the lengths of two such perpendiculars (from F, F') is b^2 $(FT.F'T' = b^2)$, a constant. From any outside point Q two tangents QP, QC, may be drawn to an ellipse, and the angles these tangents make with focal radii to Q will be equal; *i.e.*, $< PQF = < CQF$.

The area of an ellipse is πab. The problem of finding the arc length involves the use of elliptic integrals, elliptic functions and doubly periodic functions in general. (*See* ELLIPTIC FUNCTIONS.)

An ellipse may be drawn by fastening the ends of a chord $2a$ units in length at F and F', and then passing a pencil point along the chord held taut by the point. Alternatively any number of points may be found as follows: Draw concentric major and minor circles (radii a and b) about O; and any radius cutting them at P and Q. (*See* fig. 3.) Through P draw a vertical line, and then through Q a horizontal line meeting at M, a point on the ellipse. To show this, one notes that the x co-ordinate of the point M is the same as that for P, a point on the major circle, while the y co-ordinates have been shortened in the ratio b/a (fig. 3).

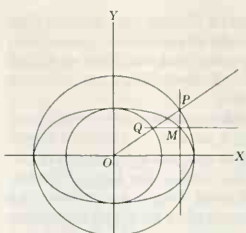


FIG. 3.—CONSTRUCTIONS ON AN ELLIPSE

Certain points and lines associated with an ellipse may be determined geometrically. To find the axes of an ellipse, bisect two pairs of parallel chords; the bisectors will be diameters meeting in the centre O. With O as centre, draw any circle cutting the ellipse in four points. These four points are symmetrical in pairs with respect to the axes, and determine a rectangle. The lines through O parallel to the sides of this rectangle are the axes. The foci are F and F' on the major axis AA' $(2a)$, each at a distance of a from the ends B, B' of the minor axis.

The ellipse occurs in astronomy as the path of a body moving about another body while obeying the Newtonian law of motions (the attracting force is proportional to the product of the masses of the two bodies, and inversely proportional to the square of the distance between them).

From the projective standpoint, ellipses, hyperbolas and parabolas are equivalent curves; *i.e.*, can be transformed into one another by projective transformations. For these curves, it is known that the opposite sides of any simple hexagon inscribed in the curve meet in points that are collinear, as in fig. 4 (Pascal's theorem); and dually, that the lines joining opposite vertices of any simple hexagon circumscribed about a conic meet in a point (Brianchon's theorem). Other projective properties of conic sections (including ellipses) may be found in standard treatises on projective geometry.

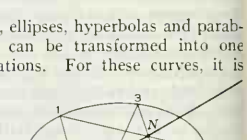
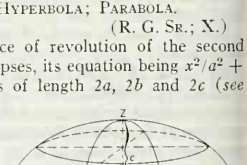


FIG. 4.—ELLIPSE WITH INSCRIBED HEXAGON

See also ELLIPSOID; ANALYTIC GEOMETRY; PROJECTIVE GEOMETRY; CONIC SECTION; CIRCLE; HYPERBOLA; PARABOLA.

(R. G. SR.; X.)

ELLIPSOID, a closed surface of revolution of the second degree. Its plane sections are ellipses, its equation being $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$. The segments of length $2a$, $2b$ and $2c$ (*see* figure) are called the principal axes of the ellipsoid, and the point O is its centre. If $a = b = c$ the ellipsoid is a sphere. If $a = b \neq c$ the ellipsoid is an oblate spheroid, while if $b = c \neq a$ it is a prolate spheroid. An oblate (prolate) spheroid is a surface of revolution obtained by rotating an ellipse about its minor (major) axis. The volume of an ellipsoid is $4\pi abc/3$. *See* CONOID; ELLIPSE; SOLIDS, GEOMETRIC; SPHERE; SURFACE.



ELLIPSOID (*see* TEXT)

ELLIPTIC FUNCTIONS. It is a familiar result of elementary calculus that if R is a linear or quadratic function of a variable x, then the integral of any rational function of x and $\sqrt{R}$ is an elementary function. In the next higher case, when R is a cubic or quartic with no repeated factor, no such result is possible. It was gradually recognized by mathematicians of the 18th century that these integrals $\int f(x,\sqrt{R})dx$, where f is any rational function of x and $\sqrt{R}$, were essentially new functions requiring special investigation. They were named elliptic integrals because a special case of such integrals occurred in the problem of finding the length of arc of an ellipse. An important formula connecting such integrals, equivalent to one of what are now called addition theorems, was discovered as early as 1761 by Leonhard

Euler. But the first systematic treatment of the new integrals is due to Adrien Marie Legendre, who proved, among other results, that any elliptic integral can be reduced to the sum of an elementary function and of constant multiples of integrals in the three standard forms

$$\int \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}, \int \sqrt{\frac{1-k^2x^2}{1-x^2}}\,dx,$$

$$\int \frac{dx}{(1+nx^2)\sqrt{(1-x^2)(1-k^2x^2)}}$$

where k and n are constants called the modulus and the parameter, respectively. These three integrals are called Legendre's standard (or normal) elliptic integrals of the first, second and third kinds, respectively.

But before Legendre finished his *Traité des fonctions elliptiques et des intégrales Eulériennes* in 1826, Niels Henrik Abel, working closely with Carl Gustav Jacob Jacobi, revolutionized the subject. Let

$$u = \int_0^x \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}$$

If the parameter k is 0 in this integral, then $u = \sin^{-1}x$ or $x = \sin u$. In 1825 Abel recognized this fact and inverted the relation between u and x (for k not necessarily 0) by treating x as a function of u instead of following Legendre in treating the integral u as a function of x. The function x obtained in this manner is denoted by sn u or, more precisely, by sn (u, k). Two more functions, cn u and dn u, are defined similarly; together with sn u they are called Jacobi's elliptic functions, as distinguished from Legendre's elliptic integrals which, however, Legendre himself called *fonctions elliptiques*. From the analogy with trigonometric functions Abel and Jacobi immediately obtained important properties of the new functions: their double periodicity (see *Properties*, below) and the addition theorems which express sn $(u+v)$, etc., as rational functions of sn u, cn u, dn u, and sn v, cn v, dn v. It is now known that many of the fundamental results of Abel and Jacobi had been anticipated, but not published, by Carl Friedrich Gauss.

Properties.—Jacobi's sn u, cn u, dn u are simple cases of a more general class of functions having the same or similar properties, namely, elliptic functions. Let ω_1, ω_2 be any pair of complex numbers such that ω_2/ω_1 is not real. A function $f(u)$ of a complex variable u is said to be doubly periodic in ω_1, ω_2 if $f(u+\omega_1) = f(u+\omega_2) = f(u)$. An elliptic function is a single-valued doubly periodic analytic function $f(u)$ (*see* FUNCTIONS, ANALYTIC) which has no singularities other than poles for any finite value of u. To obtain the properties of such a function it is convenient to use the geometrical representation of complex numbers (*q.v.*) by points in a plane (the Argand diagram). Any parallelogram with vertices at α, $\alpha+\omega_1$, $\alpha+\omega_1+\omega_2$, $\alpha+\omega_2$ (α arbitrary) is called a parallelogram of the periods; if the properties of $f(u)$ are known throughout any one parallelogram, then from the periodicity its properties are known everywhere, so that it suffices to study one parallelogram. The function $f(u)$ may also be considered as an analytic function on the torus obtained by identifying opposite sides of a parallelogram of periods. This interpretation of elliptic functions as functions on tori is the starting point of the modern analytic approach to the theory.

A number of important general properties of elliptic functions were established by Joseph Liouville (1809–82), mainly by the method of contour integration, including the following:

1. An elliptic function $f(u)$ which has no pole in a parallelogram of the periods is a constant.
2. The sum of the residues of $f(u)$ at the poles in a parallelogram is zero.
3. The function $f(u)$ has at least two poles, or a multiple pole, in a parallelogram.
4. The number of poles in a parallelogram is equal to the number of zeros and is also equal to the number of points at which the function has any assigned value.
5. The sum of the values of u at the poles in a parallelogram is either equal to the sum of the values of u at the zeros or differs from it by some period.

These theorems require obvious modifications if the poles, etc., are multiple or if any of them lies on the boundary of the parallelogram, and in the trivial case when $f(u)$ is merely a constant.

Theta Functions.—To study the properties of his elliptic functions, Jacobi introduced four allied functions $\theta_j(u, \tau)$ $(j = 1, 2, 3, 4)$, called theta functions. The first, $\theta_1(u, \tau)$, is defined by

$$\theta_1(u,\tau) = 2\sum_{n=0}^{\infty} (-1)^n q^{(n+1/2)^2} \sin(2n+1)\pi u,\ q = e^{i\pi\tau}$$

and θ_2, θ_3, θ_4 are defined by similar series. When u is replaced by $u + \frac{1}{2}$ or $u + \frac{\tau}{2}$, any θ is converted into another θ multiplied by a simple factor so that the quotient of any two θ's is doubly periodic in 1, τ. Thus Jacobi could express his elliptic functions as quotients of θ's and obtain most of the fundamental properties of elliptic functions by using theta functions. The series defining the θ's converge with great rapidity when $|q|$ is small, so that they are conveniently used in numerical calculations of elliptic functions.

The Weierstrass Theory.—Karl Theodor Wilhelm Weierstrass (1815–97) founded his theory of elliptic functions on a function $\mathfrak{p}(u)$ defined by

$$\mathfrak{p}(u) = \frac{1}{u^2} + \sum{}' \left(\frac{1}{(u-\omega)^2} - \frac{1}{\omega^2} \right)$$

where the sum is taken over all numbers $\omega \neq 0$ of the form $m\omega_1 + n\omega_2$, m, $n = 0, \pm 1, \pm 2, \ldots$, ω_1 and ω_2 being any given complex numbers such that ω_2/ω_1 is not real. The function $\mathfrak{p}(u)$ is an elliptic function with periods ω_1, ω_2; it has poles of order 2 at the points ω and at 0, is regular elsewhere and satisfies the differential equation

$$\mathfrak{p}'^2(u) = 4\,\mathfrak{p}^3(u) - g_2\,\mathfrak{p}(u) - g_3$$

where, g_2, g_3 are constants depending only on the periods. Thus $\mathfrak{p}(u)$ is the inverse function of the elliptic integral

$$u = \int_\infty^x \frac{dx}{\sqrt{4x^3 - g_2x - g_3}}$$

Closely connected with $\mathfrak{p}(u)$ are the functions $\zeta(u)$, defined as a doubly infinite series with simple poles, and $\sigma(u)$, defined as a doubly infinite product with no singularities. The three functions are connected by the equations

$$\zeta'(u) = -\mathfrak{p}(u),\ \sigma'(u)/\sigma(u) = \zeta(u)$$

Neither ζ nor σ is periodic, but they undergo simple changes when u is increased by a period; ζ can be taken as the integral of the second kind, and the integral of the third kind is simply expressible in terms of $\sigma(u)$.

The Liouville theory leads readily to the addition theorem for $\mathfrak{p}(u)$, to formulas for the addition of half periods and to many other results, the development being in general much more symmetrical and elegant than that for the Jacobian functions. With $\sigma(u)$ are associated three other functions $\sigma_j(u)$, $(j = 1, 2, 3)$, these four functions being closely related to the four θ's. The Weierstrassian and Jacobian functions are connected by equations of the form

$$\mathfrak{p}(u) = A + B/\mathrm{sn}^2(bu)$$

$$\sigma(u) = Ce^{au^2}\theta_1(cu)$$

(with similar formulas for the other σ's) A, B, C, a, b, c being constants, and the properties of either set of functions can be derived from the others.

Any elliptic function $f(u)$ can be expressed in any one of three standard forms: (1) as a product of σ's divided by another product of σ's; (2) as a linear combination of ζ's and their derivatives; (3) as a rational function of $\mathfrak{p}(u)$ and $\mathfrak{p}'(u)$. From the second form the integral of $f(u)$ is at once expressible in terms of ζ and σ.

Let $x = \mathfrak{p}(u)$, $y = \mathfrak{p}'(u)$. Then $u \leftrightarrow (x, y)$ defines a one-to-one correspondence between points u in a parallelogram of periods and points (x, y) on the plane algebraic curve defined by $y^2 = 4x^3 - g_2x - g_3$, and the elliptic functions with the given periods may be considered as rational functions on this curve. This observation enabled mathematicians to develop the theory of elliptic functions entirely within the realms of algebra and algebraic geometry without recourse to the theory of functions of a complex variable. Thus the scope of the theory of elliptic functions can be broadened considerably.

Applications.—Since elliptic integrals or functions are required for the integration of the square root of a cubic or quartic they necessarily occur in many mathematical problems. Some representative applications are as follows:

In differential equations, elliptic functions or integrals occur in the integration of Lamé's equation and of Briot and Bouquet's equation $f(y, dy/dx) = 0$.

In geometry elliptic functions or integrals arise in determining the arc of an ellipse, hyperbola or lemniscate; the parametric representation of the co-ordinates of a point on a nonsingular, plane, cubic curve or, more generally, on any curve of genus 1; the surface of an oblique cone and of an ellipsoid; geodesics on a quadric of revolution and umbilical geodesics on a general quadric.

In arithmetic elliptic functions and theta functions are used to solve problems on laws of reciprocity, higher congruences and quadratic forms.

In physics elliptic functions or integrals appear in elasticity theory, rigid-body dynamics, potential theory and many other problems.

BIBLIOGRAPHY.—F. Bowman, *Introduction to Elliptic Functions With Applications* (1854); P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Physicists* (1954); A. Fletcher, J. C. P. Miller and L. Rosenhead, *An Index of Mathematical Tables* (1946); E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 4th ed. (1940). (O. F. G. S.; K. Iw.)

ELLIS (originally SHARPE), **ALEXANDER JOHN** (1814–1890), English philologist, mathematician, musician and writer on phonetics, was born at Hoxton, London, on June 14, 1814. He was educated at Shrewsbury, Eton and Trinity college, Cambridge, and took his degree in high mathematical honours. He was the first in England to reduce the study of phonetics to a science. His most important work, to which the greater part of his life was devoted, is *On Early English Pronunciation, With Special Reference to Shakespeare and Chaucer* (1869–89). He had long been associated with Isaac Pitman in his attempts to reform English spelling, and published *A Plea for Phonotypy and Phonography* (1845) and *A Plea for Phonetic Spelling* (1848). In 1825 he adopted the name of Ellis from a relative who willed him enough money to allow him to pursue a life of study. He died in London on Oct. 28, 1890.

ELLIS, GEORGE (1753–1815), English student of early literature and writer of satirical verse who helped to found the *Anti-Jacobin*, a weekly paper opposing the spread of republican ideas, was born in the West Indies in 1753. Educated at Westminster school and at Trinity college, Cambridge, he was employed on several diplomatic missions. He contributed to the *Rolliad* and the *Probationary Odes*, collections of Whig satires against Pitt's administration, but in 1797 he and George Canning founded the Tory *Anti-Jacobin* with William Gifford as editor, to which he contributed for many years. He was one of the first to arouse interest in early English literature with his *Specimens of Early English Poets* (1790) and *Specimens of Early English Metrical Romances*, 3 vol. (1805). He also edited G. L. Way's translation of *Faibliaux* (1796). Ellis was a close friend of Sir Walter Scott, who called him "the first converser I ever knew." He died on April 10, 1815.

ELLIS, (HENRY) HAVELOCK (1859–1939), British man of letters and writer on social, individual and sexual psychology, whose work deeply influenced views about sex, was born at Croydon, Surrey, on Feb. 2, 1859, of East Anglian stock. His father was a sea captain, and his childhood was largely spent on the Pacific. After teaching for four years in New South Wales, Austr., Ellis returned to London to study medicine at St. Thomas's hospital, and qualified in 1889. He soon gave up his medical practice for literary and scientific work. He began with miscellaneous essays in an attractive style, and edited first the Mermaid series of old dramatists and later the Contemporary Science series (1889–1914). In 1898 he published the first of his *Studies in the Psychology of Sex* (completed in 7 volumes during the following 30 years). The book became the subject of legal proceedings, but succeeded in breaking the conspiracy of silence which for so long had stifled discussion of such problems. Ellis' work was based on a wide knowledge of the relevant literature and a vast collection of original data, interpreted from a biological rather than a clinical standpoint. The material contained in it and in *The World of Dreams* (1911) was freely used by Sigmund Freud, who describes Ellis' conclusions as containing "happy anticipations of our own deductions." His book on *The Criminal* (1899) formed the earliest British attempt at a psychological investigation of the causes and treatment of crime, and his *A Study of British Genius* (1904; new ed., 1927) extended Francis Galton's inquiries into the inheritance of superior mental ability. His views were expounded in a lucid and entertaining style, permeated with human sympathy and an incisive wit. Ellis died at Hintlesham, Suffolk, on July 8, 1939.

Other chief works of Havelock Ellis are: *Affirmations* (1898), *The Soul of Spain* (1908), *Sex in Relation to Society* (1910), *Impressions and Comments* (3 series, 1914–24). *The Dance of Life* (1923), *Man and Woman* (1926), *The New Spirit* (1926), *Marriage Today and Tomorrow* (1929).

See D. I. Goldberg, *Havelock Ellis: a Biographical and Critical Survey* (1926); H. Peterson, *Havelock Ellis: Philosopher of Love* (1928). (CY. B.)

ELLORA, a village in Aurangabad district, Maharashtra state, India, lies 6 mi. N.N.W. of Daulatabad on the Secunderabad-Bombay main line. Pop. (1951) 920. It is famous for the rock-cut temples just to the southeast. These "caves" differ from those at Ajanta and Elephanta in being hewn from the sloping side of a hill and not from a sheer cliff. In the case of the unique Kailasanath temple, the rock has been cut away both externally and internally from the remaining monolithic replicas of structural temples that stand in a three-sided pit. The shrines extend along the hillside for a little more than a mile and comprise: (1) caves I–XV, Buddhist, probably 6th–7th centuries A.D.; (2) caves XVI–XXIX, Brahmanical (Hindu), mainly dedicated to Siva, 7th–8th centuries, and minor adjacent caves, 10th–12th centuries; and (3) caves XXX–XXXIV, Jain, following the main Brahmanical group chronologically and architecturally, and also having lesser caves nearby. Of the Buddhist caves: I–V are *viharas*, monasteries with cells; no. X, the Visvakarma, is a splendid apsidal *caitya* hall; and nos. XI and XII, the Do Tal and Tin Tal, are masterly *viharas* on more than one level. Of the Brahmanical group the Kailasanath temple (no. XVI) stands alone. It was probably constructed during the reign of the Rashtrakuta ruler Krishna I (*c.* 756–773) in emulation of the structural Kailasanath temple of Rajasimha, the Pallava ruler at Kancheepuram (*q.v.*). A portico gives access to the 25-ft. high podium on which the pile rests. The main temple comprises a pillared hall, a small *cella* housing the *lingam* or phallic emblem of Siva, and an ambulatory with subsidiary chapels. The pavilion for Siva's bull Nandi stands in front of the central temple flanked by massive pylons and stone elephants. All the members of this great design bear rich carvings, mainly relating to the story of Siva and to the Ramayana epic. Of the other shrines, no. XV, the Dasavatara, is a spacious hall related to the later Buddhist *vihara;* while no. XXI, the Rameswara, and no. XXIX, the Dumar (correctly Dhumar) Lena, contain masterpieces of the sculptor's art.

See also INDIAN ARCHITECTURE; TEMPLE ARCHITECTURE.

BIBLIOGRAPHY.—J. Fergusson and J. Burgess, *Cave Temples of India* (1880); P. Brown, *Indian Architecture (Buddhist and Hindu)*; H. Zimmer, *The Art of Indian Asia*, 2 vol. (1955). S. K. Saraswati, *History and Culture of the Indian People*, vol. ii and iii (1951 and 1954) contain useful chapters on the architecture of the period. (F. R. A.)

ELLSWORTH, LINCOLN (1880–1951), U.S. explorer, engineer and scientist whose flights from Spitsbergen to Alaska in 1926 and from Dundee Island to the Bay of Whales in the

Antarctic in 1935 made him the first person to accomplish a transarctic and transantarctic air crossing.

Ellsworth was born in Chicago, Ill., on May 12, 1880, and was educated at Columbia and Yale universities, Kenyon college, McGill university and the London School of Mines. In 1902 he joined the first survey party of the Canadian Pacific railway as axman, exploring and surveying for the first transcontinental railway, and in 1906 he became resident engineer at Prince Rupert. He later worked for the U.S. biological survey in the Rocky Mountain states, and during World War I trained as a pilot and combat observer. In 1924 he headed an expedition under the auspices of Johns Hopkins university which made a topographic survey of trans-Andean mountains from the Amazon basin to the Pacific shores of Peru.

In 1925 an expedition with Ellsworth and Roald Amundsen as co-leaders reached 87° 44′ N. latitude in two amphibian planes; an emergency landing without radio caused the world to give them up for lost. After 30 days of grim effort carving a take off field on the rough polar ice pack, one plane was abandoned while the other returned to Spitsbergen overloaded with all six men. In 1926 the Amundsen-Ellsworth-Nobile transarctic expedition in the dirigible "Norge" made the first crossing of the polar basin, a distance of 3,393 mi. and a feat which received world-wide acclaim. In 1931 Ellsworth crossed the central part of Labrador by canoe, a distance of 800 mi., and later that same year, representing the American Geographical society, accompanied Hugo Eckener in the "Graf Zeppelin" on flights over Franz Josef Land and Novaya Zemlya.

In 1935 on the third of four private expeditions to the antarctic, with Herbert Hollick-Kenyon as pilot, Ellsworth flew from Dundee Island, off the Graham-Palmer peninsula, across the continent to the Bay of Whales, making four landings en route. On this flight the Sentinel and Eternity mountains were sighted, and all this newly discovered area was named James W. Ellsworth Land in honour of the explorer's father. (The name was subsequently changed to Ellsworth Highland.) In 1939 Ellsworth flew over and named the American Highland in the Indian Ocean quadrant, bringing the total antarctic terrain he claimed for the United States to 300,000 sq.mi.

Ellsworth wrote *Our Polar Flight* (1925) and *First Crossing of the Polar Sea* (1927), both with Amundsen; *Search* (1933); and *Beyond the Horizon* (1938). He died in New York city on May 26, 1951. (F. Ro.)

ELLSWORTH, OLIVER (1745–1807), U.S. statesman and jurist, chief author of the 1789 act establishing the U.S. federal court system and third chief justice of the United States, was born at Windsor, Conn., on April 29, 1745. He attended Yale and Princeton, from which he graduated in 1766. After pursuing theological and legal studies, he was admitted to the bar in Hartford, and represented Hartford in the Connecticut general assembly. He was subsequently state's attorney for Hartford county (1777), a member of the continental congress (1777–83), of the governor's council of Connecticut (1780–85) and a judge on the state superior court (1785–89).

In 1787 Ellsworth, together with Roger Sherman and William Samuel Johnson, represented Connecticut at the Constitutional Convention in Philadelphia, serving as a member of the important committee on detail. In the convention he proposed with Sherman the decisive "Connecticut compromise," by which the federal legislature was made to consist of two houses, the upper having equal representation from each state, the lower being chosen on the basis of population. This bargain, which insofar as it applies to the U.S. senate cannot be amended, is a keystone of the U.S. federal system. Ellsworth supported free international trade in slaves for as long as possible, insisting that moral responsibility for the trade must rest with the people of those states in which it was legally sanctioned. He vigorously defended the constitution in the Connecticut ratifying convention. His "Letters to a Landholder," printed in the *Connecticut Courant* and the *American Mercury*, had a broad influence during the ratification debates, much as the *Federalist* papers did in New York.

In 1789 Ellsworth became one of Connecticut's first United States senators and the acknowledged Federalist leader in the senate. He reported the first senate rules and suggested a plan for printing the journals, shaped the conference report on the Bill of Rights, framed the measure of admission for North Carolina, helped devise the government of the territory south of the Ohio and drafted the first bill regulating the consular service. He was chairman of the committee to establish the federal court system and the chief author of the Federal Judiciary act of 1789, the principal basis ever since of the U.S. court structure.

In 1796 Washington appointed him chief justice of the United States supreme court after John Rutledge had failed to receive senate confirmation and William Cushing, the senior associate justice, had declined. Ellsworth's service on the high court was cut short in 1800 by ill-health. His more important decisions were given on circuit such as his opinion in the Williams case (1799), which applied in the United States the common-law rule that the citizen may not expatriate himself without the consent of his government.

In 1799 he reluctantly accepted Pres. John Adams' request to join William Vans Murray and William R. Davie as commissioners to France to negotiate a new treaty. In Oct. 1800 Ellsworth persuaded Napoleon to accept a compromise convention which provided for freedom of commerce between the two nations and gave promise of preventing war between the United States and France.

From France he sent his resignation as chief justice. Until his death on Nov. 26, 1807, he lived a life of retirement in Windsor. Though his career included few acts of genius or public acclaim, Ellsworth's political skill, balanced judgment and clarity of purpose entitle him to recognition as a founding father of highest stature.

BIBLIOGRAPHY.—W. G. Brown, *Oliver Ellsworth* (1905), remains the standard life. The best short treatment is G. E. Woodbine's article in the *Dictionary of American Biography*. There is also material in H. Flanders, *Lives and Times of the Chief Justices* (1875), and in G. van Santvoord, *Sketches of the Lives and Judicial Services of the Chief Justices* (1882). A most important secondary source is Charles Warren, "New Light on the History of the Federal Judiciary Act of 1789," 37 *Harvard Law Review* (1923). (J. F. Ky.)

ELLWOOD, THOMAS (1639–1713), English author, is known for his autobiography, which is informative about Milton, and for his work in producing the first edition of George Fox's *Journal* (1694). Born into a family of Oxfordshire gentry at Crowell in Oct. 1639, he became a Quaker in 1659 through his acquaintance with the Pennington family and the preaching of James Naylor, and as a result was several times imprisoned. In 1662 the Penningtons introduced him to Milton, to whom he went frequently as Latin reader. His account of the chronology of the production of *Paradise Regained* has been doubted (see *Cambridge History of English Literature*, vol. ii [1951], p. 120), but his claim to have suggested its idea remains. Ellwood wrote polemical works such as *The Foundation of Tythes Shaken* (1678) in defense of Quakerism; two works on biblical history; and some poetry, including *Davideis* (1712). He died on May 1, 1713.

See *The History of the Life of Thomas Ellwood* (1714), ed. by S. Graveson (1906); B. S. Snell, *Thomas Ellwood* (1949).

ELM, a group of 18–20 species of deciduous (rarely semievergreen) trees belonging to the genus *Ulmus* of the family Ulmaceae (*q.v.*). While widely distributed throughout the north temperate zone, a few of them extend into the tropics at high altitudes. Elms are distinguished by alternate, simple, short-petioled leaves with doubly (rarely singly) toothed margins and oblique bases. The small, inconspicuous flowers, devoid of petals, are borne in fascicles or racemes from separate buds before leaf emergence, or in the leaf axils in the early autumn. Each flower features a bell-shaped, three- to nine- (mostly four- to eight-) lobed calyx with as many stamens as calyx lobes, and a one-celled pistil terminating in a two-parted style. The fruit is an oval to circular samara with the centrally placed seed cavity encircled by a thin, membranous or papery wing. The twigs of several species are furnished with corky ridges or wings. Six elms are indigenous to North America, all occurring east of the Rocky mountains. Three are large, important timber trees; the others are small in stature, but are well-suited for ornamental use.

J. HORACE MCFARLAND CO

LEAVES AND TREE OF THE AMERICAN ELM (ULMUS AMERICANA)

Until dutch elm disease seriously decimated elm plantings, few species were better known or enjoyed such widespread use as shade trees on streets, college campuses, town squares and commons or in parks than the American elm (*U. americana*). Its gracefully arching and vaselike crown, which even in its winter nakedness lends charm to an otherwise often bleak landscape, is most distinctive. It is found also in every state east of the Rocky mountains and occurs through southeastern Canada. The largest trees frequently surpass a height of 100 ft. and a diameter of more than 48 in., and occasionally exhibit more than 300 growth rings, which attest to their longevity. The American elm has been intimately associated with the growth and development of the United States; many a venerable tree bears the name of some famous person or event. The wood, which is strong and durable as well as beautifully grained, is used in quantity for flooring, boxes and crates, furniture and veneer, musical instruments, handles of tools and woodenware.

The slippery elm (*U. rubra*) ranges from Quebec to South Dakota, thence south to Florida and Texas. It has neither the grandeur nor the pleasing form of the foregoing species, but produces timber that is similarly used. Its fragrant, mucilaginous inner bark contains therapeutic principles, which, in the days before antibacterial drugs, were used in treating throat inflammations and cholera. The cork or rock elm (*U. thomasii*) is a large tree restricted to the northern states and adjacent Canada, and the only northern elm featuring twigs with corky wings. Its hard, heavy wood is the mechanical equal of white ash (*Fraxinus americana*). The three remaining species, all small trees, are restricted to southern forests. All feature twigs with corky wings. The winged elm (*U. alata*) is usually encountered on dry upland sites, while the cedar elm (*U. crassifolia*) and the September elm (*U. serotina*) exhibit a marked preference for soil of high water retention. Unlike other American elms, the last two species flower in the autumn.

The English elm (*U. procera*) is the most stately and widely planted of the European elms. The famous "long walk" of Windsor park is lined with this most attractive tree. A doubtful native of England, it is found throughout the greater part of Europe, thence through northern Africa, Asia Minor and across northern Asia to Japan. There are many varieties in cultivation, including forms with variegated, golden or purplish foliage and weeping habits.

The Scotch or Wych elm (*U. glabra*), a northern species, ranges from Great Britain across Europe and Siberia to Japan. Its drooping branches and smooth, thin bark serve to distinguish it from the English elm. Several varieties are prized for landscaping, the best known of which is the camperdown elm (*U. g. camperdownii*), with a small, compact, globose head of pendulous branchlets. Columnar crowns, variously coloured or shaped leaves and dwarfed or bushy habits characterize other forms.

The smooth-leaved elm (*U. carpinifolia*), a tree of the Mediterranean basin, includes many horticultural varieties. The most widely used of these is a dwarf form (*U. c. suberosa*), with corky-winged twigs. The Dutch elm (*U. hollandica*) is believed to be a hybrid between *U. glabra* and *U. carpinifolia*. The Chinese elm (*U. parvifolia*) is a commonly used ornamental in eastern and central United States. Its singly serrated leaf margins serve to distinguish it from the native species. The water elm, a small Floridian tree, is known botanically as *Planera aquatica* (Ulmaceae) and is the only species in the genus. (E. S. Hr.)

Diseases and Pests.—*Dutch Elm Disease.*—So-called because it was first described in the Netherlands in 1921, this is a very serious disease of the elm. The causal agent is the fungus *Ceratostomella ulmi*. It was recognized in North America in 1930, apparently having been brought there on elm logs imported from Europe for making furniture veneer. Some of these logs were infected with the fungus and some also carried the European elm bark beetle (*Scolytus multistriatus*), which emerged and spread the disease to nearby elms in several localities in the north central and eastern states. It became widespread in many areas from Illinois eastward to New England and Canada. The American bark beetle (*Hylurgopinus rufipes*) is also a carrier, and the disease may spread from tree to tree by root grafts.

All American and European species of elms appear susceptible to the disease except for a small number of seedlings derived from *Ulmus carpinifolia*. One of these, the Christine Buisman selection, had been rather widely planted in both Europe and America. Hybrids between resistant clones and desirable susceptible species were encouraging, although none were found completely immune. Among the Asiatic species *U. pumila*, the Siberian or dwarf elm, shows promising resistance.

It is recommended that transport of elm logs and trees be restricted from infested areas so that local spread may be reduced. Dead limbs and diseased trees should be removed and burned before the start of growth in the spring in order to destroy the overwintering bark beetles before they emerge. It is further recommended that all exposed trees be given a protective spraying with a suitable insecticide, preferably one containing DDT, and that trees be supplied with the best growing conditions.

(J. M. Bl.)

Elm Phloem Necrosis.—This viral disease, which quickly defoliates and kills elm trees, is spread by certain leafhoppers (Cicadellidae; *see* Hemiptera: *Suborder Homoptera*). An infected elm shows, besides defoliation, a characteristic slimy, tannish discoloration (so-called butterscotch symptom) of the inner bark associated with a faint wintergreen odour. Research is continuing into means of controlling this as yet unmanageable disease.

Other Diseases.—Diseases less widespread and not so destructive as the foregoing are elm mosaic (a virus causing yellow mottling of leaves); Dothiorella wilt (fungal), with symptoms resembling Dutch elm disease; canker (fungal); and several leaf spots (fungal) and rots (bacterial and fungal).

Pests.—Insect pests that directly damage elms include some of the vectors of disease that are listed above: scale insects, aphids, leafhoppers, cankerworms and several leaf and bark beetles. *See* also Plant Diseases.

ELMACIN, GEORGE (Jirjis al-Makin) (1205–1273), Christian Arab historian, known in the east by his patronymic Ibn-al-'Amid, was the author of a world chronicle *Majmu'-al-Mubarak*, the second part of which deals with Muslim history from the time of Mohammed to the end of the Abbasid dynasty (A.D. 1258) and contains important references to the eastern Christians. He is said to have been born in Egypt (1205) and to have died at Damascus (1273). His chronicle was published in Arabic and Latin in 1625, the Latin version being a translation by T. Erpenius entitled *Historia Saracena*. An abridged English translation appeared in *Purchas His Pilgrimage* (1626).

ELMHAM, THOMAS, (d. *c.* 1427) English chronicler, was a monk of St. Augustine's abbey, Canterbury, who in 1414 became prior of the Cluniac house of Lenton, near Nottingham, and a royal chaplain a year later. He resigned his priory in 1427, and probably died soon afterward. He wrote a history of St. Augustine's (ed. by T. Hardwick, 1858) and a verse life of Henry V, the *Liber Metricus* (ed. by C. A. Cole, *Memorials of Henry V*, 1858),

in which he stated that he had also written a prose life of that king. Hence he has been identified with the author of a valuable narrative covering the years 1413–16, the *Gesta Henrici Quinti* (ed. by B. Williams, 1850), written by a chaplain who took part in the Agincourt campaign; but this attribution may be regarded as doubtful. Thomas Hearne mistakenly believed him to be the author of the *Vita Henrici Quinti* (ed. by T. Hearne, 1727) now known as "the Pseudo-Elmham." (C. D. R.)

ELMHURST, a city of Du Page county in northern Illinois, U.S., 16 mi. W. of Chicago and a part of the Chicago metropolitan area. Elmhurst is primarily a residential community; its only industry is quarrying limestone. Its name is a compound of "elm," from the trees along the main avenue, and the German cognate of "forest." Its former name was Cottage Hill, coming from Hill Cottage tavern, an inn built in 1843 on a knoll midway between Chicago and the Fox valley settlements to the west and a convenient stopover for farmers from those settlements making the two-day journey to Chicago with heavily loaded wagons. In 1849 the Chicago and North Western railway was built, and German labourers brought to Elmhurst saw logs from nearby groves for the wood-burning engines. Elmhurst college, a coeducational liberal arts college, governed by the Evangelical and Reformed Church, was established in 1871, primarily for pretheological training and for preparing teachers for children of German immigrants. The city of Elmhurst was incorporated in 1910. For comparative population figures *see* table in ILLINOIS: *Population.* (H. L. SM.)

ELMIRA, a city in southern New York, U.S., the seat of Chemung county, is on the Chemung river, 60 mi. W. of Binghamton. The first settlement in the area, begun in 1788, was incorporated in 1815 as the village of Newtown. It was reincorporated and the name changed to Elmira in 1828; in 1864 it was chartered as a city. During the American Revolution the battle of Newtown took place 5 mi. S.E. on Aug. 29, 1779. In this engagement Gen. John Sullivan, commanding an American army, defeated a combined British force of Tories, under Col. John Butler, and Indians, under Joseph Brant. During the American Civil War, a prison camp was located in Elmira, 1864–65, and the graves of 2,963 Confederate soldiers who died there are in Woodlawn National cemetery. Elmira reformatory, for first offenders (male), was opened in 1876 (*see* PRISON: *United States*). Among the cultural features of the city are Elmira college, a private college for women, founded in 1855; the Arnot Art gallery; the Steele Memorial library; and the Chemung County Historical society. Mark Twain had a summer home in Elmira and did some of his writing there; his grave is in Woodlawn cemetery. Since 1930 nearby Harris hill, 859 ft. above the valley floor, has been the site of many soaring contests at which national and international glider flight records have been set. Among the more important products of Elmira's highly diversified industry are prefabricated homes, fire engines, electronic devices, television tubes, glass bottles, automobile and airplane parts, machine tools, dental supplies and business forms. Elmira has a council-manager form of government, in effect from 1934. For comparative population figures *see* table in NEW YORK: *Population.* (C. C. MA.)

ELMORE, FRANCIS EDWARD (1864–1932) and **ALEXANDER STANLEY** (1867–1944), English brothers who were the first to separate minerals successfully by oil flotation, were born at Liverpool, Francis on Nov. 9, 1864, and Stanley on Jan. 1, 1867. In their early days they were engaged, with their father, in the electrolytic refining of copper and in the production of copper tubes by electrodeposition. The "bulk oil process," the first flotation process commercially employed, was invented by Francis, patented in 1898 and brought into industrial use by his brother through the Ore Concentration syndicate. In this process the oil-coated minerals were separated from the gangue after mixing the ground ore with a sufficiently high proportion of oil (*see* also METALLURGY: *Flotation Process*). Subsequently this syndicate was taken over by the Ore Concentration company, which also acquired a number of other Elmore patents, including that for the widely used Elmore vacuum process. This required less oil, the mineral being levitated by the air bubbles released on applying vacuum to the surface of the pulp. Both brothers died at Boxmoor, Hertfordshire, Francis on July 26, 1932, and Stanley on March 4, 1944. (C. W. D.)

ELMSHORN, a town of Germany, which after partition of the nation following World War II was included in the *Land* (state) of Schleswig-Holstein, Federal Republic of Germany. It is situated on the Krückau, 34 km. (21 mi.) N.W. of Hamburg by road. Pop. (1959) 35,836. Elmshorn is an industrial town lying on the borders of the fertile marsh and the higher sandy geest. The chief buildings of interest are the Nicolaikirche and the neighbouring Möhring's wine shop, and the skyscraper of Teppich-Kibek, one of the largest carpet-exporting firms in the world. There are a wide variety of educational establishments, parks, a museum and an open-air swimming pool. Elmshorn is a junction for railways from Hamburg to the north and Denmark. There is a harbour and dockyard connected to the Elbe. Foodstuffs, leather and textiles are produced.

ELMSLEY, PETER (1774–1825), English classical scholar, best known for his valuable textual work on the plays of Sophocles and Euripides. Born in London on Feb. 5, 1774, and educated at Westminster school and at Christ Church, Oxford, he took holy orders in 1798 and was appointed principal of St. Alban hall, Oxford, and Camden professor of ancient history in 1823. He died at Oxford on March 8, 1825. He edited the *Oedipus Tyrannus* (1811) and *Oedipus Coloneus* (1823) of Sophocles; the *Heraclidae* (1813), *Medea* (1818) and *Bacchae* (1821) of Euripides; and the *Acharnians* of Aristophanes (1809). He also contributed to the *Edinburgh Review*. His work is marked by a fine sense of Attic idiom and the use of fresh textual evidence derived from the collation of manuscripts in continental libraries.

See J. E. Sandys, *History of Classical Scholarship*, vol. iii, 3rd ed. (1921). (M. L. CE.)

ELMWOOD PARK, a village of Cook county, Ill., U.S., 12 mi. N.W. of the Chicago loop, and a part of the Chicago standard metropolitan statistical area. It was incorporated in 1914 and named for the number of elm trees which had been planted there. It is primarily a residential area whose citizens commute to Chicago for employment, local business interests being confined to services and small shops. A change in zoning laws (1954) permitted and encouraged the location of light industry. For comparative population figures *see* table in ILLINOIS: *Population.* (T. J. T.)

ELONGATION, strictly "lengthening"; in astronomy, the apparent angular distance of a heavenly body from its centre of motion, as seen from the earth; designating especially the angular distance of the planet Mercury or Venus from the sun, or the apparent angle between a satellite and its primary. The greatest elongation of Venus is about 45°; that of Mercury generally ranges between 18° and 27°.

EL ORIENTE (ORIENTE or LA REGIÓN ORIENTAL), the general regional name given to the eastern part of Ecuador. It includes all the territory east of the Andes, bounded on the north by Colombia (the headwaters of the Putumayo river) and on the east by Peru. For many years this thinly settled, little-explored lowland east of the Andes was claimed by both Ecuador and Peru. The boundary was settled, on paper at least, by a treaty between Ecuador and Peru signed in Rio de Janeiro in 1942. In general the boundary was drawn some 50 mi. above the head of launch navigation on the tributaries to the Amazon. A part of the boundary, from the Putumayo to the Santiago rivers, was demarcated in the field; but south of this a new dispute arose because the Cordillera del Cóndor was found to be incorrectly located on all available maps.

El Oriente was not organized politically until 1925, when it was divided into the provinces of Napo-Pastaza (area 33,237 sq.mi.; pop. [1959] 31,012; capital Tena) and Santiago-Zamora (area 18,358 sq.mi.; pop. [1959] 24,569; capital Macas). Santiago-Zamora was later subdivided into Morona-Santiago and Zamora-Chinchipe provinces.

The whole region is covered with a tropical rain forest, about which little is known. Geologic structures suggest the possibility of oil deposits. The forests yield small quantities of rubber,

balata, tagua nuts, vanilla, medicinal and insecticidal plants and timber. Most of the settlements along the eastern rivers, which can be navigated by canoes, are maintained by missionaries. (P. E. J.)

EL ORO, the southernmost province of Ecuador, bounded north by Guayas, east by Azuay and Loja, south by Loja and west by Peru and the Gulf of Guayaquil. Area 2,288 sq.mi.; pop. (1959) 123,571. The provincial capital and commercial centre is Machala (pop. [1959] 13,100). Most of the area of the province is on the lower western slopes of the Andes and the lowlands that border the Gulf of Guayaquil, in the zone of transition between the wet country to the north and the desert to the south. In the mountains to the east is Ecuador's most productive gold mining region, and also smaller mines of copper, silver and lead. The chief crops are cacao, coffee, sugar cane, rice, fibres, tropical fruits, tagua nuts and rubber. Some pastures are used for beef cattle. In the 1950s bananas were planted under irrigation near Machala. (P. E. J.)

EL PARAÍSO, department in southern Honduras bordering Nicaragua. Pop. (1961) 106,891; area 2,787 sq.mi. The population, which is 77% rural, is located in productive highland basins and valleys in the central and southern parts. The chief products are beans and potatoes (with half the national output), harvested forage grasses, cattle, coffee, tobacco, corn, swine, poultry and vegetables. Yuscarán (pop. 1,612), the departmental capital, Danlí and several other small towns in central El Paraíso are connected by all-weather roads with Tegucigalpa. (C. F. J.)

EL PASO, a city at the extreme western tip of Texas, U.S., on the Rio Grande, opposite Ciudad Juárez in Mexico; the seat of El Paso county and a port of entry. The city lies at the foot of Mt. Franklin, below a pass where the Rio Grande issues from the bare southernmost spurs of the Rockies. The altitude is 3,762 ft.; the climate is sunny, mild and dry, with annual precipitation averaging about 8 in. Municipal deep wells supply the city with adequate water. The largest city on the Mexican border, the population in 1960 was 276,687; population of the El Paso standard metropolitan statistical area (El Paso county) was 314,070. The city and county together with Ciudad Juárez create an international community of more than 500,000 inhabitants, a metropolitan oasis hundreds of miles from any equal concentration of population and commerce. The Spanish language and the Latin heritage are important elements in the city's personality. Old adobe buildings in parts of the city are distinctly Mexican, yet in general appearance El Paso is a bustling modern American city.

Both its original establishment and its growth are attributed essentially to a strategic site upon a transcontinental crossroad. It is an important foreign trade and transportation centre on several federal highways and is served by both U.S. and Mexican railroads. Trucking systems augment these facilities, and commercial airlines operate from a municipal airport.

El Paso is a tourist gateway to Ciudad Juárez and the interior of Mexico on a Pan-American highway and to attractions in a scenic area of the United States, which include Carlsbad Caverns National Park and Big Bend National park. An estimated 1,000,000 tourists visit El Paso annually.

It is the commercial and financial centre for a widespread trade territory where livestock ranching, irrigated cotton farming and mineral production are chief resources. Located at El Paso are a copper-lead custom smelter and an electrolytic copper refinery, both among the world's largest. There are petroleum refineries, a cement plant, extensive carshops and home offices of the world's largest natural-gas distributor. El Paso is a centre for distributors and jobbers of manufactured products and goods. Many diversified small industries are components of the city's economy. A needle-goods industry, mainly the manufacture of cotton outdoor clothes, employs several thousand workers. El Paso stockyards and meat-packing plants process cattle and sheep from ranges in the southwestern U.S. and northern Mexico. El Paso is the centre of a reclamation area comprising 178,000 ac. of Rio Grande valley farmlands irrigated from Elephant Butte and Caballo reservoirs. Two-thirds of annual production is cotton, much of it high quality long-staple. It is processed by the valley's ginning and cottonseed milling industries.

Military installations are important. Fort Bliss, established as a frontier post at El Paso in 1849, was in the early 1960s the United States army air defense centre, a station of about 25,000 men on a military reservation of about 1,182,000 ac. The army's William Beaumont General hospital adjoins Fort Bliss. Also at El Paso is Biggs air force base.

City parks include a public golf course and a scenic driveway around Mt. Franklin, affording fine views of the city and valley. El Paso has public, parochial and private schools and is the seat of Texas Western college of the University of Texas (formerly College of Mines, established 1914). A symphony orchestra is supported by El Paso's citizenry. The public library, built in 1954, houses over 240,000 volumes and is of outstanding architectural merit. A museum of art built in 1960 holds a collection of old masters and exhibits contemporary works. From Dec. 26 through Jan. 1 each year the city celebrates a sun carnival.

The strategic situation of the narrow pass on the Rio Grande was recognized in 1598 by Juan de Oñate, colonizer of New Mexico. He called it El Paso del Norte ("the Pass of the North"). At this place Franciscans established a mission in 1659; the original church stands in the centre of Ciudad Juárez. Spanish survivors of the Pueblo revolt (1680–92) took refuge there and built other missions nearby. Known as Paso del Norte, the string of hamlets and riverside farms by 1776 constituted "a continuous settlement 7 leagues long" inhabited by "5,000 souls." Not until after 1827 did any settlement exist on the north bank at the present site of El Paso. This became United States territory in 1848, an army post soon thereafter and a way station for travelers on a southern route to California. The first survey of town lots was platted in 1859; incorporation did not follow until 1873. El Paso was a picturesque dusty adobe village with 736 inhabitants in 1880. Four railways arrived in 1881 and by 1890 the population was 10,338. Order replaced lawlessness before the turn of the 20th century, and the city launched upon six decades of steady growth and development. (T. Le.)

EL PETÉN (Petén), the large northern department of Guatemala. Area 13,843 sq.mi. Pop. (1957) 20,090. Its capital, Flores, is located on Lake Petén. Flores is virtually the only permanent settlement in the department, most of which is occupied by a shifting population of chicle gatherers and wood cutters. Chicle, for chewing gum, is exported partly by canoe but chiefly by airplane from Flores or from other smaller collecting stations. Throughout the dense tropical forests there are ancient Mayan ruins. (P. E. J.)

ELPHINSTONE, MOUNTSTUART (1779–1859), British colonial administrator, famous for his enlightened government of the Deccan and Bombay, was born in Dumbartonshire, Scot., on Oct. 6, 1779, the fourth son of the 11th Baron Elphinstone. He reached Calcutta in the civil service of the East India company in 1796 and entered the "diplomatic line" in 1801 as assistant to the resident at Poona, at the court of Peshwa Baji Rao II, titular head of the Maratha confederacy. He won distinction in 1803 as political agent and aide-de-camp to General Wellesley (later duke of Wellington) in the Maratha war, and was appointed resident at Nagpur in 1804. He was transferred to Sindhia's court in 1807, but was sent in 1808 to negotiate an alliance with the Afghan ruler Shah Shuja, to prevent any Napoleonic advance upon India. He returned to Poona as resident in 1811. There he skilfully performed his task of keeping the Marathas disunited, used the murder of Gangadhar Sastri, the envoy from Baroda, to force a humiliating treaty on the peshwa, and when in Nov. 1817 Baji Rao took to arms played a leading part in crushing him.

Elphinstone was largely responsible for the creation of a British administrative system in the Maratha territories annexed in 1818, first as Deccan commissioner and then, from 1819 to 1827, as governor of Bombay. Disliking the anglicized system of government introduced by Lord Cornwallis, he sought to preserve the good in Maratha institutions and to make allowance for Maratha sentiment. To the raja of Satara he restored a kingdom, to the great *jagirdars* ("territorial magnates") their lands, privileges and judi-

cial powers, to the Brahmans their temple lands and awards for learning. He tried to maintain the authority and usefulness of the village headmen and of the *panchayats* ("tribunals"), wherein village elders could administer the law locally. He made much use of Maratha officials, and sought to improve their pay and status.

Elphinstone tried to improve what he conserved. He introduced a *ryotwari* land revenue settlement with rents paid directly to the government. He swept away hampering internal tolls. He assisted the foundation of the Poona college for Hindu learning, used the Bombay Native Education society to spread an improved vernacular education, and urged the teaching of the western sciences. He planned a Benthamite code of Hindu and customary law, and produced a digest of regulations, the "Elphinstone code." After retirement he was often consulted and was twice offered the governor generalship of India. His *History of India,* two volumes (1841), and *Rise of British Power in the East* (1887), by their genuine sympathy with and appreciation of the native character, did much to correct the harsh European bias of James Mill's *History of British India.* His earlier *Account of the Kingdom of Caubul* (1815), was long a standard work. Elphinstone died in Surrey on Nov. 20, 1859.

BIBLIOGRAPHY.—Sir T. E. Colebrooke, *Life of the Hon. Mountstuart Elphinstone,* 2 vol. (1884); K. A. Ballhatchet, *Social Policy and Social Change in Western India, 1817–1830* (1957). *See* also G. W. Forrest (ed.), *Selections From the Minutes and Other Official Writings of Mountstuart Elphinstone* (1884); R. D. Choksey (ed.), *The Last Phase* (1948). (J. B. HA.)

ELPHINSTONE, WILLIAM (1431–1514), Scottish bishop and statesman, founder of Aberdeen university, was born in Glasgow, probably the son of a priest, and was educated at the university there. He was ordained priest (*c.* 1456) and after four years as a country rector went abroad to the University of Paris where he became reader in canon law; later he moved to Orléans. He returned to Scotland (*c.* 1471), was appointed rector of Glasgow university (1474) and served as official general of Glasgow diocese. He became bishop of Ross (1481) and was translated to Aberdeen (1483), but he was not consecrated until some years later. Elphinstone was sent by James III on embassies to Louis XI of France and to the English kings Edward IV, Richard III and Henry VII and served James IV on similar diplomatic missions. He was lord chancellor for a few months in 1488 and became keeper of the privy seal in 1492.

Elphinstone's later years were devoted mainly to the establishment of Aberdeen university and to the building of St. Mary's college, afterward King's college. The papal bull for the foundation was obtained in 1494 and the royal charter confirming Old Aberdeen as the university seat was granted in 1497. The university, modeled on those of Paris and Bologna, and intended to include teaching in law, medicine and divinity as well as in arts, soon became famous and popular under its first principal, the historian Hector Boece. Elphinstone was partially responsible for the introduction of printing into Scotland (1507), and the *Breviarium Aberdonense,* a liturgical collection, was compiled at his direction. He did not, however, write the lives of Scottish saints included in the *Breviarium,* nor a continuation of the *Scotichronicon.* He died at Edinburgh on Oct. 25, 1514.

See Hector Boece, *Murthlacensium et Aberdonensium episcoporum vitae,* ed. and trans. by J. Moir (1894); J. Dowden, *The Bishops of Scotland* (1912). (G. S. P.)

EL PROGRESO, a department in the upper Motagua river valley of Guatemala, northeast of Guatemala City. Area 742 sq.mi. Pop. (1964) 66,734. Its capital is El Progreso. This is the dry part of the Motagua valley, used in the colonial period for the production of nopal cactus as food for cochineal insects from which a red dye was made. The department produces sugar cane, and some maize and beans on the bordering slopes. Much land is used to pasture beef cattle. The area is served by railroad and highway. (P. E. J.)

EL QUICHÉ (QUICHÉ), a department in north-central Guatemala. Area 3,235 sq.mi. Pop. (1964) 247,775. Its capital is Santa Cruz del Quiché. The population of the department consists largely of the descendants of Maya Indians and is concentrated in the southern part of the area. The northern third of the region borders Mexico and is made up of unoccupied lowland plains. The highland Indians raise maize, beans, tobacco and potatoes. Chichicastenango, near the southern end of the department, is a colourful Indian market town and a major tourist attraction. (P. E. J.)

EL SALVADOR: *see* SALVADOR, EL.

EL SEIBO, a province in eastern Dominican Republic. Area 1,154 sq.mi.; pop. (1960) 115,604. It is a region of moderate relief, occupying the northern section of the broad eastern peninsula. Extensive fields of sugar cane on its fertile plains make it one of the chief sugar producers of the nation. Cacao, rice, corn and cattle are also significant. The province has the largest gold deposits of the country. El Seibo was established in 1845 and later reduced in size by the formation of Samaná and La Altagracia provinces. The capital, Santa Cruz de El Seibo (pop. [1960] 4,621), was founded in 1502. (D. R. D.)

ELSHEIMER (or EHLSHEIMER, as he wrote his name himself), **ADAM** (1578–1610), German painter who was one of the founders of modern landscape painting, was christened in Frankfurt am Main on March 18, 1578. A pupil of Philipp Uffenbach, he was influenced by the Frankenthal school, by Gillis van Coninxloo and by the Valckenborghs. From 1598 Elsheimer visited Munich and Venice, where he met J. Rottenhammer, arriving in Rome by 1600. Among Italian painters who had an effect on him were Giovanni Bellini, Giorgione, Veronese and Tintoretto, whose influence was still felt in Rome. In Rome, where he spent the rest of his life, Elsheimer's work and nobility of character secured him the regard of a wide circle of Italian and Dutch painters, including Rubens, who became his friend and admirer. Elsheimer's influence on the Dutch and Italian schools was considerable, both Rembrandt and Claude Lorrain being affected by him. He died in Rome, it is said after a period of imprisonment for debt, and was buried on Dec. 11, 1610.

Elsheimer's works are small, and include landscapes containing mythological or biblical figures. They are painted on copper, are undated, so that it is difficult to establish an accurate chronology, and only one is signed ("The Flood," Frankfurt). During the Roman years Elsheimer's composition became simpler and clearer, his form more plastic, his use of space, in the modern view, freer and more unified. In the neighbourhood of Rome he obtained a fresh, immediate view of nature, in consequence of which his landscapes have gained a place in the wider development of landscape painting ("Myrrha," Frankfurt). His famous "Philemon and Baucis" (Dresden) is a matchless 17th-century interior. Elsheimer's meticulous treatment of both his religious themes and his scenes from Ovid's *Metamorphoses* conform to the new naturalism. His colours, at first Venetian, became in Rome more luminous and balanced, his technique being essentially a bold treatment of light and shade. He is best represented in the Städel institute at Frankfurt am Main, in the Pinakothek of Munich and in the National gallery in London. In many ways Elsheimer can be compared with Michelangelo da Caravaggio, and is one of the founders of baroque painting. He was a fine draftsman, executing his designs with a vigorous stroke. His *gouache* landscapes have the freedom of originality. He also executed a few etchings. Elsheimer's paintings were early copied, imitated and reproduced as prints.

See H. Weizsäcker, *Adam Elsheimer der Maler von Frankfurt,* pts. i, ii (1936, 1952). (ET. H.)

ELSINORE (officially HELSINGØR), a seaport of Denmark on the east coast of the island of Sjaelland (Zealand), 45 km. (28 mi.) N. of Copenhagen by road and rail, stands at the narrowest part of the Sound. Pop. (1960) 26,658. It is opposite the Swedish town of Hälsingborg, with which it is connected by train and car ferries. Elsinore received town privileges in 1426. Kronborg castle, the Elsinore castle of Shakespeare's *Hamlet,* to the east of the town, was built by Frederick II in Dutch Renaissance style and restored by Christian IV in 1637. It played an important part in the collection of the Sound toll until 1857 and was used as a barracks from 1725 to 1929. Notable are the banqueting hall, the chapel and the museum of trade and shipping. *Hamlet* is occasionally performed in the castle. The town still possesses a Car-

melite monastery and St. Maria church, the old chemist's shop, Marienlyst castle and many old half-timbered houses. Elsinore is now a commercial and industrial town with a large shipyard, and is also noted as a tourist centre. (Ro. H. T.)

ELSSLER, FANNY (1810–1884), a brilliant and beautiful Austrian ballerina, was celebrated for her fiery performances of such spectacular dances as the Spanish cachucha and the Polish cracovienne. The warmth and spontaneity of her dancing, in marked contrast to the ethereal lightness of her greatest rival, Marie Taglioni, made the writer Théophile Gautier call her "the Spaniard from the north."

Born in Vienna, June 23, 1810, she studied under Jean Aumer and made childhood appearances at the Kaerntnerthor theatre with her sister Theresa, also a dancer. Engagements in Naples, Berlin and London brought her international fame. She made her Paris Opéra debut in Jean Coralli's ballet *La Tempête* in 1834. Her sensational success two years later in *Le Diable Boiteux*, in which she introduced her cachucha, was a serious challenge to Taglioni's supremacy. In *La Tarentule* and *La Gypsy*, she revealed extraordinary pantomimic ability. Her appearance in Taglioni's favorite role, *La Sylphide*, was less successful.

In 1840–42 Elssler toured the United States, winning extravagant adulation wherever she danced. Returning to Europe, she won further acclaim with distinguished performances of *Giselle* and *Esmeralda*. Extended tours in England, Germany, Italy and Russia preceded her retirement in 1851. She died Nov. 27, 1884 in Vienna.

See A. Ehrhard, *Une Vie de danseuse: Fanny Elssler* (1909). (Ln. Me.)

ÉLUARD, PAUL (pseudonym of Eugène Grindel) (1895–1952), French writer, one of the founders of the surrealist movement and its truest poet, was born of humble parents in Saint-Denis, Paris, Dec. 14, 1895. Illness interrupted his education, but during 18 months in a Swiss sanatorium he read poetry —Rimbaud and Lautréamont, Guillaume Apollinaire and Charles Vildrac—and thereafter began to write. Experience in the trenches in World War I, where he was badly gassed, deepened his sense of community with suffering. He was further influenced by reading Walt Whitman and by his friendship, in Paris after the war, with writers of the Dadaist movement. With them he published the first surrealist manifesto (1924). His first works appeared in 1917 and 1918; *Les Necessités de la vie et les conséquences des rêves* (1921) developed the surrealist theories of the relation between dream and reality and the free expression of thought processes.

In 1924, Éluard disappeared from Paris on a world tour, but on his return continued to write and to propagate surrealist ideas by editing the movement's two reviews. Most of his poetry of this period (*Capitale de la douleur*, 1926; *L'Amour, la Poésie*, 1929) was concerned with the relationship between men and women. An interesting experiment, with André Breton, was *L'Immaculée Conception* (1930), which attempted to convey a verbal impression of different types of mental disorder. The Spanish civil war inspired Éluard to write more militant poetry, and during World War II he became a member of the underground Communist party and took an active part in the resistance movement, in addition to writing poems clandestinely circulated and intended to strengthen morale. These were published in the collections *Poésie et Vérité* (1942), *Au Rendez-vous allemand* (1944) and *Dignes de vivre* (1944). Postwar "poems of circumstance" were collected in *Poémes politiques* (1948). He continued to be active as a Communist and traveled widely to cultural congresses. He died at Charenton-le-Pont, Nov. 18, 1952.

Whether expressing his personal dream world or his involvement in politics, Éluard's poetry is characterized by its concern for truth, and by simplicity of language and vivid imagery.

See *Selected Writings of Paul Eluard* (1951).

ELURU (formerly called Ellore), a town in Andhra Pradesh, India. Pop. (1961) 108,321. It is the headquarters of West Godavari district, and is situated on the Madras-Calcutta railway about 306 mi. N. of Madras. The canal systems of the Godavari and the Krishna deltas meet there. The town is an important grain market, and has several rice mills and a jute mill; the jute is imported chiefly from Cuttack and Vizianagram, and the markets for the jute products (including gunnies and jute twine) are the Northern Circars, Madras and parts of Andhra Pradesh. Eluru also has woolen and hosiery manufactures, a tannery, saltpetre works, oil mills and cotton ginning and pressing factories. (S. Ah.)

ELVIRA, SYNOD OF, an ecclesiastical synod held in Spain, probably in 305 or 306, at Elvira, not far from or perhaps identical with the modern Granada. There, 19 bishops and 24 presbyters, from all parts of Spain but chiefly from the south, assembled, with a view to restoring order and discipline in the church. The 81 canons which were adopted reflect with considerable fullness the internal life and external relations of the Spanish church of the 4th century.

See A. W. W. Dale, *The Synod of Elvira and Christian Life in the Fourth Century* (1882).

ELWES, GERVASE CARY (1866–1921), outstanding English tenor oratorio singer of his day, was born near Northampton on Nov. 15, 1866, and educated at Birmingham and Christ Church, Oxford. While on diplomatic service he studied music in Vienna, Paris and Brussels, and in 1903 made his professional debut at the Westmorland music festival. His sincerity and intelligence were particularly admired in Bach's *St. Matthew Passion* and Elgar's *Dream of Gerontius*. In recital he often collaborated with the pianist Fanny Davies and toured Europe and the United States. He was killed in an accident in Boston, Mass., Jan. 12, 1921, while on his way to fulfill an engagement at Harvard university.

See W. and R. Elwes, *Gervase Elwes* (1935). (Wi. S. M.)

ELY, RICHARD THEODORE (1854–1943), pioneer in economic reform, was born in Ripley, N.Y., April 13, 1854. After graduating from Columbia college in 1876, he earned his Ph.D. in Heidelberg under the noted German economist Karl Knies, who had great influence on him. He received honorary degrees from several U.S. institutions. He was the first teacher of economics (1881–92) in Johns Hopkins university, Baltimore, Md., and was a founder of the American Economic association (1885). He held a professorship (1892–1925) in the University of Wisconsin, Madison, and (1925–33) in Northwestern university, Evanston, Ill. A strong religious background influenced his writings. His best-known publications include *Monopolies and Trusts* (1900); *Outlines of Economics* (6th ed., 1937), the most widely used principles text for many years; *Elements of Land Economics* (1926); and *Ground Under Our Feet* (1938), an autobiography. He was a critic of classical economics and of big business, and a defender of the labour movement. (J. F. Be)

ELY, a cathedral city, urban district and market town, in the administrative county of Cambridgeshire and Isle of Ely, Eng., 16 mi. N.N.E. of Cambridge by road. Pop. (1961) 9,803. Ely stands on the west bank of the Ouse, on a hill 7 mi. long and 4 mi. wide known as the Isle of Ely, which rises above the surrounding fens. The cathedral, towering above the town, is a landmark far over the level countryside. Ely lies at the junction of a network of railway branch lines (to Huntingdon, Peterborough, Norwich and Newmarket). By means of the Great Ouse, on which there has been a quay since the 15th century, and the Cam it has a wide inland navigation. A large beet-sugar factory was erected in 1925, and jam, baskets and agricultural machinery are also made. There is yachting on the river and coarse fishing. St. Mary's parish church ranges in style from Norman to Perpendicular, but in the main represents the Early English structure of Bishop Eustace (d. 1215). The Lady chapel of the cathedral was also a parish church from 1566 until 1938, when it reverted to its original purpose. In Cromwell house, adjoining St. Mary's churchyard, Oliver Cromwell lived from 1636 till 1647 as farmer of the cathedral tithes. The grammar school, known as the King's school, was founded in the 11th century and refounded in 1541. Its assembly hall is the old gatehouse of the monastery, known as the Ely Porta, which is near the motte or mound where a 12th-century castle once stood. Its chapel was built by Prior John de Crauden (1321–41). Among the other educational institutions are Needham's school, founded in 1734 and reorganized in 1933, when it became a school for senior boys of the city, the High School for

Girls (1905) and a theological college (Church of England) founded in 1876 and opened in 1881.

The Cathedral.—The cathedral church of the Holy and Undivided Trinity is a cruciform structure 537 ft. long and 190 ft. across the great transepts (exterior measurements). A relic of the Saxon foundation is preserved in the cross of St. Ovin (*c.* 670), and a pre-Norman window is preserved in the triforium, which was unearthed near the cathedral. The nave, which is Norman throughout, is 208 ft. in length, 72 ft. 9 in. to the top of the walls and 77 ft. 3 in. broad, including the aisles. The great western tower and the transept, which are also Norman, were completed in the 12th century. The tower, which is 215 ft. high, is surmounted by a Decorated octagon with partly detached side turrets, and underwent alteration and strengthening in the Perpendicular period. The Galilee or western porch, by which the cathedral is entered, is the work of Bishop Eustace, and is a perfect example of Early English style. In 1322 the Norman central tower, erected by Abbot Simeon, fell. Alan de Walsingham, sacrist of the church, designed its restoration (1322–28) in the form of the present Decorated octagon, a beautiful and unique conception. Instead of the ordinary four-arched central crossing, an octagon is formed at the crossing, the arches of the nave aisles and choir aisles being set obliquely, while upon the ribbed vaulting of wood above it rose the lofty lantern, octagonal also, with its angles set opposite to those of the octagon below. The total height is 170 ft. 7 in. Under Bishop John of Hotham (1316–37) the Lady chapel, a beautiful example of Decorated work, was begun in 1321. It is attributed to John Wisbech, a monk, working under Alan de Walsingham, and was completed in 1349.

Of the seven bays of the choir the four easternmost (as well as the two beyond forming the retrochoir) were built by Bishop Hugh of Northwold (d. 1254), and the three western bays which were destroyed by the fall of the tower were rebuilt by Bishop Hotham and completed by his successors. The earlier portion is a superb example of Early English work, while the later is perhaps the best example of pure Decorated in England; the elaborate choir stalls are attributed to Richard de Saxmundham (1338–46). The Perpendicular style is represented by the windows of the aisles and certain other details. There are also some splendid tombs, such as that of Bishop John Redman (d. 1505), and the two chapels of Bishop John Alcock (1486–1501) and Bishop Nicolas West (d. 1534) in this style. Among earlier monuments the canopied tomb of Bishop William de Luda (1290–98) and the effigy of Bishop Northwold (1254) are notable.

Restoration, after 1845, included the erection of the modern reredos and choir screen, both designed by Sir George Gilbert Scott, and the painting of the nave roof by H. Styleman le Strange (d. 1855), who was succeeded by T. Gambier Parry. Parry also richly ornamented the octagon and lantern in 14th-century style.

The bulk of the remains of the monastic buildings other than the Porta and Prior Crauden's chapel, are incorporated in the deanery and canons' and other residences. Many are of very early date, such as the Norman undercroft of the prior's hall and the notable Transitional Norman chancel of the infirmary chapel. The remnants of the cloisters show a 15th-century reconstruction, but the prior's and monks' doorways from the cloisters into the cathedral are Decorated late Norman. The bishop's palace, now a school for crippled girls, has towers erected by Bishop Alcock. In the muniment room of the chapter is preserved the *Liber Eliensis*, a history of the monastery written mainly by the monk known as Thomas of Ely (d. *c.* 1174), of which the first part, which extends to the year 960, contains a life of St. Etheldreda, while the second is continued to the year 1107.

A. F. KERSTING

CATHEDRAL CHURCH OF THE HOLY AND UNDIVIDED TRINITY, ELY, ENG.: 11TH CENTURY

History.—The architectural foundations of the present cathedral were laid by its first Norman abbot, Simeon (1081–94), but the reputation of Ely had been established long before, when Etheldreda (Aethelthryth), daughter of Anna, king of East Anglia, and wife of Egfrith, king of Northumbria, founded a religious house there, in all probability a mixed community. In 673 she became its first abbess and gave the whole Isle of Ely to the foundation. In 870 the monastery was destroyed by the Danes and remained in ruins till 970, when Aethelwold, bishop of Winchester, founded a new Benedictine monastery there. King Edgar in 970 endowed the monks with the former possessions of the convent and certain other privileges and revenues. The wealth and importance of Ely rose, and its abbots held the post of chancellors of the king's court alternately with the abbots of Glastonbury and of St. Augustine's, Canterbury. In the struggle against William the Conqueror, of which Hereward (*q.v.*) was the hero, Ely again became a scene of contest. Finally, in 1071, the monks agreed to surrender the Isle of Ely to the king on condition that he confirm all the possessions and privileges held by them in the time of Edward the Confessor. Abbot Simeon, who now began the reconstruction of the church, was related to William and was the brother of Walkelin, first Norman bishop of Winchester. Under Abbot Richard (1100–07) the translation from the Saxon church of the bodies of St. Etheldreda and the two abbesses who had followed her, and their enshrinement in the new edifice, took place. The feast of St. Etheldreda, or St. Audrey as she was generally called, was the occasion every year for a large fair at which "trifling objects" were sold to pilgrims by way of souvenirs; whence the word "tawdry." In about 1240 the two ancient hospitals of St. John the Baptist and St. Mary Magdalen were united and placed under the control of the sacrist of the priory. Remains of the buildings form part of St. John's farm.

In 1108–09 the Isle of Ely, most of Cambridgeshire and the abbeys of Thorney and Cetricht were separated from the diocese of Lincoln and converted into a new diocese, Ely being the seat of the bishopric. After the dissolution of the monasteries Henry VIII converted the conventual church into a cathedral (1541), granting to the dean and chapter the cathedral with its revenues and precincts. The diocese was enlarged in 1837 and 1839, and eventually covered nearly the whole of Cambridgeshire, Huntingdonshire and Bedfordshire, part of Suffolk and small portions of Essex, Norfolk, Northamptonshire, Hertfordshire and Buckinghamshire. Subsequent changes reduced the diocese to the counties of Cambridge and Huntingdon and three rural deaneries in west Norfolk. The parish of Heydon, in Cambridgeshire, is, however, in the diocese of Chelmsford.

Ely, which according to Bede derives its name from the eels in the waters about it, was a borough by prescription at least as early as William I's reign. For long the abbot, and later the bishop, had almost absolute power. The bailiff was chosen by the bishop until 1850, when a local board was appointed. The bishop was granted a fair by Richard I. In 1319–20 Bishop Hotham was also given the right to hold a fair, and the bishop still holds a fair in May and October (St. Etheldreda's) in the marketplace. The markets are claimed under a bishop's undated charter. In 1295 Ely sent two members to Parliament but was not represented thereafter as a township.

ISLE OF ELY, covering geographically the northern half of Cambridgeshire, was amalgamated with that county on April 1, 1965, to form one administrative county—Cambridgeshire and Isle of Ely. Area 375 sq.mi. Pop. (1961) 89,180. The Isle of Ely has one municipal borough, Wisbech (*q.v.*), and four urban districts: Ely, Chatteris, March, the administrative headquarters, and Whittlesey. Market gardening, horticulture, the cultivation of sugar beet, and the raising of pigs and poultry predominate.

See *Victoria County History of Cambridgeshire*, vol. iv, *Isle of Ely* (1953); N. Pevsner, *Cambridgeshire* (1954).

ELYMUS, a genus of about 40 species of grasses of temperate regions, commonly known as wild rye grass or lyme grass. One

species, *E. arenarius*, sea lyme grass or dune grass, especially when intermixed with *Ammophila arenaria*, beach grass, is much employed as a soil binder on coastal sand dunes. (J. M. Bl.)

ELYOT, SIR THOMAS (*c.* 1490–1546), English author, landed gentleman and administrator, is memorable as a man who used English prose for subjects then customarily treated in Latin, handling the vernacular tenderly and, both as moral and political philosopher and as lexicographer, endeavouring to "augment our Englysshe tongue" as a medium for ideas. He was the son of Sir Richard Elyot, judge of common pleas. The date and place of his birth are unknown, those of his education uncertain; he claimed to have been his own teacher from the age of 12 (preface to his *Dictionary*) except for some instruction in medicine (preface to his *Castel of Helthe*). In 1511 he accompanied his father as clerk of assize on the western circuit. He was justice of the peace for Oxfordshire (1522), clerk to the privy council (1523–30) and was knighted in 1530. He married Margaret Abarrow. Both were of Sir Thomas More's circle; Elyot afterward deprecated the consequences of his friendship with More. In 1531 he published *The Boke Named the Governour*, dedicated to Henry VIII, and that autumn went as the king's envoy to the court of Charles V, with instructions to further Henry's divorce from Catherine of Aragon and to apprehend William Tyndale. He served on the commission on the monasteries, represented Cambridge in parliament (1542) and was at various times sheriff of Oxfordshire and Berkshire, Cambridgeshire and Huntingdonshire. He died on March 26, 1546, and was buried at Carleton in Cambridgeshire.

Elyot's very popular *Governour*, a plan for the upbringing of gentlemen's sons who were to bear authority in the realm, was compiled from many sources, ancient and contemporary, and modeled on Francesco Patrizi's *De regno et regis institutione*. In 1536 he almost certainly wrote and probably published his *Castel of Helthe*, though the earliest extant edition is 1539; it is a popular regimen of health which, written by a layman and in the vernacular, incurred censure but was widely read. His *Dictionary* (Latin into English), the first of its scope, was published in 1538; revised and much enlarged by himself, as *Bibliotheca Eliotae*, in 1542; edited after his death by Thomas Cooper and eventually incorporated into his own *Thesaurus* (1565).

Elyot's translations include: *The Doctrinall of Princis*, from Isocrates (1534); *A Swete and Devoute Sermon of Mortalitie of Man*, from Cyprian, together with *The Rules of a Christian Lyfe*, from Pico della Mirandola (1534); *The Education or Bringinge up of Children*, from Plutarch (*c.* 1535). He wrote or compiled: *Pasquil the Playne* (1532); *Of the Knowledge Whiche Maketh a Wise Man* (1533); *The Bankette of Sapience* (1539); *The Defence of Good Women* (1540); *The Image of Governance Compiled of the Actes of Alexander Severus* (1541); and *A Preservative Agaynste Deth* (1545). Roger Ascham recollected Elyot's having "in hand" a historical work: *De rebus memorabilibus Angliae*.

His *Image of Governance*, purporting to be principally a translation from a memoir of the emperor by his supposed secretary, was popular and frequently imitated until it was denounced as imposture by William Wotton and others. It is didactic fiction, substantiated largely from the *Historia Augusta*, and modeled on the Marcus Aurelius romance of Antonio de Guevara (whom Elyot may have known); reference to a Greek manuscript may be mere literary convention. The end, as with all his works, was evidently usefulness; the means were probably clearer to his contemporaries than they are to us.

The edition of the *Governour* by H. H. S. Croft, 2 vol. (1880), with its biographical introduction, remains the principal source of information. *See* also D. T. Starnes, *Renaissance Dictionaries* (1954). (M. M. Ls.)

ELYRIA, a city of northern Ohio, U.S., at the confluence of the east and west branches of the Black river, 9 mi. S. of Lake Erie and 25 mi. W.S.W. of Cleveland; the seat of Lorain county. Products of the city's industries include heating and air-conditioning units, automotive air brakes, foundry and machine-screw products, lace, chemicals, glass-lined steel equipment, springs, power shovel and crane parts, pipe-welding machinery, plastics, chrome products, golf balls, aircraft pumps, electric motors, steel tubing and auto accessories. Elyria was first settled in 1817 by Heman Ely of West Springfield, Mass., on Connecticut Western Reserve land owned by his father. Its centre was a mill site. It became the county seat in 1823 and a chartered city in 1892. English, German, Polish and Hungarian nationalities are represented in the population. For comparative population figures *see* table in Ohio: *Population*. (R. C. Do.)

ELYSIUM (Elysian Plain or Fields), the pre-Hellenic paradise, identified by the Greeks with their own Isles of the Blest (*q.v.*). In Homer the Elysian Plain is a land of perfect happiness, ruled by Rhadamanthus, at the end of the earth, on the banks of the Oceanus river. Much the same description is given by Hesiod of the Isles of the Blest. In the earlier authors, only those specially favoured of the gods enter, and in Homer they are carried there body and soul and made immortal. In Hesiod Elysium is already a place for the blessed dead, and, from Pindar on, entrance is gained by a righteous life, a conception partly at least due to Orphism. In later authors, such as Virgil, Elysium is part of Hades; originally it had nothing to do with it.

ELZEVIR (Elsevier), the name of a famous family of Dutch booksellers, publishers and printers, 15 members of which were in business between 1581 and 1712. Louis (1546?–1617), son of a printer of Louvain, settled in Leiden as a Protestant *emigré* (about 1581), set himself up as a bookbinder and bookseller, published more than 100 books and began the family specialization in learned books. The business in Leiden enjoyed its greatest success between 1622 and 1652 under the management of his son Bonaventure (1583–1652) and grandson Abraham (1592–1652). In 1626 they purchased the printing equipment, including a valuable Oriental press, of Isaac (1596–1651), another grandson, and became printers to the university (an office held by their successors until 1712). They were enterprising in their publishing, and printed with taste and accuracy. Two notable and widely imitated series were their *Petites Républiques*, 35 volumes in 24mo, concerned with different countries and published between 1625 and 1649; and their literary classics in small 12mo. The Leiden bookshop was closed in 1659 but publishing and printing continued, although in declining quantity and quality, until 1712.

The earliest branch of the Elzevirs was established in The Hague about 1590 by Louis (1566–1620), son of the founder, and continued until about 1665, chiefly as booksellers. Less important was the bookselling branch at Utrecht, begun about 1603 by Joost (1597–1652?), another son, and revived between 1667 and 1675 by his grandson, Pierre (1643–96). The largest and most successful branch was in Amsterdam, started in 1638 as a bookshop by Louis (1604–70), a grandson of the founder. By 1640 it had acquired a printing press, and after the arrival in 1655 of Daniel (1626–80), another grandson, it became the chief Elzevir establishment. It closed in 1681.

After having enjoyed an almost legendary reputation among bibliophiles for excellence of typography and design, the Elzevirs' work is now regarded as only typical of the high quality of their day in Holland. Among their four chief typographical devices, the "solitary," consisting of an elm tree, a fruitful vine and a man alone, with a motto *Non solus*, is perhaps the best known.

Bibliography.—A. Willems, *Les Elzevier* (1880); bibliographical studies by G. Berghmar (1885 and 1897) and H. B. Copinger (1927); E. Kelchner (ed.), *Catalogus librorum officinae Elsevirianae 1628* (1880); D. W. Davies, *The World of the Elseviers, 1580–1712* (1954); S. L. Hartz, *The Elseviers and Their Contemporaries* (1955).

EMANATION is a term rather loosely and often inaccurately applied, in describing various metaphysical theories, to mean the production of reality from a supreme principle in ordered stages of diminishing excellence. Several such theories were current in the last century B.C. and the first centuries A.D. The oldest seems to stem from an attempt to combine Stoic and Platonic doctrine; it dates back at least to the 1st century B.C. and may be due to the Stoic philosopher Posidonius (*c.* 135–*c.* 50 B.C.). It represents the soul as a material outflowing from the divinity, which, as in all Stoic thought, is itself material, an intelligent fire. It has two distinctive features found in all later theories of emanation: the productive principle is undiminished and unchanged by the giving

out of its product (this distinguishes theories of emanation from pantheism); and the product is necessarily inferior to the producer.

The Gnostic systems, pagan and Christian or Christianized, of the first centuries A.D. give many accounts, often fantastically elaborate, of how a hierarchy of realities, in descending stages, is produced from the supreme principle. In the pagan Gnosticism of the Hermetic treatises, especially the *Poimandres*, the Stoic-Platonic influence is marked; the emanation is real, an outflowing of the divine light. In some of the Christian Gnostic systems, notably that of Valentinus, the production of the vast succession of "aeons" or spiritual beings is described in sexual terms. The first principle has a consort in whom he deposits his seed, and the aeons are generated in pairs, male and female, which in their turn have children; how far this is meant literally and how far metaphorically is difficult to determine. The Gnostic stories of the origin of reality differ from Neoplatonism by being violently dramatic and arbitrary. They are works of a somewhat over-excited religious imagination rather than of reason, though that of Valentinus, at least, has great poetic beauty and power.

The Greek philosophy of the first centuries A.D. which led up to Neoplatonism uses the idea of a hierarchy of divine principles, two or three in number; the Neoplatonism of Plotinus has a doctrine of the generation of the divine Intellect from the first principle, the One, and of Soul from Intellect: this is often referred to as "emanation." Plotinus indeed uses the metaphors of the old Stoic-Platonic emanation theory, especially that of the radiation of light, to describe this generation, and he maintains the principles of the immutable producer and the inferior product. But he views the divine generations as timeless and immaterial; no ancient philosopher purged his thought so well of unconscious material and spatial conceptions as Plotinus. Moreover, for him the moment of conversion, in which the reality produced assumes its own being by returning in contemplation on its producer, is at least as important as the moment of generation; this is alien to "emanational" ways of thinking.

In striving for a perfectly smooth transition between levels of reality the later Neoplatonists over-elaborate Plotinus' simple hierarchy of derived being. They assert a distinction as real wherever it can be made in thought, and try to accommodate in their system every divine and semidivine being of classical and oriental mythology. But their principles remain those of Plotinus. Though their systems look like those of the Gnostics, they are very different; here it is not religious imagination which runs mad, but the distinguishing and classifying reason.

See also GNOSTICISM; NEOPLATONISM; PLOTINUS; POSEIDONIUS; STOICS. (A. H. AG.)

EMBA, a river in the Aktyubinsk and Guryev *oblasts* in the Kazakh Soviet Socialist Republic of the U.S.S.R. The river itself is of little importance since in dry years its waters do not reach the sea; but its basin contains the important Emba oil fields which extend from the Mugodzhar hills in the east to the lower reaches of the Volga in the west, and from the southern spurs of the Ural mountains in the north to the northern shores of the Caspian and the Ustyurt plateau in the south. The leading fields—Dossor, Makat, Iskininski, Baychunas, Koschagyl and Shubar-Kuduk—produce one of the highest-grade oils in the U.S.S.R. and reserves have been estimated at 1,000,000,000 tons. (G. E. WR.)

EMBALMING, the art of preserving dead bodies from decay, had an early origin in ancient Egypt and reached a high degree of perfection there. For details of the Egyptian methods, *see* MUMMY.

Human remains, preserved naturally or artificially, have been found in countries other than Egypt. Burial in sandy soil in hot climates has produced natural mummification by desiccation in North, Central and South America and elsewhere, and well-preserved bodies also have been recovered from the bogs of Denmark, Ireland and Scotland in situations where air and decay organisms did not come in contact with the bodies. Artificial mummification was achieved early by packing the body in salt, immersing it in brine or in honey, or coating it with wax.

The Greeks and Romans practised embalming. Embalming of kings and other notabilities started perhaps in the 4th century A.D., but it was still inefficient in the middle ages and after. Many physicians of the period A.D. 500–1700 wrote on the subject, notably Rhazes, Peter Forestus, Girolamo Segato, P. Sue, Louis Penicher, J. Lanzoni, Ludwig de Bils and Gabriel Clauderus. Sue lists many aromatic substances reputed to resist putrefaction and Penicher recommends balsams, while Clauderus injected a fluid containing tartar into the body cavities and then immersed the body in the fluid, later removing it and drying it by heat. Other methods included evisceration, insertion of herbs into the soft tissues through incisions, immersion of the body in spirits of wine and bandaging of the body with waxed or tarred sheets. Frederik Ruysch (1638–1731) of Amsterdam is credited with being the first to embalm by injecting a preservative solution into the blood vessels, though his technique is unknown. Pickling of the body in brine or alcohol in a sealed container was used successfully in the 18th and early 19th centuries. William (1718–83) and John (1728–93) Hunter and their nephew Matthew Baillie (1761–1823) used comparable methods in embalming. William Hunter forced into the arteries a mixture of oils of lavender, camomile and turpentine, venice turpentine and also vermilion; the viscera were then removed, the vessels reinjected and ligatured, the cavities washed and the viscera replaced surrounded by camphor, resin and nitre. Baillie did not eviscerate but injected the stomach, bowels and air passages with the preservative fluid.

European embalmers of the 19th century and the chemicals they used for vascular injection included: J. N. Gannal, a French chemist, who used alum, common salt and potassium nitrate; Tranchini, of Naples, who used arsenic; Sucquet, of Paris, zinc chloride; Louis Jacques Thénard, a chemist, and François Chaussier, also in France, who used bichloride of mercury. Experimental methods included the vascular injection of disinfectant gases and/or the exposure of the body to an atmosphere of similar gases.

Embalming, by modern processes, is carried out on the great majority of the dead in North America. It is much less common in Great Britain but was becoming more frequent in the second half of the 20th century. In the United States it originated in the care given to the dead of the American Civil War (1861–65) by such men as Thomas Holmes, August Renouard and Joseph Henry Clarke. They, and others, became traveling teacher-embalmers.

Modern embalming consists of the injection of several gallons of a fluid into one or more arteries. The fluid generally contains formaldehyde as a preservative, glycerol to counteract dehydration, borax to keep the blood liquid before it is drained, and sometimes dyes such as eosin. The commonest sites of injection are the armpits, the neck and the groins. The solution spreads through the vascular system and reaches the tissues. The blood displaced by it is drained off by opening a vein at the place of injection. The abdominal wall is incised and gases are allowed to escape and fluids are aspirated. The hollow organs are perforated, as is the diaphragm on each side, so giving access to the chest cavities. About 1–2 pt. of a disinfectant and preservative fluid, containing a wetting agent to facilitate penetration, is then introduced into the body cavities and the incision closed by stitching with twine. Areas of the body that have not become firm from the action of the arterial solution receive local injections. When an autopsy has been made prior to embalming, a modified form of the same technique is used: injections are made into each limb and into either side of the neck through the arteries, the organs being treated in preserving fluid and the body wall being injected locally. When preservation for a period of more than about two months is required, arterial injection is made at multiple sites (commonly six). Cosmetics, modeling wax and other materials may be used, particularly in the case of injured bodies.

Laws in all states of the United States require the incorporation of a definite percentage of formaldehyde in all embalming fluids and prohibit the use of arsenic and usually the salts of the heavy metals such as mercury, lead, zinc, etc., and various alkaloids such as strychnine. Without such a prohibition the investigation of deaths in which poisoning by these substances was suspected would be rendered impossible if embalming had taken place.

BIBLIOGRAPHY.—R. Harlan, *Gannal's History of Embalming* (1840); A. O. Spriggs, *Champion Textbook on Embalming* (1944); C. J. Polsen,

R. P. Brittain and T. K. Marshall, *The Disposal of the Dead* (1953).
(E. C. J.; Ro. P. B.)

EMBARGO, the detention of merchant vessels or other property in ports within a state's domain to prevent their departure for foreign territory. To achieve its objective, an embargo should also prohibit shipment by air and, if necessary, by land. The term "embargo" comes from the Spanish word *embargar,* "to arrest" or "to restrain." Embargo should be distinguished from *arrêt de prince,* a temporary detention of ships or aircraft; from angary (*q.v.*), an emergency requisition of the use of foreign merchant vessels; and from boycott (*q.v.*), an interruption of commercial and financial relations not confined to exports. An embargo is not imposed against enemy ships or property, because enemy character subjects these to other action.

Embargo may be civil or hostile. A civil embargo is the detention of national vessels in home ports either to protect them from foreign depredation or to prevent goods from reaching a particular state. The U.S. embargo of 1807 was a civil embargo (*see* UNITED STATES [OF AMERICA]: *History*). A hostile embargo is the detention of the transport or property of a foreign state.

An embargo may be employed either as a reprisal or for political purposes. As an act of reprisal an embargo may be imposed against a state which is felt to be violating its international legal duties. Of such nature was the U.S. embargo of 1807 against France and Great Britain. An embargo may also be employed for the political purpose of prohibiting exports of arms and other war materials to belligerent states or to states in rebellion, either in an attempt, usually collective (*see* SANCTION AND GUARANTEE), to force a cessation of hostilities or in an individual state's effort to preserve its neutrality. An example of the political embargo was the U.S. embargo of 1937, intended to promote U.S. neutrality during the Spanish civil war. An embargo, such as that imposed on the shipment of strategic materials to Japan by Pres. Franklin D. Roosevelt under the Export Control act of 1940, may also be invoked against one party to a conflict. Such an embargo constitutes intervention. An older practice, condemned by the second Hague conference but revived during World Wars I and II, is to detain merchant ships of the probable enemy.

Embargoes have been imposed by belligerents on neutral ships and by neutrals on belligerent ships. For example, during World War I, British, French and even neutral embargoes were imposed in preparation for the exercise of the right of angary. In 1918 both Argentina and Chile, which had adopted a neutral policy, detained German ships. In 1940 Norway's prohibition against the carriage of oil to Italy led to an Italian embargo against Norway. In 1941, before it became officially a belligerent, the United States seized German, Italian, Danish and French ships lying idle in U.S. waters and also froze Axis assets.

When employed as a collective measure of enforcement or prevention by an international organization, the effectiveness of an embargo depends upon co-operation both by members of the organization and, to prevent transshipment to the state subjected to such coercion, by nonmembers. The Geneva Narcotics convention of 1921 allowed employment of the embargo as a collective measure to prevent any state from acquiring excessive quantities of narcotics. An abortive effort was made in the League of Nations to impose an oil embargo upon Italy during its war against Ethiopia (1935–36). During the hostilities between Bolivia and Paraguay over the Chaco area, Bolivia accepted but Paraguay rejected League of Nations recommendations for a cease fire in 1934. Approximately 30 states responded to the League's request that existing embargoes, applicable against both states, be lifted in respect to Bolivia. In 1951, during the Korean War, the United Nations general assembly adopted a U.S. recommendation that members of the UN prohibit shipments of arms and strategic materials to territories controlled by the Chinese and North Korean Communists. Thirty-three members and five nonmembers soon declared their intention to give effect to the resolution, and later the 14 states principally involved formed an international coordinating committee to implement their decision. This embargo, a means of supplementing UN military action in Korea, continued in effect after the 1953 Panmunjom armistice. As it coincided with a redirection of Communist Chinese trade from the west to the U.S.S.R. and its satellites, it was difficult to judge the effectiveness of this measure.

See also ALIEN PROPERTY; BLOCKADE; VISIT AND SEARCH.
(W. L. GD.)

EMBASSY: *see* AMBASSADOR; DIPLOMACY; EXTERRITORIALITY.

EMBER DAYS AND EMBER WEEKS, the four "times" set apart by the Roman Catholic and Anglican churches for special prayer and fasting and the ordination of the clergy, known in the medieval church as *quatuor tempora* ("four seasons"). The derivation of the word from Anglo-Saxon *ymbren,* "a circuit" or "revolution," brings out the seasonal and cyclic aspect of the liturgical practices connected with these periods, which occur in autumn, winter, spring and summer. The ember weeks are the complete weeks following (1) the Feast of the Exaltation of the Cross (Sept. 14); (2) the Feast of St. Lucy (Dec. 13); (3) the first Sunday in Lent; and (4) Pentecost (Whitsunday). The Wednesdays, Fridays and Saturdays of these weeks are the ember days.

The exact origin of the ember seasons is not certain. Pope Leo the Great (5th century) considered them a synagogue practice continued by the apostles. The Dead sea scrolls mention special exercises of prayer for the seasons of the year. There were feasts in ancient Rome in June, September and December for the harvests of wheat, wine and oil. The ember seasons of the early church were limited to three and may have been the Christian transformation of the pagan festivals. They were dedicated primarily to recollection and spiritual renewal but also included thanksgiving for the harvests. Wednesday, Friday and Saturday were days of fast ending with a vigil service on Saturday night at which there were ordinations. At first, ordinations occurred only at the end of the ember week in December, but Pope Gelasius I (A.D. 492–496) extended the practice to the four ember seasons.

From Rome the observance of the ember weeks gradually spread throughout the Western Church, but it did not reach Gaul or Spain until the 8th century. Augustine of Canterbury (597) took the practice to Britain, from which it eventually spread to Germany. Gregory VII (1073–85) fixed it for the entire church, and this was confirmed in 1095 by Urban II in the decree *Statuimus.*

BIBLIOGRAPHY.—Dom Gregory Dix, *The Shape of the Liturgy,* 2nd ed., pp. 342 ff. (1945); J. A. Jungmann, S.J., *Public Worship,* pp. 226–229 (1957), *The Early Liturgy,* pp. 271–274 (1959); Archdale A. King, *Liturgy of the Roman Church,* p. 207 (1957); *Dictionnaire de théologie catholique,* vol. xiii, ch. 1,447–1,455. (C. J. BO.)

EMBEZZLEMENT, a statutory offense, may be described generally as the fraudulent misappropriation of goods of another by a servant, an agent or other person to whom possession of the goods has been entrusted. Such a description, however, by no means constitutes a precise definition of the offense in all of its many forms. Being a creature of statute rather than of common law, embezzlement has been variously defined by legislatures; and even similar statutory provisions have been variously interpreted by courts of different jurisdictions.

Typically, embezzlement presents a situation in which the offender first gains possession of the goods lawfully and subsequently misappropriates them. In this respect the offense is to be contrasted to the common-law crime of larceny (*q.v.*). Larceny requires an initial trespassory taking; *i.e.,* the taking of goods from the possession of another without the latter's consent. Although the trespassory-taking requirement has persisted in the modern law of larceny, the scope of larceny has, in fact, been expanded by various manipulations of the concept of possession. One of the most important of these was codified by a statute of 1529 (21 Hen. VIII, c. 7) which in effect provided that larceny may be committed when goods are taken by a servant from the home of his master or when entrusted by the master to the servant. This expansion of doctrine was based on the theory that, although the goods are in the custody or physical control of the servant, possession is in the master; and hence their misappropriation by the servant constitutes a trespassory taking. "Larceny by servant" was, however, subject to a significant limitation. Goods intended for the master but delivered to his servant by a third party were regarded as not yet in the possession of the master. Hence a con-

version of the goods by the servant at this stage did not constitute a trespassory taking, and the servant was therefore guilty of only a tort or a civil wrong.

The failure of the law of larceny to provide adequate protection for the property of employers against the depredations of servants and employees became increasingly serious with the rise of commerce and modern business organization. By the end of the 18th century the situation had become critical. In the famous case of *The King* v. *Bazeley* (168 E.R. 517), decided in 1799, the defendant, a teller in a banking firm who had pocketed a bank note handed him by a customer of the bank for deposit, was acquitted. The basis of the decision was that since the defendant misappropriated the note before possession of the master had attached, only a civil breach of trust occurred. In the same year, parliament enacted the first general embezzlement statute (39 Geo. III, c. 85). The act of 1799 applied only to servants and clerks. In 1812 the law was expanded to cover cases of agents, such as stockbrokers. Throughout the 19th century the English statutes were gradually broadened to cover an increasing number of situations. The Larceny act, 1916, contains the embezzlement provisions in effect in mid-20th century.

A wide range of embezzlement provisions may be found in the U.S. statutes. Some of the provisions contain limitations similar to the English act of 1799 that restricted the offense to misappropriation of goods received by a servant or clerk "by virtue of his employment." Others broaden the offense to any property of the principal received by the defendant employee. The Massachusetts act was held inapplicable to a case in which the property was delivered to the defendant by mistake, the provision being interpreted as requiring the property to be deliberately entrusted to the defendant by the owner. This requirement has been rejected in some other jurisdictions. Characteristically, U.S. jurisdictions have enacted long series of special provisions applying to agents, bailees, attorneys, trustees, executors, guardians and the like. The intent required to be established is generally that of depriving the owner of his goods permanently. An exception to this requirement is to be found in widely enacted provisions applicable to custodians of public funds. Under these laws defendants may be convicted and made subject to severe penalties even in the absence of an intent to steal when public funds are lost as a result of improper administration of the funds.

Viewed as a social phenomenon, embezzlement appears to be an offense frequently committed. Although sometimes classified as a white-collar crime, offenders at all levels of the social scale have been found to be involved in thefts of property entrusted to their care. U.S. criminal statistics reveal a relatively high incidence of the crime, but there is reason to believe they considerably understate its occurrence. Embezzlements are frequently undetected when property of trifling value is involved. Even when detected, the victims are often more interested in restitution of the stolen goods or money than in pressing criminal charges. The complexities of embezzlement prosecutions often induce law-enforcement agencies to withhold prosecution, particularly when the stolen goods have been restored. Even when the offender is brought to trial and convicted, he is often placed on probation upon the condition that restitution be made. Modern methods of business management have to some degree discouraged embezzlement and assisted its detection. Embezzlement of property, however, continues to present a substantial problem to all advanced economic systems. This appears true of nations like the U.S.S.R. as well as those in which property is predominantly in private hands.

See also FRAUD. (F. A. A.)

EMBIOPTERA (EMBIIDINA), an order of small, slender, chiefly tropical colonial insects remotely related to termites. Numerous silk glands, located in the enlarged basal segment of the tarsi of the first pair of legs, are present in all embiids, or web-spinners as they are commonly called. Silken strands are spun through hollow hairs on the same tarsi and are used to form a protective network of galleries that ramify through or on dead leaves, bark, lichens, moss, etc., which also constitute the food. Embiids seldom leave their galleries and usually escape their enemies by running backward. Females are wingless, males generally winged. The wings, folded over the back in repose, are stiffened for flight by blood pressure. An estimated 1,000 species exist, but many of these await discovery and description.

See also INSECT: *Insect Classification.*

See E. S. Ross, "A Revision of the Embioptera, or Web-spinners, of the New World," *Proc. U.S. Nat. Mus.*, 94:401–504 (1944).

(E. S. Ro.)

EMBLEM BOOKS, composed of emblematic pictures and passages of prose or verse to explain each picture and point its moral lesson, had a considerable vogue throughout Europe during the 16th and 17th centuries.

The first printed emblem book, Andrea Alciati's *Emblematum Liber*, published in Italy in 1531, became immensely popular—90 editions being printed before 1600—and was extensively imitated, adapted and translated. French writers quickly responded to the fashion set by Alciati. Of the more notable French emblem books Guillaume de la Perrière's *Theatre des bons Engins* appeared in 1539 and Claude Paradin's *Devises Heroïques* in 1557.

In the Low Countries emblem pictures designed to satisfy a characteristic taste for realistic domestic scenes, and the then unrivaled skill of Dutch and Flemish engravers, gave the emblem book greater artistic merit. The emblems of Jakob Cats (*q.v.*) were among the early models of Sir Joshua Reynolds.

Added vitality was given to the emblem book by the Jesuits, who adopted it for educational and missionary purposes. One very popular and influential Jesuit emblem book was Hermann Hugo's *Pia Desideria* (1624), afterward translated into English for Protestant use by Edmund Arwaker (1686).

English emblem books are comparatively few and, since their authors borrowed largely from the continent, generally of no marked originality. The earliest, Geffrey Whitney's *A Choice of Emblemes*, 2 parts (1586), is virtually an anthology. Even the best English emblem book, Francis Quarles' *Emblemes* (1635), was derived from two Jesuit works, Hugo's *Pia Desideria* and *Typus Mundi* (1627) by several hands.

But Quarles (*q.v.*) and his followers, *e.g.*, Christopher Harvey, whose *Schola Cordis* (1647) was similarly derivative, and Edmund Arwaker, produced emblem books of a new kind, remarkable for their poetic qualities. When Quarles' *Emblemes* appeared the emblem book could, for the first time, claim serious consideration as literature.

Many emblem books were used by contemporary writers and craftsmen. For example, some of the devices borne in the tournament in Shakespeare's *Pericles* probably come from Paradin and the splendid Vyvyan salt in the Victoria and Albert museum, London, is decorated with designs from Whitney.

A few emblem books (including those by Hugo and Quarles) continued to find some favour as devotional or instructional works, but in the general change of taste that occurred at the end of the 17th century popular demand for the emblem book vanished.

BIBLIOGRAPHY.—L. Volkmann, *Bilderschriften der Renaissance* (1923); M. Praz, *Studies in Seventeenth-Century Imagery*, 2 vol. (1939, 1947); R. Freeman, *English Emblem Books* (1948); R. J. Clements, *Picta Poesis: Literary and Humanistic Theory in Renaissance Emblem Books* (1960). (JN. HN.)

EMBOLISM: *see* THROMBOSIS AND EMBOLISM.

EMBOSSING, the art of producing raised patterns on the surface of metal, leather, textile fabrics, cardboard, paper and similar substances. Strictly speaking, the term is applicable only to raised impressions produced by means of engraved dies or plates. Thus raised patterns produced by carving, chiseling, casting and chasing or hammering are not embossed work. Crests, monograms, addresses, etc., are embossed on paper and envelopes from dies set either in small handscrew presses or, when large quantities are desired, in ordinary letter press. The blocked ornaments on cloth covers for books, and the blocking or imitation tooling on the less expensive kinds of leather work, are effected by means of powerful embossing or arming presses. (*See* BOOKBINDING.) For impressing embossed patterns on wallpapers, textiles of various kinds, and felt, cylinders of copper, engraved with the patterns to be raised, are mounted in calender frames, in which they press against rollers having a yielding surface, or so constructed that depressions in the

engraved cylinders mesh with corresponding elevations in those against which they press. The operations of embossing and colour printing are also sometimes combined in a modification of the ordinary rotary printing press.

A modified embossing technique came into use in the 1950s for producing maps. The maps were first printed in full colour on flat sheets, usually of heavy paper impregnated with a plastic compound; each sheet was then accurately positioned, face down, over a heated die and drawn against the die by vacuum, thus reproducing relief features such as mountains in raised form on the finished map. *See* also METALWORK, DECORATIVE: *Techniques of Metalworking*.

EMBROIDERY is the decoration of textile fabric (and occasionally of leather or other material) by means of a needle and thread (and sometimes fine wire). The basic techniques are described in the article NEEDLEWORK. Like other crafts, embroidery probably had its origin in the daily needs of humanity. The problem of how to join two edges of a fabric together must have arisen very early, and even a seam, as so many oriental embroideries show, may be treated decoratively. Such work as the quillwork of North American Indians and the decoration of bark cloth by African tribes also may represent primitive origins of embroidery. Different styles of embroidery, depending on the materials available and on the local culture, have evolved in different parts of the world.

Paintings and carvings from Egypt indicate that garments were embroidered in very early times; tomb paintings show that clothes, couch covers, hangings and tents were decorated, and since foreigners are portrayed in ornate clothes, the art of embroidery must have been highly developed in other countries as well as Egypt. Sails for ships and for ceremonial boats were probably made of a simple form of patchwork with small embroidered details of design. In the Old Testament, the Books of Exodus and Ezekiel record that among the Hebrews gold plate and wire, similar to those used in the 20th century, were worked on fine linen in rich colours for liturgical garments, and ordinary clothes and sails were also decorated. Embroidery is thought to have been executed in China by 1000 B.C., and some of the traditional ornaments are described in the *Shu Ching* (*c.* 6th century B.C.).

Excavations in the Altai mountains of Mongolia have revealed woolen tunics and jackets or jerkins of Scythian make, the tunics neatly stitched and covered with ornate and spirited appliqué designs, the leather or felt jerkins also lavishly trimmed. Women's clothes and also belts, bags, satchels and saddlecloths were embroidered with appliqué. The Coptic period of Egypt (1st century A.D.) saw the introduction of Christian symbols into embroidery, particularly on tunics. Designs were always executed in purples with the details picked out in stem stitch with natural-coloured linen thread. By the 4th–5th centuries A.D., the designs were finer and more elaborate, and small touches of other colours were introduced, mainly pinkish-red, yellow and green.

After this period there is enough material extant so that the history of embroidery in many countries may be traced.

MEDITERRANEAN WORLD

Greek Islands.—No embroidery from ancient Greece has survived. The origin of that in the islands is very uncertain, but it probably took place sometime between the 12th and 15th centuries; it was discontinued during the Napoleonic Wars when the mulberry trees, the ultimate source of the islanders' silk thread, were cut down and afterward replaced by olive trees. The embroideries were executed with untwisted silk thread on a linen ground, never more than 20 in. wide, and soft blues and reds are characteristic colours. The Ionian Islands were on a trade route that brought Persian and Italian influence into their designs, particularly the pot of flowers and wide border patterns on dresses, curtains and valances. Cretan embroideries used, among other colours, a sharp acid yellow and black and were sewn with a stitch now called Cretan stitch. In the Cyclades, designs were based on an irregular hexagon, known as the king pattern, and a diamond, known as the queen pattern, which were subdivided in various ways and arranged to form an all-over pattern; the design was carried out in darning stitch, as is also the star design of Naxos. Rhodes, as the headquarters of the Knights Templars, was subject to European influence. Bed furnishings were tent shaped, with embroidery round the slit entrance to give a doorlike effect, and borders were worked on pillowcases. Scyros shows Turkish influence, with semi-oriental designs of branching flowers, birds and ships, often with amusing figures perched in the rigging; in this area, the traditional embroidery lasted almost to the mid-20th century.

Italy.—Many varied styles of peasant and fashion embroideries are found in Italy which, in addition, is especially rich in embroidered liturgical vestments. Panels are embroidered by a method considered to be typically Italian: gold thread is laid in rows and couched down with coloured silk, the spacing and colour of the silk stitches being arranged to shade the design. The flesh is worked in long and short stitch. During the 15th century, Antonio Pollaiuolo designed for vestments small rectangular pictures that were worked in this fashion by male craftsmen; a series of these relating to the life of John the Baptist is in the Museo dell'Opere del Duomo, Florence. The more generally accepted methods of silk, gold and appliqué were also used through the centuries and are of a very high standard.

White work and tulle embroideries found a place, along with varied coloured work, in peasant costumes. Among the linen methods used were cutwork, drawn thread work and the style known as Assisi work, in which the background of the design is worked solidly in cross-stitch or long-armed cross-stitch (frequently in red silk) and the details of the design are picked out in double running stitch. Formalized birds and leaf designs are common. Cutwork developed into reticella lace in the 15th and 16th centuries and was fashionable trimming. Dresses and other women's garments were also embroidered elaborately, as were men's waistcoats and coats. Florentine canvas embroidery was used for upholstery.

By the 20th century folk embroideries were dying out, but other types of embroidery of a high standard continued to be executed by both professional and amateur craftsmen.

Turkey.—Turkish embroideries are known from the 15th century onward. Their characteristic quality is their richness, achieved by use of colourful silks and gold and silver threads. Women's festal chemises are embroidered with vertical stripes or powderings (small, independent floral sprays or leaves) in gold work; red or purple velvet dresses are similarly trimmed. On the men's large red velvet cloaks, powderings were worked in gold cord. During the 17th and 18th centuries, when the emphasis on decoration reached a climax, even horse harness was heavily decorated with gold and silver embroidery. Among the most important embroidered articles used in domestic life were towels made of linen with sprays and borders embroidered in silk and metal threads. The basic double running stitch was worked so perfectly that both sides of the work were identical.

In the 19th century embroidery declined somewhat, but a revival that began in the 1930s continues to produce excellent work.

Algeria.—Nothing is known of Algerian embroidery prior to the period of Turkish domination (16th–19th centuries), but the work of that period is very fine; it survived the French occupation. The typical design closely resembles the pomegranate or artichoke pattern that is found in Turkey, but it is worked in soft, lightly spun floss silk in the peculiar shade of mauve or violet that distinguishes the work of north Africa from that of the rest of the Mediterranean. The stitches were based on the counted thread of the background fabric, the main part of the design being worked in eyelet stitch with a small portion in satin, brick and double running stitches. In the 19th century herringbone, laid and couching stitches were also introduced. Door curtains were made of three widths of linen, embroidered as individual panels and joined by an uneven number of French silk ribbons sewn together. Headdresses were made of long, narrow strips of linen with an embroidered centre panel, a narrow border at each side of the lappets and a wider border at the ends. Handkerchiefs were embroidered in colours and designs similar to the curtains, sometimes also with narrow bands of white embroidery.

Iberian Peninsula.—Spain, because of the years of Moorish

domination, has been much influenced by oriental styles. The oldest extant embroideries are of the 17th century, but, as the designs changed very slowly, they may well be reproductions of much earlier work. Linen weaving has always been a Spanish handcraft, and linens were embroidered, particularly bedding, pillows and curtains for the marriage chest. A favourite colour scheme, black on white, is said to have been taken by Catherine of Aragon to England, where it developed into the popular Elizabethan black work. In Spain the thread most often used was undyed wool from black sheep, and samplers are an important record of this type of work. Bedspreads, valances and table mats were embroidered with characteristic arrangements of hemmed appliqué, chiefly in salmon pink or black taffeta on a rather heavy linen with borders of drawn thread. Felt appliqué on wool was also worked. In rural districts costumes were carefully designed and executed. Black work was most popular with bright-coloured embroidery second and white work the least favoured.

The Portuguese, as the pioneer explorers of the far east, were the first Europeans to absorb oriental influence into their art. From the end of the 15th century onward imported oriental embroideries, particularly bedspreads, led the Portuguese to develop their own style, and many bedspreads of this period are national treasures. Alentejo is celebrated for its bridal bedspreads. In the district of Arraiolos needlework carpets using long-armed cross-stitch were made; until the 18th century Persian carpet designs were copied, but after the turn of the century a national character was apparent and the colour was generally restricted to blues and yellows. The gayest and most original costumes are found in the rural areas, and as late as the 1960s these were still embroidered in the Minho district. Bridal costumes are of fine black cloth covered in beadwork, and the square white headkerchief is decorated with motifs and border. The islands of the Azores archipelago used the designs of the local china industry, carrying them out in long-and-short embroidery in two shades of blue. From Madeira comes the *broderie Anglaise* white work which is widely known through its large commercial market. (D. N.)

WESTERN EUROPE

France.—The art of embroidery began to develop in France after the crusades. Both Louis XI and Charles VIII were interested in this art, and they summoned Italian embroiderers to France; early French embroidery was thus much influenced by Italian. Most of the early work is seen on vestments and hangings, and much of it was made in monasteries and convents as well as by guilds of embroiderers. Work done by noble ladies was usually under the supervision of the ecclesiastical authorities. After the Reformation there was a tendency to secularization in ornament, even in work intended for ecclesiastical purposes. Thus scrolls and bird forms are used on vestments, usually, however, surrounding sacred monograms. By the 17th century many altar frontals were designed without any religious symbolism.

French needlework of the 16th century is closely related to that of England. Both had a tendency to adapt engraved and other designs for embroidery patterns, and embroiderers fell under the influence of successful illustrators, though they invariably took liberties with the designs, transforming fantastic biblical landscapes into domestic woodlands and concentrating on storytelling elements. Pattern books for embroidery were made in France and distributed throughout Europe.

In the Cluny museum may be seen one of the most remarkable productions of French embroidery of this period, an oval medallion, representing the Hebrews adoring the Golden Calf, the cartoon for which was done by Raphael at the request of Francis I. The historic meeting of Henry VIII and Francis I on the Field of Cloth of Gold in 1520 brought forth much embroidery. Tents, banners, horse trappings and men's costumes were so splendid with embroidery that, according to contemporary writers, they were beyond the power of mortals to describe. Francis I was clothed in gold and silver damask embroidered in coloured silks. His horse wore a mantle of blue and gold tissue embroidered with fleur-de-lis in trellis work.

Secular embroidery was at first a luxury set aside for persons of rank. Deep-piled velvets, rich brocades and heavy silks were embroidered with gold thread and coloured silks. Quasi-embroidery effects were obtained by sewing cords, heavy bullion, spangles, tinsel and materials on to fabrics. Such "doges" as outline patterns or gold thread tracery have been depicted by Holbein and Bronzino in their portraits of European rulers, nobles and other distinguished persons. In men's clothes black velvet, with threads of gold tracery touched here and there with bits of colour, was much admired. An equestrian portrait of Henry II by François Clouet shows the king wearing a black velvet doublet with gold embroidery; the most interesting details are the elaborately embroidered white boothose showing above his jack boots. Women's long stomachers were encrusted with jewels and embroidery.

A magnificent cope (in the Metropolitan Museum of Art, New York city), which according to tradition was given to the cathedral of Metz by Anne of Austria, is embroidered all over with flowers and foliage sprawling across a silver ground; the metal threads are laid in whirling curves and spirals. The general effect is one of unrestrained exuberance. This quality in embroidery was carried to such an extreme in both men's and women's dress that in 1629 Louis XIII issued an edict forbidding its use. Though embroidery continued, white work and laces began to take the place of the earlier, more sumptuous work.

Louis XIV, unlike his predecessor, encouraged the art, and embroidered furniture hangings as well as costumes received his support. A set of valances and curtains in the Metropolitan Museum of Art (dated about 1700) are embroidered in silk and wool on canvas in tent stitch of two sizes. The ground is black and the design a blaze of vivid colours with red, blue and white predominating; the medallions, framed in red, show would-be classical figures. In the interstices of the brilliant foliage that surrounds the mythological scenes are little beasts and birds, some representing the fables of Aesop and La Fontaine, some, such as a fox in a pulpit preaching to birds, reverting to medieval themes. The design is symmetrical, but the riotous colour has been applied without any thought of symmetry. A later set of curtains, worked in a large cross-stitch in wool and small tent stitch in silk, show a design that at first sight appears to be haphazard but that on close inspection is seen to be fairly symmetrical. The figures are Chinese, examples of the style known as *chinoiserie*, and the ground is of brilliant yellow silk, the embroidery colours chiefly blue, green and red.

During the reign of Louis XV, style rejected symmetry, and in the change from the baroque to the rococo there was a beautiful disorder and wild asymmetry; the designs became gay, frivolous and gracious. After the king's death, Madame de Maintenon established a school for girls at St. Cyr, where she devoted much time to instruction in needlework. Still hanging in the palace at Versailles are elaborate pieces executed by these girls. Many of the treasured 17th-century embroideries of the church and palace were destroyed during the French Revolution, when the embroiderers were ordered to prick out the gold and silver threads.

In the 18th century elaborate patterns gradually gave way to dainty design. Design in that century was much influenced by China. The superiority of French work became so widely recognized that there was a demand for it in all other European countries. Madame de Pompadour set the fashion for tambour work.

After the Revolution embroidery all but vanished from men's clothes, except for short embroidered vests, but women wore sheer fabrics embroidered with scattered flowers, and during the consulate hemlines were bordered in polychrome silks. At the beginning of the 19th century the beautiful *kasida* embroidered muslins of Dacca (*see* below, *India*) became the vogue for women's Empire-style dresses in France, England and America. From the mid-19th century, embroidery in women's clothes was largely confined to the underlinens.

Germany.—German embroidery of the 10th and 11th centuries showed marked Byzantine influence and was done for ecclesiastical purposes. A distinctly German feature was the treatment of draperies of figures as flat surfaces covered with diaper patterns. From the Reformation on, embroidery was done for secular uses. Later embroidery design showed a preference for

bold floral patterns, often mingled with heraldic emblems. Crewel (worsted) embroidery, which became popular in England in the late 17th century, received its impetus from Germany, where there was a great fondness for wool embroideries for domestic purposes.

During the later part of the 18th century Saxony became outstanding for its white embroidery. The daintiest of muslins, cambrics and linens, embroidered in white cotton, were used for collars, cuffs, scarfs, handkerchiefs and other accessories. Marie Antoinette chose this delicate and beautiful embroidery for her own wardrobe and thus set the fashion for the world.

Berlin woolwork, brilliant worsteds on canvas, became popular during the Victorian era. The craze, which developed about 1856 and spread from Europe to America, was the most important needlework fashion for the next 30 years. Motifs used were pet dogs, lions, deer, parrots perched on wreaths, sprays or bouquets of roses. (*See* below, *British Isles: England* and *United States.*)

Low Countries.—In the 14th and 15th centuries Bruges was to the north of Europe what Florence and Venice were to the south. The luxury of the courts of the counts of Flanders and dukes of Burgundy encouraged manufacture and handicrafts of all kinds. It was during this period that embroidery lost its distinctly religious character and came into common use in lay dress. Paintings by the Van Eycks and Hans Memling show the embroidery of the period in scrupulous detail. The Virgin in Hubert van Eyck's "Adoration of the Sacred Lamb" wears a magnificent robe, the hem of which is embroidered in gold, pearls and jewels; Mary of Burgundy lies wrapped in an almost identical state mantle in her brass effigy on the altar tomb at Notre Dame in Bruges.

A Flemish triptych of the early 16th century, showing "Scenes of the Childhood of Christ," was worked on linen in gold thread and coloured silk with translucent, split, cross and brick stitches and couching. The design was based on a painting by an Antwerp mannerist of about 1510, possibly a lost painting by Hugo van der Goes. The technique, with translucent stitch prevailing, was introduced by Burgundian embroiderers for the vestments of the Order of the Golden Fleece. This technique permits the ground of couched gold to remain visible, producing an effect similar to that of the translucent enamels of the period, which reveal their precious metal ground. Over gold thread, coloured silks are applied at right angles; faces and hands, reserved in gold, are covered with split and satin stitches. In some ecclesiastical embroideries there is an effect of shot gold, obtained by the grounding being entirely of double gold thread sewn down with silks of different colours. This laid gold grounding originated in Flanders, from where it spread to Italy.

Another example of fine Flemish embroidery of about a century later is a hood for a cope with a representation of the martyrdom of St. Margaret. Above the figures is embroidered a canopy of late Gothic architectural ornament in couched gold threads. The figures are worked with coloured silks in long and short stitches. Parts of the costumes are worked in gold threads and fine coloured silks, with seed pearls worked into the edgings. This is an important example, for it is typical of the most elaborate finished picture embroidery that was produced in Europe. Great quantities of this type of embroidery were done for church purposes, and it occupies a position in embroidery comparable to that of the fine 15th- and early 16th-century Flemish tapestries in tapestry making.

During the 16th century most of the continental embroidery designs were of Flemish origin.

Scandinavia.—Norwegian embroidered picture tapestry dates from Romanesque times, and ornamental embroidery on clothing and household linens has been popular throughout the ages. The Renaissance stimulated embroidery into a new growth with the Renaissance geometric (number) stitch, which is called the *tellesøm* and has been used in Vestland up to modern times. In Trondelag and Ostland, the freer forms of the rococo style were adopted. Wool embroidery freely executed was a characteristic of peasant dresses, though some localities used silk. Colours were mostly homemade, extracted from local plants; bright red and blue dyes had to be imported. Using *fargesøm*, a colour stitch built up of small stitches resembling commas and added one over another in a regular fashion, it was easy to create a design without a pattern. Sometimes *fargessøm* was combined with bead embroidery.

Rennesøm, running stitch—better known as *Holbeinsøm* after the German painter Hans Holbein the younger, whose paintings of the Virgin Mary often show similar embroidery—is done with continuous and diagonal stitches so that only the alternate stitches are visible, the next thread being darned in the available opening so as to form a solid line. While it is mostly done in black, it is occasionally accented with colours. *Hardangersøm*, Hardanger stitch, is a composite of several types of stitches, the most characteristic being the "cut out," the open areas of which are filled in with *nettsøm*, netting stitch. This method undoubtedly came to Norway from Italy about 1700.

In Denmark there developed embroidery called Danish *hedebo*, which is done only on linen and used for men's shirts, women's underclothes and home decoration. Motifs used in the embroidery were taken from Renaissance furniture and from scrollwork in old books. It consisted mostly of drawn work surrounded by flowers and other designs worked in chain stitch. In the 19th century there was a change, and figures were represented by drawing a number of threads at intervals, lengthwise and crosswise, forming square holes. The remaining threads between these open spaces were worked in ordinary darning stitch. By the middle of the century embroiderers broke away from established forms and began to follow new and freer lines. The motifs now were leaves, stars, lotus flowers and figures taken from the pattern of Paisley shawls. At the end of the century embroidery ceased to be a home product and so lost its originality and beauty.

CENTRAL AND EASTERN EUROPE

Embroidery in central and eastern Europe is largely a folk art. Embroideries are used for decorating national costumes and for decorating the house, where embroidered sheets, valances, pillows and towels are found. Generally these embroideries are very colourful, the designs being made up of geometrical and floral motifs.

Poland.—Wool coats are finished with decorative stitches in colourful threads over the seams. White blouses are embroidered on the shoulders, at the neck and at the edge of the sleeves.

Hungary.—Hungarian national costumes are totally unlike those of any other country. White is the ground colour, and most interesting is the long coat of white felt, abundantly decorated with cotton or silk embroidery or appliqué. The favourite motifs are carnations, roses and tulips.

Rumania.—The Rumanians use both embroidery and leather appliqué in decorating the costumes worn for festival occasions. The favourite designs are geometric, worked in cross-stitch, with the addition of beads and sequins.

Bulgaria.—In Bulgaria, as in Hungary, white is the basis of the dress, but in the colourful Bulgarian embroideries red is the dominant colour. Men's trousers, shirts and coats and women's chemises, dresses, and coats are embroidered, usually with red, yellow and black as the dominant colours. (M. S. Ve.)

Czechoslovakia.—The beautiful embroidered peasant costumes of Bohemia, Moravia, Silesia and Slovakia gradually went out of use during the 19th and 20th centuries, but much is being done to revive the national folk art in all these provinces. Dance groups are encouraged to wear traditional form of dress and craft centres have been established. Throughout the centuries the embroideries show many variations of style as a result of contact with the fashions of the upper classes, which were adapted by the taste and needs of the villager to the styles of the past. With the introduction of machine-made fabrics, finer in texture than handwoven materials, geometric patterns gave way to natural (*e.g.*, floral) forms. Symbolism used at least as early as the 14th century survived in the conventional interpretation of animals such as the cock, hen and stag and also of the human figure. Colours changed with the centuries. The earliest threads were natural flax, later dyed scarlet; black embroideries date from the Renaissance and multicoloured work from the baroque period. Later, colours sometimes became typical of districts; for example, pale blue silk embroidery on fine linen was typically Bohemian.

Tailor-made outer garments were embroidered by the tailor, who worked the design freehand. Richly developed floral designs,

braiding, gold thread padded by stiff-card (cut to shape of design, tacked in place and then completely worked over and covered by gold thread) and raised work on buckskins all came within his scope. Sheepskin coats, with the wool on the inside, were decorated on the smooth side with multicoloured leather appliqué decorations. Soft washable embroideries, including childbed curtains (drawn to give privacy to the mother and new baby), bridal shawls and apron kerchiefs were made by women. Lacy effects were obtained by cutwork similar in style to Madeira work, the cut shapes sometimes being filled with needle-made lace. Tulle came into use in the mid-19th century, and various districts devised ways of embroidering this material with darning stitch, appliqué and even cut work and drawn thread fillings.

Yugoslavia.—The national costumes of the southern Slavs are very ancient in origin; a small figure dating from the Iron Age of the middle Danube is carved to represent an embroidered dress which already shows some of the principal characteristics of the southern Slav costume. In general the designs are purely geometric or of formalized flowers and animals into which ancient symbols, such as the cross or crescent, are introduced. The colours, as is usual in a peasant craft, were derived from vegetable dyes giving yellow, reddish-brown and blue. The yarn is a hard thin cord made from unwashed fleece; embroidering with this is extremely hard work, but the result is almost indestructible. Wool from the very common black sheep is used for the black embroidery, giving interesting variations of black. Gold is used in the south and in Slovenia. In modern times the braid is often manufactured, but thinly beaten bands of brass of up to $\frac{1}{4}$ in. width are frequently used.

The southern Slavs used red as a dominant colour; among the Dalmatians the prevailing colour is indigo. The background material is either bleached or natural linen (hemp in the northern areas). In regions such as Montenegro, influenced by the Turks, the background fabrics are richer—pale blue cloth and purple, mauve or black velvet. Croation and Serbian women wear a variety of caps, bonnets and headkerchiefs, always embroidered.

As a rule each district employs one basic stitch, supplemented by others. The geometric designs are worked in slanting Slav stitch (this forms the diagonal to an oblong three threads by one thread). Stroke, chain and cross-stitches decorate the edges. Sometimes darning and satin stitches are used. On the island of Pag lace trimming has remained as an independent craft at the transitional stage between needlework trimming and lace. (D. N.)

Russia.—Russian art and life were dominated by Byzantine culture, and the vestments of the Russian Orthodox Church and the long, full gowns of richly patterned fabrics and gem-studded embroideries, the elaborately conceived headdresses of pearls and precious stones of the former court dress were all inspired by Byzantine modes. Existing examples of 18th- and 19th-century liturgical vestments are as luxurious as weaving and embroidery can make them, and the heavy gold-laden materials are reminiscent of the richly ornamented figures in Byzantine mosaics.

National Russian embroidered ornamentation dates from remote times and was used to decorate sheets, aprons and towels. Towels were used from antiquity as religious accessories, for solemn ceremonies and for family festivals and carnivals; in later times they were ornamented with the cross and with images of saints and were often given as offerings to the church. Incorporated in their designs were fantastic flowers and foliage more or less stylized; human beings with arms and hands raised in a sign of adoration; symbolic animals such as lions, unicorns, horses, stags, peacocks; stars with eight points and roses with eight petals; the swastika; and fabulous birds named Sirine and Alcanoste who, according to legend, lived in a Russian pagan paradise before the appearance of Christianity. Later embroidery patterns included landscapes and scenes of festivals and processions. Very little that is representative of Russian life is seen in these designs. They derived from manuscripts of the 13th and 14th centuries. Such designs were worked in coloured silks, gold and silver threads with paillettes and precious stones on silk or linen. The manner of execution varied, but the cross-stitch and drawn work were most popular.

In the Ukraine colour is very much more developed, mystic figures disappear and the decoration forms flowers and foliage more or less stylized. In the Tatar regions, colour plays a great role on gloves, purses and velvet caps, often enriched with gold thread embroidery. In Armenia and the Crimea the tones are softer and more harmonious, the designs more oriental. They are executed in silk on white cloth with open-work embroidery; with the drawn work are slender foliated lines of silver and gold.

Russian boots for women must be mentioned, for they are famous for their colourful leather appliqué work. (M. S. Ve.)

BRITISH ISLES

England.—The earliest example of embroidery in the British Isles, dating from A.D. 903, is the stole and maniple found in the tomb of St. Cuthbert at Durham (now in the library at Durham cathedral). They are worked in very fine stem stitch, split stitch and couching, in red silk over a very fine gold thread. The Bayeux tapestry (embroidered work, not true tapestry) is the only large hanging of the period to survive (*see* BAYEUX TAPESTRY).

The 13th century was an outstanding period in the history of English embroidery. Vestments and church furnishings had backgrounds entirely embroidered with couching, in which a top thread is pulled through to the back by tightening the tie-down thread, giving the appearance of very fine satin stitch. In representations of human beings the hair and line details were worked in split stitch, and the faces in spiral stitch were very decorative. Jewels and pearls were used, though many were subsequently stolen. The designs, similar to those seen in the manuscripts and wall paintings of the period, fall into three main groups: (1) the tree of Jesse, with scrolling stems of either oak or vine leaves, which dates from the late 12th and early 13th century; (2) geometric patterns with repeating geometrical panels, dating from the mid-13th century; (3) architectural designs, usually early 14th century. Figures, often with blue and green striped hair, beards and shaven upper lip, were arranged to fill the shapes of the basic design.

In the 14th century embroideries were also worked for secular purposes although there was a decline in the second half of the century. The 15th saw a revival when velvets and damasks were introduced from the continent and embroidery was adapted to these backgrounds. The solid part of the design was first worked on linen and the linen stitched to the background; the lighter details were worked directly on the velvet or damask. A perfect set of altar hangings of this style is in the parish church at Chipping Campden, Gloucestershire.

The abolition of the monasteries in the 16th century brought much destruction to the treasures of the church; embroideries were cut up for household articles, their jewels being stolen and gold melted down. Only a few have been preserved, through extraordinary circumstances—for example, the nuns of the convent of Syon took the Syon cope with them when they fled to Flanders. A period followed when household furnishings such as panels and cushions were embroidered. By the reign of Elizabeth I. embroidery had once again reached a very high standard. The publication of books, particularly herbals, inspired designs which included the typical English plants set in a scroll arrangement of stems and worked either in coloured silks on linen or silk, or in black on linen. The latter was carried out in fine-counted geometric fillings enriched with gilt stems and a lightly spangled background. Caps, bodices, dresses and tunics were decorated, as well as such household articles as pillowcases and coverlets. Canvas work became popular and replaced tapestry for wall panels, bags and book bindings. Table carpets, either in tent stitch (petit point) or in a rug stitch to give pile, are often depicted in portraits of the period.

During the 17th century three new styles of work were introduced. Jacobean woolwork produced large decorative hangings, curtains and bedspreads, whose style of design (large trees with exotic leaves and birds) is described as *chinoiserie*. The colours were monochrome blue, red or green at first, later developing into bright and more varied colours. A variety of stitches, including chain and knotted, produced open laid fillings and heavy shaded shapes. Samplers were introduced to record stitches and patterns, but printed embroidery books were known as early as 1631 though not yet in general use. One of the earliest extant samplers is dated

1596. The third new type was stump work, which decorated mirror frames, caskets (often made with secret drawers) and small wall panels. The designs were in high relief, frequently depicting biblical scenes with the characters in contemporary dress, and were executed in detached buttonhole stitches. Metal and various types of silk thread were used.

In the next century there was a reaction against heavy woolwork, and designs became more delicate, with an increasing tendency to copy nature. The background of bedspreads, cushions and some clothing (such as waistcoats) was quilted all over and provided a rich setting for the small floral sprays embroidered in bright colours. Silk-embroidered aprons worked in coloured silks and gold thread were fashionable. Samplers became more square in shape and were used as pictures rather than patterns. The great advances in furniture design (*see* FURNITURE DESIGN) gave scope for upholstery both in tent stitch and long-and-short on satin.

Amateur embroidery was greatly influenced early in the 19th century by hand-painted charts, imported from Germany, from which cross-stitch patterns could be worked in a very soft embroidery wool, spun at Gotha and dyed in Berlin where the charts were printed and painted; the style was called Berlin woolwork. The first chart was issued in 1804 and between 1810 and 1840 no fewer than 14,000 different designs were published; by the latter year they had superseded practically all other embroidery. In the second half of the century church embroidery was revived, due to the formation of societies interested in introducing the use of vestments into the Anglican service. In 1861 William Morris opened his interior furnishings shop, whose products included embroidery (*see* ARTS AND CRAFTS MOVEMENT). The embroidery designs were carried out by his daughter May Morris and others, and also by the Royal School of Needlework, which was founded in 1872.

Scotland.—Ayrshire white work, begun by an Italian embroiderer who arrived in Edinburgh in 1782, was a special Scottish industry. There was at that time a great demand for lacy embroideries to replace the imported and heavily taxed lace from the continent, and this craftsman was given an annual grant to train workers in the art of fine muslin embroidery. By the middle of the 19th century it was estimated that 500,000 workers were employed, including a number in a branch of the industry that had developed in Belfast, Ire. In white work, satin stitch is used with contrasting openwork patterns; the designs are floral, and those who embroidered them were known as "flowerers." Many beautiful examples of christening robes and bonnets of the early 19th century may be seen. The decline in the second half of the 19th century was caused by difficulties in obtaining raw cotton from abroad, largely due to the American Civil War, and to the change in fashion to the bustle style, which required heavier materials.

Most other Scottish embroidery followed the trends seen in England. (D. N.)

Ireland.—Embroidery, called *druinech,* was practised by women in early Ireland as an art of trade. The designs were drawn on leather and the embroiderer copied the leather patterns. Embroidery was a well-recognized art from the earliest period of legend. Irish white embroidery, of later times, similar to the white embroidery of Saxony, was usually done on fine cambric, linen or muslin and the stitches used were flat and raised satin, French knots, overcast and dot. An example found in the Metropolitan Museum of Art, a handkerchief of the 19th century, has a heavy border of floral sprays embroidered with lace stitches at regular intervals connected by a design of detached flowers and hemstitching. The name of Limerick is applied in Ireland to tambour, run and appliqué work. (M. S. VE.)

UNITED STATES

Embroidery was practised in America from the time of the early settlers, who brought the art from their homelands. Pioneer women were handicapped by the lack of familiar materials, and it was necessary for them to produce all supplies by hand. Woolen embroidery yarns (from the colonists' own sheep) were skeined, dyed and cut into convenient lengths for use in crewelwork. A great variety of colour was obtained by the use of vegetable dyes, but blue was commonest, derived from home-grown indigo and tinted various shades depending upon the length of time of immersion. Linen yarn was also used for embroidery among the early settlers, but with the perfection of spinning machinery, mechanically spun cotton gradually replaced the hand-spun linen.

Elaborate crewel embroidery was fashionable in England in the 17th century, and American women followed this style, though American designs were necessarily much simpler than the English and the stitches were designed to show most of the crewel on the surface of the material. Turkey work for cushions and upholstery was also brought to the colonies. This was done by knotting wool threads on linen canvas, thus producing a pile similar to that of oriental carpets, and filling in the background spaces with cross- or other stitches.

Quilting, as practised in France and England, was done from an early period; as materials became more plentiful the appliquéd quilt was evolved in which patterns were cut from coloured fabrics and applied to a foundation with decorative stitches. This style was conducive to a wide use of embroidery to enhance the pattern. A late-19th-century development was the silk crazy quilt, made by piecing together silk scraps in haphazard fashion, covering the seams with fancy stitches and embroidering on the pieces. Materials for crazy quilts, sold commercially by American textile manufacturers, included not only the silk pieces but also designs and instructions.

As in England, a popular use of early American embroidery was in the making of samplers which, besides being decorative, were a means of teaching little girls their stitches, alphabet, numbers and often Bible verses. The execution of a sampler was a necessary part of every girl's education, and samplers came to be a sort of final examination or sign of outstanding achievement in girls' schools, where they were exhibited upon graduation. Needlework pictures of all kinds were made in 18th- and 19th-century America, in wool or wool and silk on canvas in cross-stitch, tent stitch and other canvas stitches. Scenic, classical, pastoral or biblical designs were often taken from contemporary prints and engravings. With the growing availability of materials, pictures were embroidered on thin silk or satin with silk, metal thread and chenille. The needlewoman of the early 19th century was greatly influenced by the style of contemporary painters, particularly Angelica Kauffmann, whose work was greatly admired and copied. Sometimes embroidery was supplemented by the addition of water colour, thereby conserving both material and labour.

The Germans who settled in Pennsylvania continued their special traditions of needlework. A characteristic form of Pennsylvania Dutch embroidery was the long, narrow, linen "door towel," on which were worked decorative bands of traditional designs in fancy stitches and, generally, the name of the maker and the date. These towels were for decoration only, being used to cover ordinary towels in the guest room or as a hanging on the door or wall.

The most characteristic style of embroidery during the 19th century was Berlin woolwork, which was done for every conceivable use in interior decoration and costume accessories (*see* above, *British Isles: England*). The wool for this work was dyed brilliant colours in the German taste, and the bright colours and versatility of the embroidery led to its widespread popularity. Besides the usual cross-stitch and petit point used in canvas embroidery, a raised or clipped stitch was employed which created a thick wool pile and enhanced the colour and shading of floral designs. Coloured glass beads were also introduced to accent the floral and scenic patterns. Finished with heavy silk cords, tassels, ribbons, bows and fringes, these embroideries were a colourful addition to 19th-century interiors. Berlin woolwork was also done on cardboard, known as "Bristol boards," perforated or punched with minute holes at regular distances, that could be bought in whole or part sheets and embroidered with silks or fine wools. This became fashionable for the occupation of leisure hours and was considered good work for children. Mottoes, simple designs and names were worked for bookmarks and Valentines and other greeting cards.

The popularity of embroidery extended to costumes and accessories. White muslin of every weight and kind was embroidered with cotton or silk thread and used for dresses, ruffles, flounces, handkerchiefs, shawls and underwear. Patterns for such work

were issued in most women's magazines. Hand-run net was also used extensively for costume purposes from early colonial times; in its production, net was stretched on a frame and the pattern darned in with silk or cotton thread.

Toward the end of the 19th century, the influence of William Morris had a revolutionary effect on all forms of arts and crafts, including fancy needlework. This resulted in the popularity of "art needlework," embroidery and appliqué on coarse linen crash for wall panels, hangings and portieres. The widespread appeal of this work established the fashion for ecru linen or crash, which practically usurped all other foundations for decorative embroidery. The late 19th century brought about many revivals and innovations in needlework design. Patterns were executed in "Gothic," "Renaissance," "Empire" and "oriental" styles. Kate Greenaway illustrations were transformed into embroidery patterns and applied to table linen and crazy quilts to be worked in simple outline stitch. Monogramming of household linen became fashionable at the close of the century.

In the early 20th century, following a growing trend for naturalistic design, shaded silk embroidery developed. This was worked largely on white linen and showed floral sprays, garlands, wreaths, birds and butterflies embroidered in fine silks in flat satin stitch, its delicate shadings rivaling the originals in colouring.

Many needlework groups were formed in the latter part of the 19th century to further the interest and practice of needlework, and since that time such organizations have kept its popularity alive for decoration and pastime. (M. Da.)

ASIA

Iran.—Embroidery in Iran was used widely for the decoration of prayer and bath carpets, as well as for robes, hangings, curtains, table covers and portieres. Persian embroiderers showed a decided preference for floral patterns. Though human and animal forms are sometimes seen—men engaged in hawking, hunting or feasting—the human theme is uncommon in Islamic art and birds are the only living creature frequently represented. Persian embroiderers compensated for this by the richness of their floral ornament. Naturalistic flowers are rendered charmingly, but most characteristic is the use of conventional plant motifs. An example is the radial medallion, which is nowhere rendered more splendidly than in Persia (*see* also Rugs and Carpets). The tree of life motif, used as early as the 7th century A.D. on rock reliefs in Persia, spread to other countries principally through the importation of Persian embroideries.

An important development of Iranian embroidery is the varied and subtle use of texture, which may be its greatest contribution to the art of embroidery. Texture—emphasis on which is characteristic of Persian design in various mediums—is obtained in embroidery by use of three different techniques: darning, couching and *filo tirato* (drawn work).

No known Persian embroideries can be dated before the 16th century. Marco Polo, however, told of seeing embroideries in eastern Persia in the 13th century, and his description of the designs he saw could apply as well to the work of later centuries. The embroidery patterns undoubtedly were very similar to those of the woven silks. In early examples the entire ground is covered by the design, while in later work the ground colour plays a part equally with the varied silks of the needlework. There is also an obvious similarity between Persian embroideries and carpets, the same angularity being seen in the outlines, the same detached motifs and the same stylized animal forms.

A large group of covers that date from the 17th to the 19th century have very heterogeneous patterns worked in coloured silks and metal thread on a ground of satin or silk. Unfortunately, these are often poor in execution and are not very interesting, but they are the commonest type of Persian embroidery, having been made all over the country. Another smaller group, which has much more interest, comprises covers, prayer mats and bath rugs made with floral patterns in coloured silks on a white ground. Chain stitch is frequently used and the prevailing colours are crimson and green. It was this class of embroidery that reached India and came to form the basis of many Indian printed designs; with the introduction of Indian textiles into England, many 18th-century coverlets and curtains made in that country assimilated these originally Persian patterns. This work undoubtedly originated in the southern part of Persia and has little affinity with work done in the north; the patterns are naturalistic rather than stylized. The central floral medallion is one of the commonest patterns of this group of embroidery. Around it is arranged an intersecting network of flowering plants, which in turn is enclosed with a rectangular border consisting of a running pattern of similar flowers. This pattern is fairly close to that of the so-called Polish carpets, which were thought to have been made in Poland but which, it has come to be assumed, must have been made in Persia or an adjoining region.

Up to the end of the 19th century women's trousers were decorated with elaborate embroidered patterns whose designs were always of diagonal parallel bands filled with floral ornamentation. Originally the trousers were made of woven fabrics of elaborate design, and the embroidery work was based on these original woven designs.

Embroidery has ceased to be important in Iran; indeed, in order to study the beautiful Persian embroideries of the past it is necessary to look in collections outside Iran. (M. S. Ve.)

China.—Chinese embroidery has been influenced at various times by its neighbours and the countries with which it traded, particularly Persia, but the most powerful influence has been sericulture, which for centuries equipped the worker with beautiful materials and enabled him to produce exquisite work. Among the earliest examples are some found in eastern Turkistan dating from the T'ang dynasty (618–906). The dating of early embroideries is difficult owing to the conservatism of the Chinese craftsmen both in choice of pattern and style, but after the 16th century, and particularly from the 18th century onward, more accurate information is available.

The main purpose of embroidery was adornment; the emperor wore on his robes symbols known as the 12 ornaments, which included the sun with the three-legged bird, the moon with the hare pounding the drug of immortality, the stars and mountains, etc. Civil and military officials wore other symbols. Decorative arrangements of writing were introduced into designs to convey a message, such as "double happiness." The embroidery is very fine and the technique perfect. Void satin stitch, in which a narrow strip of background material shows between the bands of satin stitch, is typically Chinese, as is Peking knot, often used to carry out a whole design. Gold and silver threads were made by covering silk with fine, untarnishable gold or silver paper which was couched in rows. Floral design, often representing the seasons, was used on many articles such as cushion covers, hangings and fan cases, and colours are varied, clear and bright, though not brilliant. The silk backgrounds were often in rich, dark colours or black. Summer robes were worked in Florentine stitch, which was counted on a fine muslin type of canvas, giving an entirely different style of embroidery. The making up of the embroideries was completed with cords and tassels. (D. N.)

Japan.—Japanese embroidery work follows that of China but is more pictorial and fanciful. The decorative patterns created for women's kimonos in the 17th and 18th centuries constitute a unique development that is more closely related to classic traditions than are the prints and paintings of the same period.

In the mid-17th century patterns were typically large scale and open, and featured simple abstract motifs and bold flowing lines. Late in the century the designs blossomed out with charming combinations of flowers with books, boats, puzzle characters, winding streams, etc. Embroidery was done in coloured silks and gold thread with untwisted silk threads generally laid in long soft stitches overlaid with wehlike filaments crossing one another on the diagonal. This, called *kiriosa* in Japan, is like the western split stitch. On leaves and flowers, veins and petals are indicated by couching threads backstitched. The design is first drawn with black ink and the embroidery stitches are worked over the lines.

Flowers were a favourite motif, and embroiderers used a variety of them, representing the four seasons: plum, cherry, violet, pawlownia, camellia, bell-flower, clover, chrysanthemum. Reddened maple leaves and snow-laden pines, bamboo and grasses were also

typical design elements. One outstanding feature of Japanese design is the spacious distribution of motifs, which lends a note of calm to the general impression. There is a feeling of restraint and yet of bold decorative quality.

As a result of sumptuary laws that went into effect early in the 18th century, patterns became temporarily smaller in scale and featured fairly simple renderings of birds, snow, pines and flowers. In the development of 19th-century design Japanese art became more realistic and less abstract and subtle.

The swatches of fabric used for gift wrapping (*fukusa*) are exquisitely embroidered, as are small objects such as pocketbooks, perfume bags and comb cases. Sashes and ribbons are ornamented with floral motifs embroidered in coloured silks. (M. S. Ve.)

India.—Embroidery has a long history in India; bronze needles of the 3rd millennium B.C., excavated at Mohenjo-Daro, were probably used for this purpose, and among figurines from this and other Indus valley sites there are a number with drapery that is apparently embroidered. Like other forms of textile decoration, Indian embroidery received much stimulus from Persia after the 16th century A.D., particularly under the patronage of the Mogul emperors, who encouraged the settlement of Persian craftsmen. In many parts of India there flourished also a vigorous tradition of peasant embroidery, the styles varying from region to region. Especially notable are the embroideries of Kutch and Kathiawar, worked in chain stitch on a satin ground and usually inset with mirror glass. Even here the sprigged patterns and floral diapers show strong Persian influence. In contrast to these, the more geometrical patterns of the *phulkari* embroideries of the Punjab (worn by women there and used in the west as wall hangings) represent what may be regarded as a more basically indigenous or pre-Mogul style. *Phulkari* work is executed in darning stitch, done from the back; in good examples the stitches are very close and short, and the regularity of the design is secured by counting. Another type is carried out in cross-stitch, with effects like those of European and American samplers. The most famous muslins of the Ganges valley were woven at Dacca, where the humidity was an advantage in the handling of delicate skeins. Those figured on the loom by a brocading technique were known as *jamdanis;* those embroidered with a needle in white cotton (in darning and satin stitch, often in combination with drawn-thread or net embroidery work) were called *chikan* work, those embroidered in gold-coloured silk (also in darning and satin stitch) were called *kasida.* Gold and silver embroidery was widely used throughout India for ceremonial purposes and in combination with a wide variety of materials, especially velvet, very rich effects being obtained in this way. (J. C. I.)

MODERN TRENDS

The introduction of machinery for weaving, lacemaking and embroidery and finally, in the mid-19th century, the advent of the sewing machine led to a great interest in the different varieties of material, while that in decoration declined. After World War II, however, a great revival of interest in needlework took place. Twentieth-century embroidery follows the general trend toward simplification of shape and decoration in design. Many countries are consciously adapting their traditional designs to the demands of a period in which labour is expensive and the amateur embroiderer has few free hours to devote to her work. At the same time, opportunities for both professionals and amateurs to develop styles sympathetic to modern architecture have been provided by the many churches built since World War II. Man-made materials such as nylon, Dacron (terylene) and untarnishable threads made from lurex have allowed a fresh approach to established technical methods. Examples of this are to be found in the use of gold and silver thread on washable materials, and in the delicate effects achieved with the transparency of nylon.

Important among the organizations that encourage embroidery are the Royal School of Needlework, London, founded 1872; the Embroiderers guild; the Needlework and Textile guild, Chicago, Ill. (1927); the Needle and Bobbin club, New York (1916); and the Royal Guild of Copenhagen (Denmark). *See* also BEADS; LACE; QUILLWORK; QUILTING; TAPESTRY. (D. N.)

BIBLIOGRAPHY.—M. Symonds and L. Preece, *Needlework Through the Ages* (1928); Barbara Snook, *English Historical Embroidery* (1960); A. G. I. Christie, *English Medieval Embroidery* (1938); Louisa F. Pesel, *Historical Designs for Embroidery* (1956); Kathleen Mann, *Design From Peasant Art* (1952); Kay Kohler, *Embroidery Designer's Sketchbook*, 2nd ed. (1953); Yvonne Hackenbroch (ed.), *English and Other Needlework*, collection of Irwin Untermeyer (1960), Pauline Johnstone, *Greek Island Embroidery* (1961); Eivor Fishes, *Swedish Embroidery* (1953); Frances Little, *Early American Textiles* (1941); Ethel S. Bolton and Eva I. Coe, *American Samplers* (1921); Georgiana Brown Harbeson, *American Needlework* (1938); Donald Key, *Samplers* (1960); Mariska Karasz, *Adventures in Stitches* (1949); Mary Hogarth, *Modern Embroidery* (1938); H. Hanpt-Battaglia, *Let's Embroider* (1960); Phyllis Ackerman, "Embroidery in Persia," *Embroidery*, vol. iii, no. 1 (1934); Lady M. Alford, *Needlework as Art* (1886); Halfdan Arneberg, *Norwegian Peasant Art* (1949); Alan S. Cole, *Ornament in European Silks* (1899) and "Some Aspects of Ancient and Modern Embroidery," *Journal of the Society of Arts*, 53:956–967 (1905); Helen C. Gunsaulus, *Japanese Textiles* (1941); George Leland Hunter, *Decorative Textiles* (1918); Rose G. Kingsley, "Flemish Embroideries Old and New," *The Art Journal*, 1887:377–380 (1887); A. U. Pope, *Survey of Persian Art* (1935); *Embroidery*, Embroiderers Guild, quarterly.

EMBRYO is sometimes loosely used to refer to the developing young of any organism, but more strictly it refers to the new individual produced by the fertilization (*q.v.*) of an egg by a sperm in animals or by the equivalent union of a microgamete ("male gamete") and a megagamete ("female gamete") among plants. It includes all the developing stages up to the time of hatching or escape from the embryonic membrane, as the case may be.

Plant.—In seed plants the embryo is always part of the seed, arising from the zygote, which resulted from the union of the egg nucleus of the embryo sac and one of the sperm nuclei of the pollen tube. It consists of four parts: the plumule (a short shoot); one or more cotyledons (seed leaves); the hypocotyl, a short section between the attachment of the cotyledons; and the rudimentary root, which is the fourth part. The seed coats and the endosperm, the remaining parts of the seed, are derived respectively from the integuments of the ovule and the fusion of the two polar nuclei of the embryo sac and one sperm nucleus with the surrounding cytoplasm. (*See* ANGIOSPERMS: *Characteristics.*)

Animal.—An embryo comes into existence as a result of sexual reproduction, but both uniparental sexual processes (*see* PARTHENOGENESIS) and biparental (in which fertilization by a sperm occurs) produce embryos. In the development of many animals, there is a close relationship between asexual and sexual reproduction. In the loose sense referred to above, the word embryo includes the young produced by both modes of reproduction, and cell division, growth and differentiation are involved in both, a fact that sometimes entails confusion regarding the meaning of the term. Alternation of sexual and asexual generations, or metagenesis, often causes further confusion. (*See* REPRODUCTION.)

Developments of the Embryo.—An embryo begins its development when the egg is fertilized by the sperm. Fertilization is, therefore, one of the most important steps in its history. After the union of egg and sperm, the fertilized egg begins the process of cleavage or segmentation, by which it is transformed from a one-celled individual into a many-celled one. The materials that were contained in the egg undergo redistribution and segregation into appropriate cells. The processes of cleavage are surprisingly regular, especially with regard to the polarity and symmetry of the original egg. The influences responsible for the differentiation of the cells of the cleaving egg are as yet elusive, but they involve both the distribution of the egg's materials within the embryo, and the presence of chemical substances, the so-called organizers, that in part condition subsequent development.

The result of cleavage is a blastula consisting of a single layer of cleavage cells, or blastomeres. This embryo becomes successively two-layered and three-layered. In the two-layered stage it is customary to distinguish ectoderm and endoderm, the primary outer and inner layers. Between these a third layer, the mesoderm, is then developed from either or both of the primary layers. These three germinal layers, as they are called, give rise by various modifications to the organs of the embryo.

Further developmental processes convert the almost undiffer-

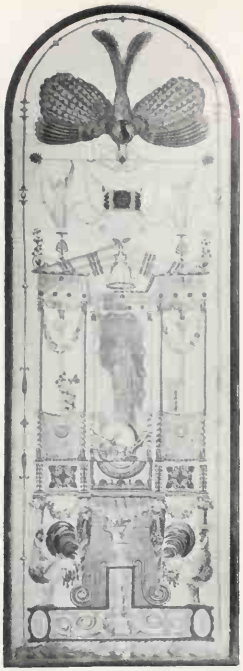
EMBROIDERY FROM THE IONIAN ISLANDS AND TURKEY: 16TH–18TH CENTURIES

Turkish towel, worked in silk and gold on fine linen, with stitches so arranged that both sides of embroidery are alike; 18th century

Detail of an embroidered bedspread from the Ionian Islands; 16th century. The design reflects Persian and Italian influence

PHOTOGRAPHS: (TOP LEFT) JACK SKEEL FOR ENCYCLOPÆDIA BRITANNICA, INC.; (BOTTOM RIGHT) THE ART INSTITUTE OF CHICAGO, GIFT OF MR. BURTON Y. BERRY

EMBROIDERY



French late 18th-century, satin panel worked in silk and metal thread; design by Jean François Bony

Detail of a silk skirt from the wardrobe of Marie Antoinette embroidered in gold and coloured silks; French, 18th century, period of Louis XVI

Silk panel embroidered in silk and cotton with a peacock motif; French, 18th–19th centuries

Embroidered wall hanging worked in the Italian shading method, design by Antonio Pollaiuolo; Italian, 15th century

Early 19th-century French dress border embroidered in a large floral design

FRENCH AND ITALIAN EMBROIDERY: 15TH–19TH CENTURIES

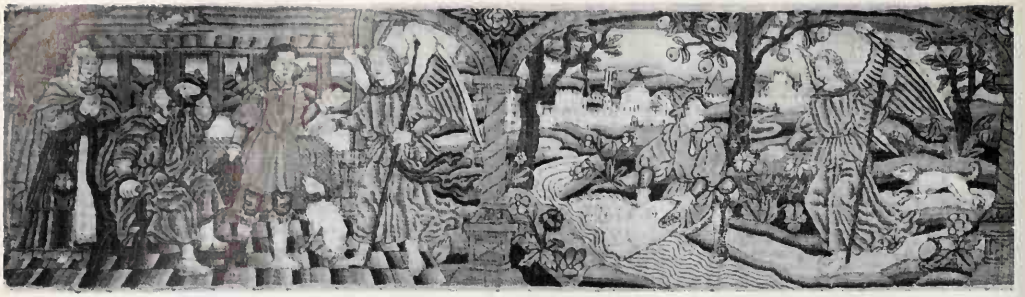
BY COURTESY OF (TOP LEFT) THE METROPOLITAN MUSEUM OF ART, ROGERS FUND, 1923, (TOP RIGHT) THE METROPOLITAN MUSEUM OF ART, GIFT OF MRS AUGUSTUS CLEVELAND, 1903, (CENTRE RIGHT) THE METROPOLITAN MUSEUM OF ART, ROGERS FUND, 1950, (BOTTOM RIGHT) THE METROPOLITAN MUSEUM OF ART, GIFT OF THE UNITED PIECE DYE WORKS, 1936, PHOTOGRAPH, (BOTTOM LEFT) MANSELL-ALINARI

BY COURTESY OF THE METROPOLITAN MUSEUM OF ART, ROGERS FUND, 1946

FRENCH EMBROIDERY: 17TH CENTURY

French canvas-work hanging representing Summer, embroidered in wool, silk and metal thread; Paris, c. 1683–84

EMBROIDERY



Detail of panel depicting the story of Tobias in seven scenes, embroidered in silk and metal thread on canvas; Flemish, mid-16th century

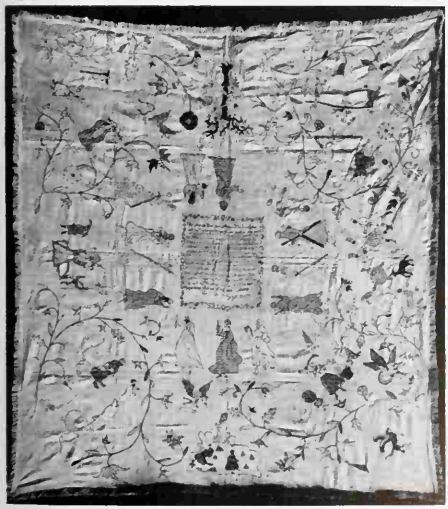
15th-century roundel showing St. Martin and the brigands, silk and metal thread on linen, 6½ in. diameter; Franco-Flemish, *c.* 1430–40

Embroidered panel representing the Flight into Egypt, worked in wool on canvas; Flemish, early 17th century

FLEMISH EMBROIDERY: 15TH–17TH CENTURIES

BY COURTESY OF (ALL) THE METROPOLITAN MUSEUM OF ART, (TOP) GIFT OF ARCHER M. HUNTINGTON, 1922, (CENTRE LEFT) THE CLOISTERS COLLECTION, PURCHASE, 1947, (BOTTOM) ANONYMOUS GIFT, 1949

18th-century patchwork cover of red, green, yellow and blue wool embroidered in coloured silk floss worked in herringbone, stem and chain stitches. The design comprises a central medallion surrounded by equestrian and other typical 18th-century figures and is bordered by strips of material and trimmed with coloured fringe



Linen cover with biblical figures embroidered in coloured silks and gold thread worked in stem, buttonhole and open chain stitches with French knots and couch work; 16th or 17th century

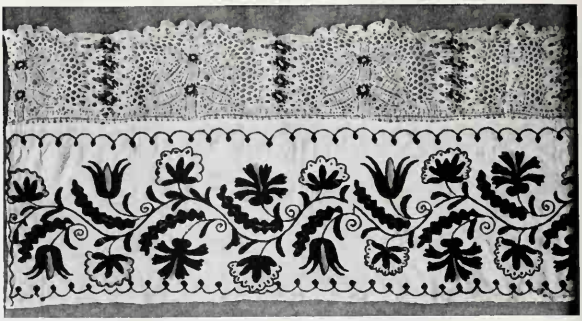
19th-century sampler of undyed linen with alphabet and ornamental designs embroidered with green, pink and brown silks in a fine cross-stitch; signed Anna Dorathea Wulfs, dated 1826

GERMAN EMBROIDERY: 16TH-19TH CENTURIES

BY COURTESY OF (ALL) THE METROPOLITAN MUSEUM OF ART. (TOP, BOTTOM RIGHT) ROGERS FUND, 1909. (BOTTOM LEFT) ROGERS FUND, 1908

EMBROIDERY

Hungarian white linen drawn work embroidered in coloured silks; 18th century



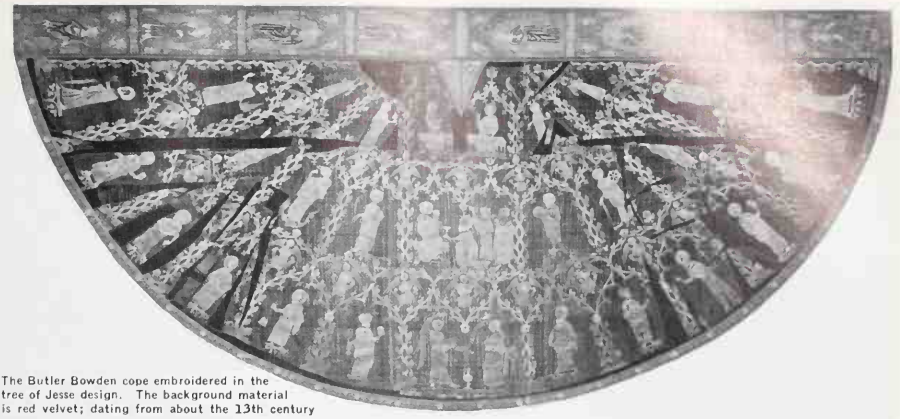
Border of linen embroidered in red cotton with touches of yellow and trimmed with bobbin lace; Hungarian, 19th century

Detail of a Georgian chalice cover of linen canvas embroidered in metal thread and coloured silks; 15th century

Danish towel made of linen embroidered in the *hedebo* style; 19th century

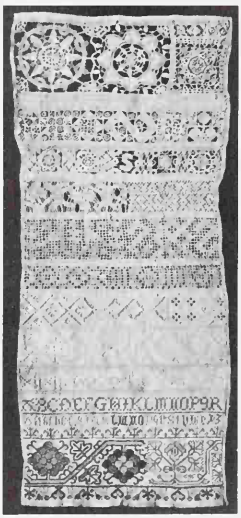
HUNGARIAN, DANISH AND GEORGIAN EMBROIDERY: 15TH-19TH CENTURIES

BY COURTESY OF (ALL) THE METROPOLITAN MUSEUM OF ART. (TOP LEFT AND RIGHT) ROGERS FUND, 1909. (BOTTOM LEFT) GIFT OF J. PIERPONT MORGAN, 1917. (BOTTOM RIGHT) ROGERS FUND, 1909



The Butler Bowden cope embroidered in the tree of Jesse design. The background material is red velvet; dating from about the 13th century

Portion of the Steeple Aston horse cope showing the background solidly embroidered in the ancient method of couching; dating from about the 13th century



Early English sampler showing needle-made lace and white surface patterns, drawn-thread, alphabets and coloured borders

Three motifs in canvas work applied to green velvet, part of the Oxburgh hangings worked by Mary Queen of Scots and Bess of Hardwick; second half of the 16th century. The motif on the right is assumed to be representative of Mary at the time of her imprisonment

EARLY ENGLISH EMBROIDERY

BY COURTESY OF (TOP, CENTRE LEFT, BOTTOM) VICTORIA AND ALBERT MUSEUM. PHOTOGRAPH, (CENTRE RIGHT) JACK SKEEL FOR ENCYCLOPÆDIA BRITANNICA, INC.

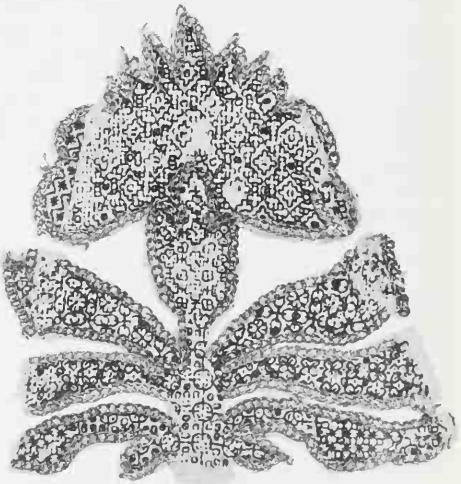
PHOTOGRAPH, W. ENTWISTLE FOR ENCYCLOPÆDIA BRITANNICA, INC.

ENGLISH EMBROIDERED SURCOAT

Replica of the surcoat of the Black Prince made by the Royal School of Needlework in 1953 and placed with replicas of other regalia above his tomb in Canterbury cathedral. The original dates from the 14th century and is preserved at the cathedral

English altar frontal from a side chapel at St. John's college, Oxford; 15th century. Design of lantern flowers is typical of this period with the fine lines softening the solid effect of the embroidery on the velvet background

Infant's gown decorated with Ayrshire white-work embroidery; Scottish, first quarter of the 19th century



Elizabethan black work, showing the typical counted patterns worked on linen and outlined in broad chain stitch in metal thread

ENGLISH AND SCOTTISH EMBROIDERY

PHOTOGRAPHS, (ALL) FOR ENCYCLOPÆDIA BRITANNICA, INC. (TOP) FRANK NEWENS, (BOTTOM LEFT) TOM SCOTT, EDINBURGH, (BOTTOM RIGHT) JACK SKEEL

EMBROIDERY

Chair seat cover embroidered with bright-coloured wool; 18th century

Detail of the bed valance, an early 18th-century example of crewelwork showing the simplicity of design characteristic of early American embroidery

U.S. EMBROIDERY: 18TH CENTURY

BY COURTESY OF (TOP) THE MUSEUM OF FINE ARTS, BOSTON (BOTTOM) THE ART INSTITUTE OF CHICAGO

Embroidered kerchief of white muslin worked by Rachel Leonard, Norton, Mass.; dated 1752

Sampler worked by Lucy Potter in silk on linen canvas using tent, satin and other canvas stitches; Providence, R.I., 1791

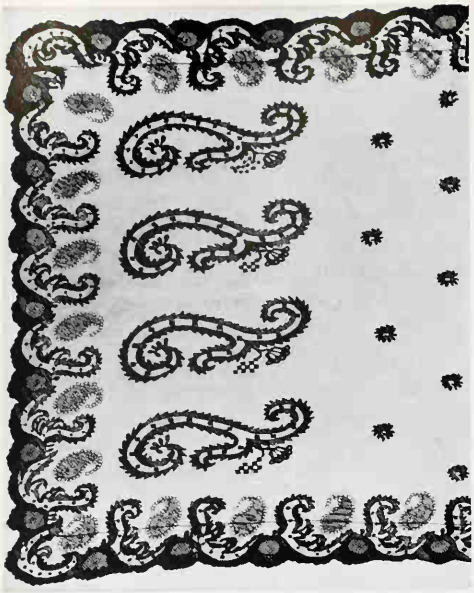
"Reclining Shepherdess," an embroidered (tent stitch) picture made in New England; mid-18th century

Brown wool bedspread embroidered in red, green and tan wool; made by the sisters of the Honorable Amos Patterson, Broome county, N.Y., c. 1790

U.S. EMBROIDERY: 18TH CENTURY

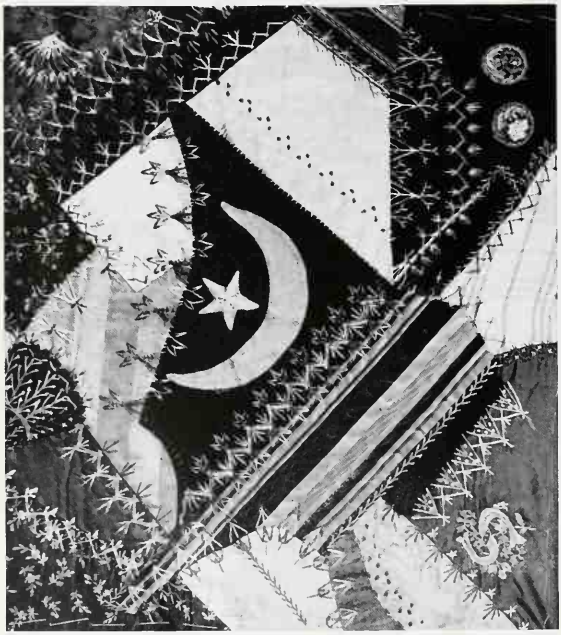
BY COURTESY OF (TOP LEFT, BOTTOM LEFT) MUSEUM OF FINE ARTS, BOSTON, (TOP RIGHT, BOTTOM RIGHT) THE ART INSTITUTE OF CHICAGO

EMBROIDERY

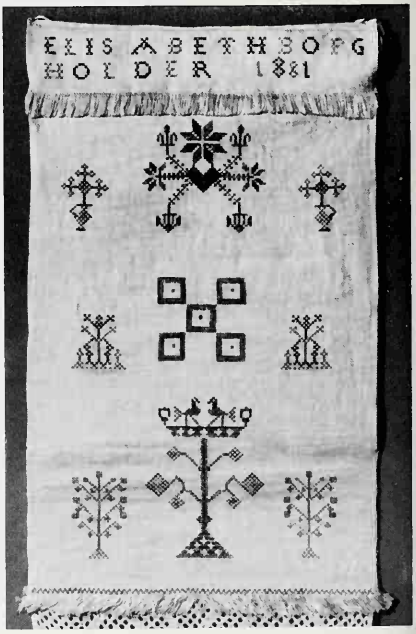


19th-century veil of black net darned with black silk thread; made in Holland, Ohio, *c.* 1835

Bride's quilt made with appliquéd designs cut from coloured fabrics and quilted in a variety of motifs; *c.* 1820



Block from a crazy quilt made by piecing together silk scraps in haphazard fashion and covering the seams with fancy stitches; *c.* 1884



Pennsylvania Dutch door towel of embroidered linen. The name of the maker and date are cross-stitched along the top; 1831

U.S. EMBROIDERY: 19TH CENTURY

BY COURTESY OF (ALL) THE ART INSTITUTE OF CHICAGO

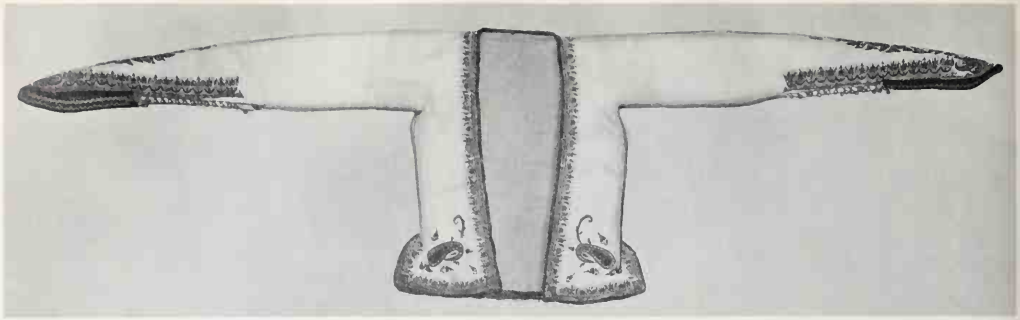
U.S. EMBROIDERY: 19TH CENTURY

Berlin woolwork, the most characteristic style of embroidery in the United States during the 19th century

19th-century rug with a typical American pastoral motif embroidered in coloured wool on canvas

BY COURTESY OF (TOP LEFT) MUSEUM OF FINE ARTS, BOSTON (BOTTOM) THE ART INSTITUTE OF CHICAGO

EMBROIDERY



White wool jacket embroidered in coloured silks worked in satin, stem and buttonhole stitches; Persian, early 19th century

Black-ground cover of thin cotton embroidered in darning stitch in red, light blue, yellow and other coloured silks; Persian, 18th century

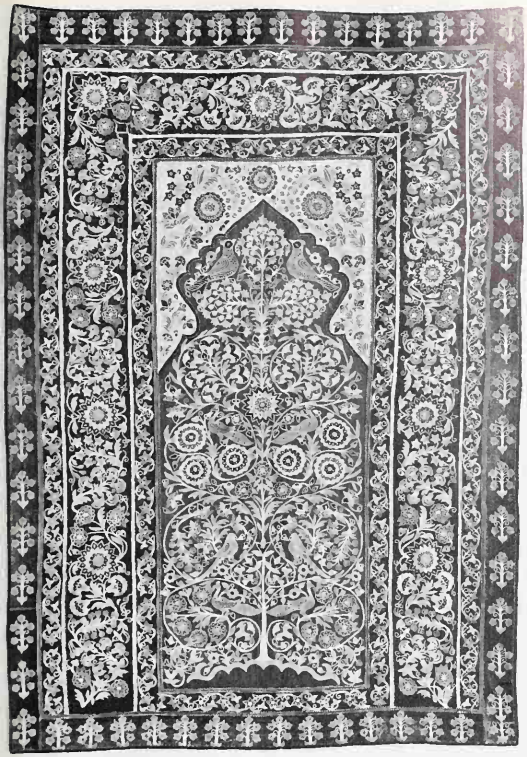
Satin coat embroidered in a fine chain stitch; India, first half of the 17th century

Persian cover of cotton embroidered in many-coloured silks on black ground; first half of the 17th century

Chinese man's wedding robe embroidered with various symbols; date unknown. The stripes at the hem represent water

PERSIAN, INDIAN AND CHINESE EMBROIDERY: 17TH–19TH CENTURIES

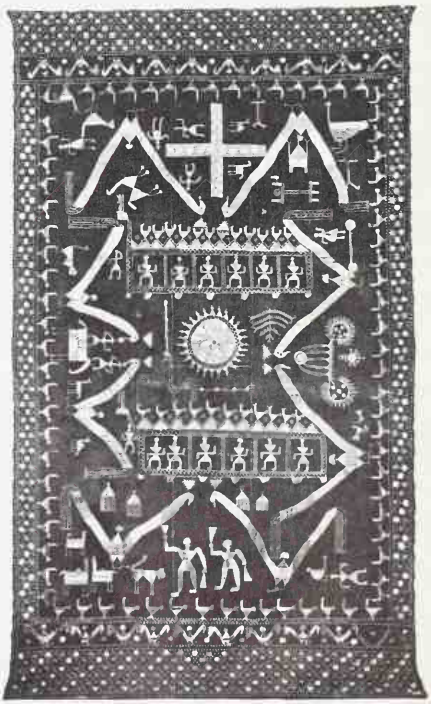
BY COURTESY OF (TOP) THE METROPOLITAN MUSEUM OF ART, GIFT OF ROBERT J. AND MARY B. MONKS, 1953. (CENTRE LEFT) THE METROPOLITAN MUSEUM OF ART, ROGERS FUND, 1908. (BOTTOM LEFT) THE METROPOLITAN MUSEUM OF ART, ROGERS FUND, 1909. (OTHERS) VICTORIA AND ALBERT MUSEUM



Persian hanging made of pieces of coloured wool cloth connected by silk thread chain stitching; 19th century

Japanese silk panel with two Bodhisattvas worked in coloured silks and gold and silver thread; 14th–16th centuries

Embroidered *fukusa*, a cloth in which to wrap presents; Japanese, 19th century



Indian cotton shawl embroidered with floss silks and cottons; from Punjab, 18th or early 19th century

PERSIAN, JAPANESE AND INDIAN EMBROIDERY: 14TH–19TH CENTURIES

BY COURTESY OF (BOTTOM RIGHT) VICTORIA AND ALBERT MUSEUM, (OTHERS) THE METROPOLITAN MUSEUM OF ART (TOP LEFT) ROGERS FUND, 1910, (TOP RIGHT) BEQUEST OF MRS. H. O. HAVEMEYER, 1929, THE H. O. HAVEMEYER COLLECTION, (BOTTOM LEFT) ROGERS FUND, 1909

EMBROIDERY

Chinese imperial robe of blue satin embroidered with the 12 ornaments; 19th century. Property of the Earl of Elgin and Kincardine

Detail of an Indian girdle, hand-painted and embroidered with red tulips on a cream-coloured foundation; *c.* 1700

CHINESE AND INDIAN EMBROIDERY

PHOTOGRAPHS (TOP) JACK SKEEL AND (BOTTOM) NORVAL, DUNFERMLINE, FOR ENCYCLOPÆDIA BRITANNICA, INC

entiated embryonic cells of the organs into the histological types peculiar to each. It is not until these histological changes are accomplished that the young embryo is ready to function fully and to take an independent place in the world. Hatching, therefore, occurs when the young embryo reaches a sufficient stage of development to maintain itself independently, at least in part. It is customary to use the term embryo to refer to the young animal during the period while its essential parts are undergoing formation. After they become established, we speak not of an embryo but of a larva, of which there are various types, or of a fetus in the case of mammalian embryos.

Specialized Structures.—Embryos display some special adaptations, the possession of which ensures that enough will survive and mature to maintain the species. The first of these adaptations is in the number produced. The most primitive environment in which development occurs is water, especially sea water. If eggs are shed directly into the water and fertilized there purely by chance, vast numbers of gametes must be produced since fertilization itself often fails and many young embryos that successfully start out, later become the food of other species of organisms, or die as the victims of physical forces. The chance that any one will reach maturity is very small. Since a single codfish can produce as many as 5,000,000 eggs in one season, the very force of this number is enough to guarantee that development will succeed in a large number of cases.

Other protective devices take the form of egg membranes which are present in almost limitless variety. All, however, come under three different heads: primary egg membranes, which are produced by the ovum itself; secondary membranes, produced by the follicle cells of the ovary; and tertiary membranes, which are added as the egg passes down the oviduct. Any or all of these membranes may be present or absent in particular cases, depending upon the habitat in which development is to occur. As the protective envelopes become more efficient the eggs increase in size and are better supplied with food materials, while the young are correspondingly stronger at birth, or at hatching.

There are also embryonic membranes that protect and aid the growth of the embryo during the period of immaturity. The four types of these membranes are formed by the growing embryo and are actually part of it. They are amnion (and serosa), yolk sac, allantois and, in part, the placenta. Although these organs are only of temporary significance, they are necessary during the stages of immaturity. Other adaptations and adjustments take the form of osmotic equilibrium between the contents of the embryo and the surrounding medium. Both structural and physiological adjustments are found in embryos upon which the maintenance of both individual animal and race depend. *See* GESTATION PERIOD; *see* also references under "Embryo" in the Index volume.

(A. Rs.)

EMBRYOLOGY is the science that treats of the development of an individual animal from its beginning as an egg to a period, arbitrarily determined, when it has attained or nearly reached its adult structure.

The Sex Cells or Gametes.—With perhaps a very few exceptions all multicellular animals, both Parazoa and Metazoa, reproduce sexually, the new individual arising by a fusion of two gametes, one, the spermatozoon, being produced by the male, the other, the ovum or egg, by the female parent. These two reproductive cells always differ greatly in size and appearance but have an equal influence on the characters of the animal arising from their fusion.

Spermatozoon.—The spermatozoon is a small, actively moving cell, which, in the majority of animals, has a very characteristic structure consisting of a head, middle piece and tail. The head is almost entirely occupied by the nucleus, capped by a body called the acrosome and surrounded by a thin layer of cytoplasm. The middle piece is cytoplasmic in nature, containing certain granules, of which the most important are a centrosome, which is responsible for the first division of the fertilized egg; and a basal granule, which controls the activities of the long tail, a thread of protoplasm, which by its spiral lashing drives the whole forward.

Ovum.—The ovum differs from the spermatozoon in being relatively large and immobile. It is a cell containing a nucleus and much cytoplasm the latter of which includes a food store of greater or lesser amount. This food reserve sustains the developing embryo until it becomes capable of feeding. As this event occurs at very different times, the amount of yolk, and thus the size of the egg, varies greatly in different animals.

Fertilization.—Fertilization (*q.v.*) is the act of union of a spermatozoon with an egg. It is followed immediately by a change of the physical condition of the egg surface that renders it impossible for any more spermatozoa to enter its substance. The entry of a spermatozoon begins a series of activities whose result is the formation from the apparently simple and uniform egg of the elaborately differentiated body of the adult. The first stages of this process are similar in essentials in all Metazoa, but the courses of development of the individuals of the different phyla of the animal kingdom part company at various stages depending on the remoteness of their relationship.

After the spermatozoon enters the egg, usually leaving its tail outside, a nucleus, reconstructed from the head, approaches the equivalent egg nucleus. Fertilization is completed when these two nuclei unite, although in some cases fertilization occurs only as the mitotic spindle is formed and the chromosomes from the two nuclei take their positions on it. The centrosome, which was brought in by the sperm, divides into two parts and forms a spindle between the nuclei; upon this spindle the chromosomes align themselves, thus forming the first cleavage mitosis. (*See* CYTOLOGY: *Cell Division.*)

Cleavage.—The originally single egg divides into two new cells, the blastomeres of the two-celled stage. These are usually, but not always, exactly similar in appearance. This is called the first cleavage. In the case of the sea urchin (an echinoderm), *Echinus esculentus*, this division takes place about one hour and ten minutes after fertilization and occupies about half an hour. In *Echinus*, an hour later each of these two blastomeres divides into two, giving rise to a four-celled embryo. The plane of this second cleavage is usually at right angles to that of the first, so that normally the four blastomeres lie in a ring, in contact with one another. Each of these cells then divides again by a third cleavage, often at right angles to the two that preceded it. The cleavage of the four cells takes place simultaneously.

Blastula Stage.—This process of cleavage is continued until there is built up a mass of some 128 cells, the blastula, in which the cells are arranged in a layer, one cell thick, surrounding a cavity, the segmentation cavity or blastocoele. Although the cells multiply in number, the embryo up to the blastula stage remains about the size of the original fertilized egg, the component cells being smaller with each successive division. The blastula is usually larger than the egg from which it arises because it imbibes water from its surroundings.

Gastrula Stage, Formation of Ectoderm and Endoderm.—The next process in development is known as gastrulation. It has as its object the formation of a primitive food cavity, the archenteron, surrounded by a double cellular wall, an inner layer of cells, the endoderm or hypoblast, which actually forms the wall of the gut and is concerned with digestion and the absorption of food, and an outer ectoderm or epiblast, which is the part of the embryo that comes into contact with the outside world. The ectoderm forms the whole external surface, and from it arise the nervous system and sense organs; in certain embryos arising from eggs that are shed into water, the ectoderm also produces cilia that serve for locomotion and feeding.

In relatively small eggs gastrulation is brought about by the actual intucking, or invagination, of the cells of one half of the blastula into the other. In larger and more yolky eggs so simple a mode of formation is impracticable, and various divergencies are found that may be regarded as modifications of the primitive method. Thus, in a frog's egg an actual intucking takes place, but is restricted to one point on the surface of the blastula, and leads to the formation of a shallow pit, subsequently lengthened by a very active localized growth of the part turned in, and a spread of organization from it forward into a mass of cells, which fills

most of what should be the blastocoele.

In such eggs as that of a bird the endoderm is formed by the splitting off of a layer of cells from a comparatively thick mass formed by the cleavage divisions, by a process called delamination. There are also other methods of endoderm formation (*see* Invertebrate Embryology).

Formation of Mesoderm.—The gastrula formed by any of these processes corresponds in its fundamental architecture with the adult animals that compose the phylum Coelenterata (*q.v.*). In the development of the members of all higher phyla a third cell layer, the mesoderm, is established. The mesoderm separates the ectoderm from the endoderm and from it are formed the majority, at any rate, of the muscles of the adult animal, its blood and blood vessels and its sex glands.

The mesoderm may arise in many ways. It may be entirely derived from special blastomeres very early set apart in cleavage, or it may come from the endoderm, or from the ectoderm.

Later Structural Developments.—In all higher phyla the mesoderm is either divided into two sheets from its first appearance, or becomes so later; one of these sheets lines the ectoderm and the other encloses the endoderm of the gut. The cavity that separates these mesodermal sheets is the coelom or body cavity (*see* Coelom and Serous Membranes). This space, which separates the gut from the body wall, is in many animals connected to the exterior by a canal or canals, the coelomoducts, whose function is to extrude the reproductive cells from the coelom, in whose wall they were formed, and into which they were set free. Subsequently, in the evolution of many forms, these coelomoducts become modified so as to subserve the function of excretion of nitrogenous waste products.

The coelom with its walls, the primordium of the mesoderm, seems to have arisen as a pouch pushed out from the archenteron, that is, as an enterocoele, but in many animals other modes of origin are found, the mesoderm also arising from as few as two cells, early set aside; from mesenchyme cells that wandered out from the endoderm; or by "delamination," being cut off as a sheet from the endoderm or even the ectoderm. Subsequently a split, a schizocoel, appears in the originally solid mesoblast so formed. The embryo at this stage possesses a gut, the enteron, surrounded by the endoderm and communicating with the exterior by a single opening, the blastopore.

In the Deuterostomia (*see* Invertebrate Embryology) the blastopore may close completely, often to reopen as the anus, the mouth being formed as a result of the perforation of an area of contact between the endoderm and the ectoderm. When the mouth and anus are established, the embryo begins to feed and passes into the larval stage.

Independent Existence.—*Oviparity.*—Oviparous animals extrude eggs that, after development, hatch in the medium in which they were placed. In many cases the hatched young is called a larva; this term properly implies an animal at a stage in its development when it can feed and move about without having acquired the structure of an adult. The larva may have a very short free life, a few minutes or hours in the case of some mollusks and tunicates, or it may cover the greater part of the individual's existence, as in the case of caterpillars.

The larva may be adapted for a mode of life and even to a medium entirely different from that of the adult animal that it will eventually become, in which case the passage to the adult condition is usually carried out rapidly by a process involving the destruction of the larval tissues by wandering phagocytic cells and the building up of their material into new adult tissues and organs. Such a transformation is known as a metamorphosis (*q.v.*). It is well illustrated by the mode in which a caterpillar becomes a pupa. All its tissues, except the nervous system and the gonads, become reduced to a structureless mass of cells, which are then built up into the body of the butterfly that will emerge from the chrysalis. The larval stage in a life history thus secures the possibility of growth under favourable conditions between the period when the food store originally laid down in the egg is exhausted and that when the adult form is reached.

Viviparity.—In animals of very many groups the egg is fertilized within the body of the female that produced it (a process usually requiring special intromittent structures in the male) and may be retained there until development has gone on to some definite point. The new individuals are then expelled either as larvae or as small copies of the parent. This process of bearing living young is called viviparity.

In its early stages this condition implies no more than a retention and eventual hatching of the egg inside the mother (ovoviviparity), but in its fully developed form it involves the continuous provision of food for the growing embryo, either in a condition in which it can be utilized by what is essentially an unborn larva, as in certain fish, or by diffusion into the embryo from the mother's blood, usually through a special organ, the placenta, as in mammals. Viviparity has the advantage that the few young that are born make their appearance at a larger size, are far more advanced in structure than would otherwise be possible, and thus have better chances of survival.

Biologic Concepts Derived from Embryology.—It is obvious that a period of development is a necessity for a metazoan; only thus can the elaborate body of one of the higher animals be built up from a single cell. It is not clear, however, that this development must follow so similar a course in all the higher forms. The only explanation is that all these animals are related by descent and owe the differences that separate them to evolution, and that the courses of their development, originally identical, have come to differ by modification of the later stages, cleavage and gastrula remaining much the same in all.

The Biogenetic Law.—The essentials of this concept were first stated by Karl Ernst von Baer in 1824. Much later Ernst Haeckel made a more rigid generalization, his so-called biogenetic law that ontogeny (the development of an individual animal) recapitulates phylogeny (the evolutionary history of the species to which it belongs). In other words, each embryo in its development passes through abbreviated stages that more or less resemble some developmental stage of its evolutionary ancestors, both remote and immediate. For example, man's embryonic development (ontogeny) indicates a lengthy and complicated evolutionary history (phylogeny). Haeckel compared the fertilized egg to a protozoon, the blastula to such a protozoan colony as *Volvox*, the gastrula to a coelenterate and so on. He applied the term caenogenetic to the characters shown by a developing embryo that could not have possessed an evolutionary significance and could never have represented the adult stage of any ancestor. Belief in the general validity of the biogenetic law led to immense activity among zoologists in the last quarter of the 19th century. More and more refined methods of observation were elaborated and applied to representatives of all the phyla. Special efforts were made to work out the development of the more primitive members of each group, with the hope, seldom justified by results, that knowledge of it would make plain their phylogenetic relationships. This work, although it can scarcely be said to have succeeded in its original aim, established morphology as a discipline with an accepted technique and theory. It made it certain that development would afford most valuable evidence about the homologies (*q.v.*) of organs, supporting the conclusions reached by comparative anatomy.

Cell Lineage.—The most interesting of the later developments of observational embryology was the study of cell lineages, which led to the remarkable discovery that the early development of a mollusk followed a most elaborate course, individual cells and small groups of cells being very early set apart for the production of definite structures in the larva and adult. Even more unexpected is the identity of this process in mollusks (snails, clams, etc.) and annelids (earthworms, leeches, etc.).

Control Mechanism.—It must have been realized from the beginning of the study of embryology that there is present in the fertilized egg a mechanism that controls subsequent development. The actual course of that development, in such forms as mollusks, suggests that it may be possible to discover its nature. The only possible method for such an investigation is an experimental one. The course of development may be modifiable by altering the environment, the temperature, the chemical composition of water in which development takes place, the space avail-

able for the growing embryo and so on. (*See* EMBRYOLOGY, EXPERIMENTAL.)

From various experiments the conclusion was reached that the cytoplasm of the egg possesses a structure that determines the course of development, and that it gains this structure before its maturation, while the oöcyte is growing. Experiments were designed and carried out with the intention of investigating the period in development when cells become committed to the formation of certain definite structures or types of structures.

Range of Study.—There are other embryological problems than those concerned with the mere sequence of stages. The alternation of asexual generations with a sexual one, metagenesis, occurs widely in some animal groups. A related phenomenon is the alternation of two types of sexual reproduction with each other, called amphigony, in which a biparental sexual generation alternates with one produced by the development of parthenogenetic ova (developing without fertilization). Parthenogenesis (*q.v.*) although involving only one parent, is a sexual process since the ova are germ cells, and in some animals they can develop with or without fertilization. Special problems grow out of this form of development and experiments on the artificial production of parthenogenesis were most instructive.

Early sexual maturity and reproduction occur in some animals that otherwise show only youthful characteristics. The scope of this kind of development is indicated by the definitions of terms used. Neoteny is the retention of ancestral larval conditions in some organs, although the animal has otherwise become sexually mature (as in the axolotl and some termite species). Paedogenesis is the sexual maturity of an animal otherwise youthful (as in larvae of some gall midges). Dissogeny involves sexual maturity at two different periods in the life of one organism, one at a youthful stage and the other when full adulthood is attained (as in the jellyfishlike ctenophores). These unusual cases occur in a wide range of the animal kingdom.

Heterochrony is the changed rate of development of different organs or systems. Modifications thus arising may influence the course of the later stages in a profound manner. The divergence in rates accounts for many modifications in later embryos of forms that started out quite similarly. The study of polyembryony, or the production of many embryos from a single egg, has thrown light on many obscure problems of development. The production of more than one embryo is no more remarkable than that of one, the how and why of which leads us through embryology into the most fundamental inquiries of the nature of life itself. *See* EMBRYO; EMBRYOLOGY, CHEMICAL; INVERTEBRATE EMBRYOLOGY; VERTEBRATE EMBRYOLOGY. For embryology in plants, *see* ANGIOSPERMS: *Characteristics*. *See* also references under "Embryology" in the Index volume. (D. M. S. W.; A. Rs.)

EMBRYOLOGY, CHEMICAL. The aim of chemical embryology is to discover the chemical mechanisms underlying embryonic development, in order to give chemical and physical basis to such terms as induction, determination, gradients and fields which are widely used in experimental embryology. This goal has not yet been reached, although marked progress had been made in several directions. Chemical embryology is one of the biological sciences closely linked with cytology, biochemistry and biophysics. It is to be expected that, in the near future, it will establish close connections with genetics also, and that it will quickly develop on a very large scale.

The main results obtained in the field of chemical embryology up to 1931 are to be found in J. Needham's important book, *Chemical Embryology*. All available data on the egg's chemical constitution and its metabolism were collected and discussed. The emphasis was definitely laid by Needham on the biochemical side; the embryo was little more than a convenient material to follow chemical changes and few results had a direct bearing on the problems of morphogenesis.

Since 1931, the general trend of research in chemical embryology has markedly changed. The embryological problems are now in the foreground, while biochemistry and biophysics are mostly used as powerful tools, in order to solve them. This recent evolution is a consequence of a considerable improvement in the techniques. Thanks mainly to K. Linderström-Lang and H. Holter, accurate and extremely sensitive analytical methods have been worked out. The present ultramicrotechniques (microtitrations, micromanometric methods) are at least 100 times more sensitive than the previous ones (*see* MICROMANIPULATION). Thus the experimenter may study a small fragment of a frog's egg instead of a batch of 50 or 100 eggs.

Cytochemical methods, which permit the detection in the cell itself of chemical constituents have also been improved and various attempts have been made, especially by T. Caspersson, to use them in a quantitative way.

With these new methods, interesting results have been obtained in several fields of chemical embryology; the most important ones are: the chemistry of the growing unfertilized egg (oöcyte), the biochemical changes occurring in the eggs at fertilization and during cleavage, the localization and metabolism of nucleic acids, the metabolism during growth and differentiation, the comparison of vertebrate and invertebrate embryos' metabolisms and the biochemistry of the organizer (*see* below).

Chemistry of the Growing Egg Cell.—Owing to the oöcyte's large size, it has been possible to solve an old problem of cell physiology: the question of whether the nucleus is the centre of the cell's metabolism. E. B. Harvey found that ultracentrifuged sea-urchins' eggs cleave into two halves, one of which is nucleated and the other devoid of nucleus. H. Shapiro and R. Ballentine have demonstrated that the nonnucleated part has the higher respiratory rate. Important respiratory enzymes (dehydrogenases) are bound to cytoplasmic granules that accumulate in the nonnucleated part obtained by such ultracentrifugation.

Nuclei directly isolated from frogs' oöcytes have only a very low respiratory activity and are poor in enzymes. It can therefore be safely concluded that the nucleus can no longer be held to be the centre of the egg's metabolism.

Biochemistry of Fertilization.—*Fertilizin.*—Much work has been devoted to the chemical nature of the substances that are given off by unfertilized eggs to the surrounding sea water (*e.g.*, F. R. Lillie's fertilizin): these substances activate the movements of the spermatozoa and agglutinate them in clumps. They are proteins present in the gelatinous coat of the egg. Sperm contain on their surface an antifertilizin that, when extracted, causes an agglutination of the eggs. In addition they contain lysins that dissolve egg membranes. The two last-mentioned substances are also proteins. Inside the egg another antifertilizin is present. Antifertilizin and fertilizin neutralize each other, just as in an antigen-antibody reaction. There is no doubt that these egg and sperm substances play an important role in the union of spermatozoon and egg.

Metabolism.—The respiratory rate, according to the species, is increased, decreased or left unchanged after fertilization. The oxygen uptake of unfertilized eggs, in various species of marine invertebrates, show wide variations; the fertilized eggs, on the contrary, respire much at the same rate. Fertilization brings the oxidation rate to the normal level, whereas the unfertilized egg often has an abnormal metabolism, either too high or too low. These chemical observations agree well with E. Bataillon's idea that the unfertilized egg is an inert cell because it is intoxicated by metabolic products.

Other purely biological theories of fertilization have found a firmer ground in recent biochemical experiments. For instance, J. Loeb thought that the spermatozoon induces first a disintegration (cytolysis) of the egg, soon followed by a reconstitution of the egg's structure. On the chemical side, A. Örström found that the large molecules (proteins, glycogen) present in the egg are broken down during the first ten minutes following fertilization; later, a resynthesis occurs. These chemical changes are what were to be expected according to Loeb's theory.

The activity of acids in inducing parthenogenesis is better understood now that J. Runnström has demonstrated an acid production soon after fertilization. The calcium ion is another efficient activating agent, and it is now known that this ion is set free from organic complexes when the egg is fertilized.

Obviously much more work on the biochemistry of fertilization

will be needed before all biological and chemical facts fit together in a general theory of fertilization.

Biochemistry of Egg Cleavage.—Biochemical studies on cleaving eggs have brought forward several interesting facts related to the physiology of cell division. First of all, the existence of cyclic variations in the oxygen uptake during cleavage in frogs' eggs has been shown. However, these variations are only of a small magnitude and they are not indispensable for cell division, since frogs' eggs undergo cleavage for a long time in the absence of oxygen. Presumably, this respiratory cycle during cleavage indicates the existence of anaerobic cyclic processes in the dividing cell.

Work on cleaving eggs has also demonstrated the necessity of —SH (sulfhydryl) groups for cell division: when these groups are blocked, mitosis is rapidly stopped. Probably, reversible denaturation of the egg's proteins, with a liberation of —SH groups, plays an important part in the spindle formation.

Nucleic Acids Synthesis During Development.—Cell divisions during embryonic development lead to a marked increase in the number of cells. This multiplication of the nuclei finds its chemical expression in a considerable synthesis of thymonucleic acid, which is one of the main constituents of chromosomes and chromatin. Thymonucleic acid arises from a cytoplasmic precursor, very probably ribonucleic acid, which is also present in the nucleolus. Cytochemical studies by Caspersson and Brachet (1942) have demonstrated that ribonucleic acid is present in large amounts wherever synthesis of proteins occurs. During growth and differentiation of the embryo, there is an active protein synthesis at the cost of the yolk. As might be expected, the ribonucleic acid content of the cells increases at that time. The behaviour of ribonucleic acid in different cells and at different stages of the embryo will thus be diverse: it decreases during mitosis, especially during cleavage stages—where very little growth occurs—and it increases as soon as new proteins are built up in the actively differentiating embryo.

Ribonucleic acid is linked to small, ultramicroscopic granules, present in the cytoplasm; these "microsomes" are very similar to some of the largest viruses, and they also contain important respiratory and hydrolytic enzymes. It is very likely that the microsomes are the site of protein synthesis.

Growth, Differentiation and Metabolism.—A. Tyler (1942) has extensively analyzed and studied the links existing between growth, differentiation, maintenance and metabolism in the embryo. Eggs are able to live for a while without exhibiting either growth or differentiation; growth is possible in the absence of differentiation, while differentiation without some amount of growth seems impossible. As regards metabolism, Tyler found, through ingenious experiments, that it is increased when the amount of work done for the differentiation and growth of the embryo also increases; growth and differentiation processes thus need energy.

Such a conclusion is in harmony with the well-known fact that the oxygen consumption steadily increases during development. In most cases, the embryo does not utilize all the energy sources that it contains at the same time. There is a succession in the use of these sources of energy, carbohydrate being first oxidized, then fats and lastly proteins. As a consequence the respiration quotient decreases in a regular way during development. In many cases, however, there is good evidence that carbohydrate metabolism starts only after cleavage is over, *i.e.*, at the onset of gastrulation (*see* EMBRYOLOGY).

Chemical Embryology of Invertebrate Eggs.—It is a well-known fact that invertebrate eggs fall into two different classes as regards the mechanism of their development. Some of them (worms, mollusks) are a mosaic of plasms giving rise to definite organs (germinal localization). The others (sea-urchins' eggs, for instance) are capable of extensive regeneration (regulation) when the egg is separated in two. Cytochemical observations indicate that regulative eggs have a much more homogeneous chemical composition than the mosaic eggs. In the latter, definite substances (enzymes, vitamins) are localized in definite regions. However, Holter and E. Zeuthen brought forward good evidence that the oxygen consumption of various parts of a mosaic egg (Ascidia) shows no significant differences.

Extensive studies on the metabolism of sea-urchins' eggs have been made by Runnström and P. E. Lindahl. It is known, since S. Hörstadius' work, that the development of animal (superior) and vegetal (inferior) isolated cells is different, and the presence of two opposite gradients in the egg has been suggested in order to explain these results. The concentration of an animal substance would progressively decrease from the animal to the vegetal pole, while a vegetal, antagonistic substance would decrease from the vegetal to the animal pole. Attempts to identify these substances or to establish the presence of qualitatively different metabolisms at the two opposite poles have so far failed.

It is certain that no simple, quantitative gradient prevails in sea-urchins' eggs, since there is no difference in the rate of oxygen uptake by animal or vegetal cells. Indirect evidence that respiratory enzymes are at work in the determination of the antero-posterior axis in the fertilized eggs has been put forward by D. C. Pease.

Chemistry of the Organizer.—This is a field where chemical embryology has been most active since the discovery in 1932 by H. Bautzmann, J. Holtfreter, J. Mangold and H. Spemann that implantation of killed organizers (dorsal lip of the blastopore) into amphibian gastrulae brings forth the induction of a secondary nervous system. The mechanism of induction, *i.e.*, the formation of neural tubes by the underlying presumptive chorda anlage (incipient notochord), is of course of fundamental importance in embryology. Four main lines of investigation have been followed by chemical embryologists: (1) comparative studies on the metabolism of the organizer and other parts of the gastrula; (2) cytochemical observations on the localization of chemical substances during gastrulation and neurulation (formation of nervous tissue); (3) attempts to identify inducing chemical substances; and (4) experiments on the physical changes leading to gastrulation and neurulation movements. Here is a summary of the main results so far obtained:

Metabolism of the Organizer.—It is now certain that the oxygen consumption of the gastrula cells decreases from the animal to the vegetal pole. The dorsal lip of the blastopore, *i.e.*, the organizer, has a lower metabolism than the future nervous system, but higher than the corresponding ventral part of the gastrula (future ventral lip). In the neurula, the dorsal (nervous system + chorda) have a much higher respiration than the ventral part, and a definite anteroposterior decreasing gradient is present in the nervous system.

Furthermore, it has been ascertained that carbohydrate metabolism initiates in the organizer and spreads out later to the remaining parts of the gastrula. The respiratory quotient and the anaerobic lactic acid production are higher in the organizer, which is also the site of considerable breakdown of glycogen (glycogenolysis).

It is, however, doubtful whether these metabolic differences are responsible for the inducing activity of the organizer. It is known that in a certain species, the obstetrical toad (*Discoglossus*), a nervous system can develop when the organizer is treated with cyanide, which strongly inhibits oxidations. Furthermore, a chemically induced increase in the carbohydrate metabolism of the ventral, inactive parts of the gastrula does not lead to the formation of the nervous system.

Cytochemical Studies.—Cytochemical methods first demonstrated the existence of a strong glycogenolysis in the infolding (invaginating) dorsal lip, which was later confirmed by quantitative methods. Of some interest also are observations on the localization of ribonucleic acid. This substance steadily decreases in concentration from the animal to the vegetal pole in pregastrula stages. In the gastrula itself, ribonucleic acid decreases during the invagination of the organizer, while it increases in the future nervous system.

In the neurula, a clear anteroposterior decreasing gradient is to be found in the nervous system; a decreasing dorsoventral gradient is also very obvious in the mesoderm which ultimately gives rise to the notochord (chorda-mesoderm). It corresponds with

the morphogenetic activity of this tissue. Cytochemical methods thus provide an easy means to demonstrate the existence of gradients in the embryo. They suggest that ribonucleic acid is somehow concerned in the chemical reactions which bring forth neural tissue induction. These nucleic acid gradients are probably gradients of protein synthesis, and they are parallel to the respiratory gradients. This was to be expected since we know that ribonucleic acid is associated with essential respiratory enzymes in small granules, the microsomes (see *Nucleic Acids Synthesis During Development*, above).

Inductions by Chemical Substances.—The fact that killed organizers are chemically active led to many attempts to isolate the "inducing substance." Unfortunately, it soon became clear that a large number of chemically unrelated substances can induce neural tubes after grafting in the gastrula; most important of these are sterols, acids, synthetic polycyclic hydrocarbons, nucleic acids, dyestuffs. They probably act in an indirect way, simply by releasing the true active substance from neighbouring disintegrating cells. Of some interest is the fact that dead tissues or nucleoproteins lose their activity after treatment with a nuclease, which removes nucleic acid. As cytolysis also induces a breakdown of ribonucleic acid in the dying cell, it might be that this acid or the products of its hydrolysis are of importance in nervous system induction.

Mechanism of Morphogenetic Movements.—A good deal of work has been done, mainly by Holtfreter (1943–44), on the complex movements (kinematics) undergone by the embryo during gastrulation and neurulation. Very important in this respect is the egg's surface layer. The egg is coated with a sticky, elastic substance. This surface coat becomes thicker wherever folding movements are occurring. Invagination movements, during gastrulation, are probably the consequence of this surface coat's existence and of surface tension differences between the different cell layers. Several lines of investigation point toward the idea that the folding and closure of the nervous system result from changes in the orientation or the configuration of the proteins, presumably those building up the cell membranes. *See* also INVERTEBRATE EMBRYOLOGY; VERTEBRATE EMBRYOLOGY; EMBRYOLOGY, EXPERIMENTAL.

BIBLIOGRAPHY.—General books: J. Brachet, *Chemical Embryology* (1950); B. H. Willier, P. A. Weiss and Viktor Hamburger (eds.), *Analysis of Development* (1955); J. Needham, *Biochemistry and Morphogenesis* (1942); H. Spemann, *Embryonic Development and Induction* (1938).

Articles: E. Bataillon, *Ann. Sci. Nat. Zool.*, 16:371 (1941); H. Bautzmann, J. Holtfreter, O. Mangold and H. Spemann, *Naturwiss.*, 20:51 (1932); J. Brachet, *Arch. Biol.*, 46:1 (1935), *Arch. Biol.*, 50:233 (1939), 53:207 (1942), *Bull. Acad. Roy. Sci. Belg.*, 29:707 (1945); T. Caspersson, *Naturwiss.*, 29:33 (1941); H. Holter and P. E. Lindahl, *C. R. Trav. Lab. Carlsberg*, 23:249 (1940); H. Holter and E. Zeuthen, *C. R. Trav. Lab. Carlsberg*, 25:33 (1944); J. Holtfreter, *J. Exper. Zool.*, 93:251 (1943), 94:261 (1943), 95:171 (1944), 98:161 (1945); F. R. Lillie, *Problems of Fertilization* (1919); R. S. Lillie, *Biol. Bull.*, 60:288 (1931); P. E. Lindahl, *Acta Zool.*, 17:179 (1936); K. Linderström-Lang and H. Holter, *Methoden der Fermentforschung* (1941); D. C. Pease, *J. Exper. Zool.*, 89:329, 347 (1942); A. Tyler, *Quart. Rev. Biol.*, 17:197, 339 (1942). (J. L. A. B.; X.)

EMBRYOLOGY, EXPERIMENTAL. All higher organisms originate from a fertilized egg, which has the value of a single cell. This cell, which has little visible structure, performs the remarkable feat of transforming itself within a reasonably short time into a multicellular complex organism. First, the egg divides by a series of cell divisions, whereby a stockpile of building blocks is created. The initial phase of these cell divisions is called cleavage. Subsequently, these embryonic cells undergo cell differentiation, and strains of specialized cell types, such as nerve or liver cells, are formed. These cells are organized in tissues and organs. Organ formation (organogenesis) often begins when the cells are still undifferentiated. Organs are first blocked out in their general form and then molded and shaped in detail by a variety of devices, such as invagination (infolding) or evagination (outfolding), whereby tubes and vesicles are formed; budding; hollowing out, etc. Processes of proliferation, differentiation and organogenesis are interwoven; they interact with each other following a precise time schedule. Generally speaking, development (*q.v.*) proceeds from relatively simple to more complex states, in a continuous flow of integrated events, until the functional organization of the adult is achieved. This general theme has innumerable variations in the life histories of the different animal groups. *See* EMBRYOLOGY; EMBRYOLOGY, HUMAN; INVERTEBRATE EMBRYOLOGY; VERTEBRATE EMBRYOLOGY.

RISE OF EXPERIMENTAL EMBRYOLOGY

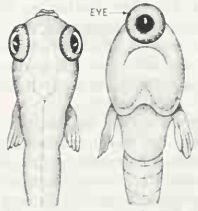
Descriptive *v.* Experimental Approach.—Descriptive embryology has practically fulfilled its immense task of assembling detailed descriptions of the development of representative animals, from the lowest invertebrates to man, and of deriving generalizations on the basis of comparative studies.

Toward the end of the 19th century it became clear that the descriptive and comparative approach was inadequate for a complete understanding of the developmental process; the need was felt for a more analytical approach. It was realized that although descriptive embryology can show in great detail how an organ such as the eye or brain develops, it fails to inquire into the causes or factors that call these structures into existence and that direct their further differentiation. Obviously, direct observation, no matter how subtle, cannot solve this kind of problem, and embryology, like other analytical branches of science, had to turn to the experimental method.

An Analytical Embryology.—The German anatomist Wilhelm Roux was among the first to envisage the scope of an analytical embryology. He became the founder and most articulate exponent of experimental embryology, which he called *Entwicklungsmechanik* ("developmental mechanics"). He formulated its program and its conceptual framework, and his pioneer experiments on the frog's egg (beginning in 1885) gave a strong incentive to further work in this field. In 1893 he founded the *Archiv für Entwicklungsmechanik der Organismen*, better known as *Roux' Archiv*, which was for several decades the leading international journal in this field. The *Journal of Experimental Zoology* became its equivalent in the United States. The leadership in experimental embryological research soon passed to Hans Spemann, also of Germany, who won the Nobel prize in 1935 for his discovery of the "organizer" (see *The Organizer, a Complex Inductor*, below) and to Ross G. Harrison, a United States embryologist, who did pioneer work in the fields of organ determination, neuroembryology and tissue culture.

INTRINSIC AND EXTRINSIC FACTORS

The Role of Milieu Factors.—One of the basic insights gained in the early phases of experimental embryology was the recognition that the fertilized egg contains all the essential and specific factors for differentiation within itself; among these factors are the all-important hereditary ones (genes) that reside in the nucleus. Yet development, like other organismic activities, takes place in an external milieu that provides the necessary general conditions for development. Different embryos are adapted to different ranges of conditions such as temperature, oxygen tension, etc. Variations within the normal range may have very general effects such as retardation or acceleration of development. Extreme or abnormal conditions usually produce specific and localized abnormalities. For instance, a short exposure of aquatic embryos of fishes, amphibians or mollusks to weak solutions of lithium chloride results in head deficiencies, including a striking abnormality known as "cyclopia," in which a single median eye develops instead of two lateral eyes (fig. 1). Many naturally occurring monsters, as well as hereditary abnormalities, are identical with abnormalities that can be produced experimentally by manipulation of milieu factors; hence these types of experi-



FROM C. R STOCKARD PUBLISHED IN THE "JOURNAL OF EXPERIMENTAL ZOOLOGY," BY THE WISTAR INSTITUTE OF ANATOMY AND BIOLOGY

FIG. 1.—(A) NORMAL FISH LARVA; (B) LARVA WITH CYCLOPEAN EYE (*see* TEXT)

ments have led to a deeper understanding of the origin of monsters (*see* MONSTER), of developmental genetics and, indirectly, of normal development. The value of extrinsic factors as analytical tools is enhanced by the fact that they can be varied quantitatively.

Some generalizations derived from these experiments may help to explain how the exposure of the whole embryo to a deleterious condition can result in localized defects. It was found that in many instances the same end result could be achieved by exposure to different and often unrelated agents. For instance, cyclopean monsters can be produced not only by lithium chloride but also by magnesium chloride, or by excessively high temperature. Conversely, the same agent can produce different effects when applied in different quantities or at different stages of development. For instance, the injection of a certain dose of insulin in a chick embryo after one day of incubation results in tail deficiencies, whereas the same dose, injected three or four days later, results in characteristic abnormalities of the legs. Likewise, the exposure of nine- to ten-day-old mouse embryos to an X-ray dose of 300 röntgens produces a high percentage of eye abnormalities, whereas the same treatment at 10–11 days results in a high percentage of leg abnormalities. Findings of this type lead to the concept of critical (or sensitive) periods: organ primordia pass through phases in their development during which they are particularly sensitive to deleterious extrinsic agents. For example, radiation may have a particularly harmful effect on structures when they pass through a stage of high proliferative activity. The concept of differential susceptibility is closely related to that of critical periods. Since different organ primordia pass through critical periods at different phases of development, the exposure of an embryo to a deleterious agent will affect selectively those structures that happen to be in their critical period at that stage, and leave others completely undisturbed. This explains how a short exposure of the whole embryo can result in lasting effects that are restricted to certain regions of the body or to certain organs or even parts of organs, such as the retina of the eye or individual skeletal elements of the limb.

However, the end effect is not determined by the time factor alone. Of even greater significance in the overall picture are physiological and biochemical aspects. Each chemical agent that produces abnormalities has its effects (inhibitory or otherwise) on one or more specific metabolic processes. Hence, the differential susceptibility of an embryonic region is closely related to the particular biochemical processes that go on in this region at the time of exposure. From this viewpoint, the embryologist can expect the most interesting information to come from experiments with those chemical agents whose effects can be understood, to some extent, in biochemical terms, as enzyme inhibitors, vitamins, antimetabolites. (*See* EMBRYOLOGY, CHEMICAL.)

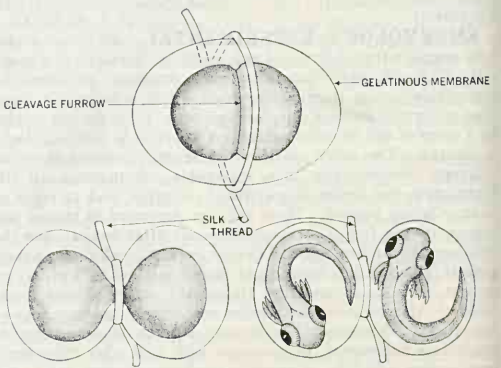
Regulative Capacity and Egg Organization.—For centuries, two alternative theories concerning the origin of organs in the embryo have been under discussion. One, known as the preformation theory, assumes that all organs are preformed in the egg as distinctive, though invisible entities; the egg would be, so to speak, a miniature adult. The theory of epigenesis, on the other hand, contends that no preorganization exists in the egg, and that the organs come into being, one after another, as the result of interaction of parts, or through other devices. Passionate, but largely speculative controversies centred round these theories, particularly during the 17th and 18th centuries. Experimental embryology has, however, placed the problem on a sound experimental basis.

The first classical experiment concerning egg organization was done on the egg while it divides into two cells, or blastomeres. (This division, known as first cleavage, is followed by the second cleavage, in which the first two blastomeres divide into four, and so forth.) Any idea of a stable preorganization in the egg had to be discarded when the German zoologist and philosopher Hans Driesch made the momentous discovery (in 1891) that each of the two first blastomeres of the sea urchin egg, when separated from the other, can give rise to a whole, though undersized, larva. If organ precursors had existed in the egg, then one would have expected that each blastomere, representing one-half of the egg, would have formed a half larva. Apart from demonstrating a remarkable regulative property of egg fragments, this experiment definitely implies an epigenetic concept of development. In the following decades similar results were obtained by blastomere separations of other animal groups, including bony fishes (teleosts) and amphibians among the vertebrates. An elegant method of obtaining twins from an amphibian egg (salamander or frog) was by the constriction of the egg in the plane of the first cleavage with a fine hair or silk thread (fig. 2). Clearly, if one-half of a salamander egg can give rise to a whole salamander, then any preexisting organization must have been of a very general nature, at best, and sufficiently labile to reestablish itself when disturbed.

These findings explain the occurrence of identical twins and double monsters (the latter originating from incomplete separation) in many animals and in man (in birds and mammals identical twins can likewise be produced experimentally). Natural twinning in birds and mammals usually does not occur during cleavage, but in later stages. The factors that cause it are not known. Multiple births such as triplets or quintuplets originate by fractionation of an early embryo into several parts. Their identity with respect to physical and other characters is due to their common origin from a single, fertilized egg; they all share in the same complement of hereditary factors.

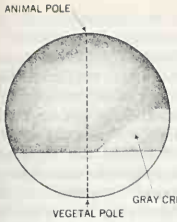
Not all animal eggs behave in this fashion, however. Isolated blastomeres of the eggs of annelid worms, mollusks and ascidians (sea squirts) produce only fragments of embryos. It appears in this group of eggs that the precursors of organs are already localized in early cleavage stages and that the capacity for regulation is lacking. On the basis of these differences embryologists used to distinguish between regulative and mosaic eggs (the latter term referring to a mosaic of precursors present in the egg). However, this distinction is unwarranted; the differences are in degree rather than in principle. First, the so-called mosaic eggs are not entirely devoid of regulative capacity, although it is limited to stages preceding cleavage. Second, the so-called regulative eggs are not entirely devoid of qualitative regional differences, which impose definite limitations on their regulative powers.

The Swedish embryologist Sven Hoerstadius demonstrated in a series of masterfully planned and executed transplantation and isolation experiments that the sea urchin (*Paracentrotus*) egg contains two qualitatively different agents or substances that are distributed in the egg in a gradient fashion. The peak of one gradient is at the upper (animal) pole and the other at the lower (vegetal) pole of the egg. A proper ratio of both is necessary for the formation of a normal larva; only those egg fragments that contain both in the required quantitative balance can form a typical whole larva. Consequently, splitting of the egg in a vertical (meridional) plane, as was done by Driesch, results in twins. If the egg is cut in a horizontal (equatorial) plane, the two halves give only fragmentary larvae.



BY COURTESY OF V. HAMBURGER

FIG. 2.—INDUCED TWINNING OF SALAMANDER EGG (*see* TEXT)



BY COURTESY OF V. HAMBURGER

FIG. 3.—FERTILIZED EGG OF FROG SHOWING ORGAN PRECURSOR MATERIALS

The amphibian egg shows a different type of early regional difference. An important organ precursor material becomes localized shortly after fertilization near the equator on the side of the egg that is destined to become the dorsal side of the future embryo. This early segregated material becomes visible in some frogs' eggs as the so-called "gray crescent" (fig. 3); its existence can be demonstrated experimentally in other amphibian eggs in which it is not visible. This region turned out to be the precursor of the organizer and as such indispensable for the formation of the embryo. Only egg fragments that contain at least part of this region can form an embryo.

In summary, two points of importance for a theory of development have emerged from these and similar experiments: First, all eggs possess, in varying degrees, a remarkable regulative capacity, which, in some, is limited to a short period before fertilization. Second, all eggs show the beginnings of differentiation in the form of some sort of qualitative regional difference. In most eggs these incipient differentiations do not represent actual organ precursors but differences of a more general nature, which mark the first steps in a sequence of progressive differentiation processes. These early segregations of qualitative differences are probably chemical in character, and referred to as chemodifferentiation. However, this term may be too restrictive, because structural differences on the molecular level may also be involved. It may be mentioned that some investigators (C. M. Child, A. Dalcq and others) consider the very first differentials in the egg to be purely quantitative differences of some basic physiological processes or substances distributed in gradient fashion.

EXPERIMENTAL METHODS AND CONCEPTS

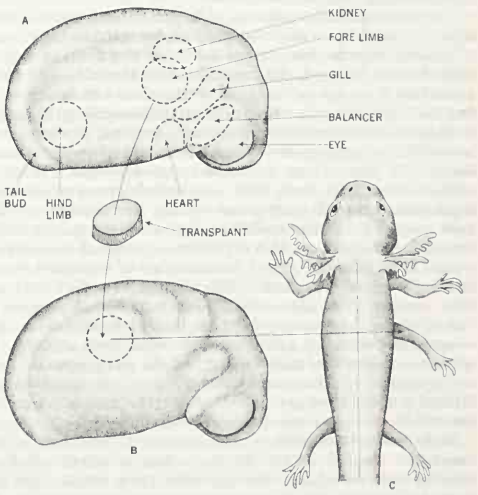
Transplantation.—Organ primordia as such do not preexist in the early stages. The when and how of their origin were answered through a new type of experimental procedure. The basic experiment, embryonic transplantation, is based on the distinction, made by Roux, between self-differentiation and dependent differentiation. A given embryonic area may be self-sufficient in the sense that it contains within itself all specific factors necessary for its further differentiation; it is then said to be capable of self-differentiation, that is, its fate is "determined." On the other hand it may require essential, specific stimuli from other embryonic parts in order to proceed with its differentiation; in this case, the region is said to be capable of dependent differentiation. Both terms are not sufficiently concise for modern use, but they are helpful in the first steps of analysis. The crucial test for the state of determination of an embryonic area is its transplantation to a foreign site. If it proceeds with its typical normal differentiation in the abnormal environment, then the transplant, at the time of transplantation, was "determined," that is, in possession of all agents necessary for its development. If not, then further experiments would have to be designed to search for the source and nature of the extrinsic stimuli that were missing. The area designated as "fore limb" in the five-day salamander embryo (fig. 4[A]) may be transplanted to the flank of another embryo (fig. 4[B]) and produce a normal fore limb (fig. 4[C]). Hence this area proves to be self-differentiating. Similar tests show self-differentiation of other areas as well.

This type of experiment, particularly when performed in much earlier stages, is greatly facilitated if one knows beforehand where to look for a particular organ-forming area, such as that of the brain, eye or limb. For this purpose, topographic "maps" of organ-forming areas have been constructed in the following way. Very small pieces of agar, serving as carriers of a red or blue vital (*i.e.*, nontoxic) dye are pressed briefly against the surface of an early embryo. Small areas can thus be marked on the surface of a spherical amphibian blastula or gastrula that has few natural landmarks or on the sheetlike blastoderm of a chick embryo; individual blastomeres can be stained in a sea urchin or sea squirt egg. These marks are then traced through subsequent stages and eventually located in a definite organ. In this way the entire organ pattern can be projected back onto an early stage. Fine carbon particles have served the same purpose as the vital dyes. Of course these maps give information only on the fate of an area in normal, undisturbed development, not on its inherent state of determination.

The eggs and embryos of amphibians are particularly favourable for all kinds of experiments: they are relatively large, develop outside the maternal body and are enveloped by gelatinous membranes that can be removed without disturbing development. Furthermore they are available in abundance the year round, since ovulation (egg release) can be induced by implantations of additional hypophyses (pituitary bodies) into the females. Together with chicken eggs and marine sea urchin eggs, amphibian eggs are the material of choice for the experimental embryologist. However, eggs and embryos of other animals have also been used, as, for instance, those of fishes, mammals, insects and numerous marine invertebrates. Each egg type and each operation requires its own technique.

The most widely used instruments for transplantation are glass needles drawn out to a very fine point and bent at the tip, or finely sharpened steel needles. Some experimenters use very fine iridectomy scissors (designed originally for surgery on the eye). A hair loop (that is, a fine hair sealed in a capillary tube) serves to manipulate the embryo. The embryos are placed in a glass dish with agar or wax on the bottom, and the operation is performed in a balanced salt solution under a binocular dissecting microscope. The remarkable healing power of embryos aids the experimenter greatly. The transplantation experiment has much broader applications than merely in testing for self-differentiation. It permits an almost unlimited variety of combinations of different embryonic parts and has thus become a major tool in the analysis of tissue interactions in development (see *Embryonic Induction*, below).

A more radical test for self-differentiation than transplantation to a foreign site is the method of explantation whereby the embryonic part is removed completely from the influences of a living environment. Again, amphibian material has been widely used for this type of experiment because the embryonic cells contain sufficient nutritive material in the form of yolk platelets to permit their differentiation in a nonnutritive balanced salt solution. J.



BY COURTESY OF V. HAMBURGER

FIG. 4.—LIMB TRANSPLANTATION IN SALAMANDER EMBRYO

Holtfreter, who did the pioneer work in this field, following an earlier experiment of Harrison, succeeded in rearing small isolates from very early embryos to advanced stages of differentiation and organ formation. Isolated fragments of embryos from warm-blooded animals (birds and mammals) require special nutritive solutions and have to be reared in the incubator (*see* TISSUE CULTURE). The chick embryologist has another method of isolation at his disposal: transplantation to the chorio-allantoic membrane. This is a highly vascularized double membrane that is closely applied to the inner surface of the shell of the egg and serves the embryo as a respiratory organ during incubation. Embryonic parts placed on it are completely isolated from the embryo but undergo further development, nourished by the blood vessels of the membrane. The chorio-allantoic membrane is also used widely for basic medical research as a substrate for the growth of mammalian tumors and viruses.

Morphogenetic Fields.—In the early salamander embryo (fig. 4), the contours of the head and eye vesicles and the trunk and tail bud are distinguishable, but gill, heart and other organs are not yet recognizable. However, transplantations of the embryonic areas outlined in fig. 4(A) demonstrate that they are already self-differentiating units. For instance, the cell group labeled as "fore limb" will differentiate to a normal limb when grafted to the flank; the gill area will form gills in an atypical position, and so forth. One can derive from these and corresponding experiments in other animal embryos the generalization that the first decisive steps of organ determination are completed long before visible differentiation sets in. This implies that the visible or microscopic differentiations are preceded by more subtle chemical or ultrastructural regional differences. Refined cytochemical, immunological and other techniques are now being developed to characterize these differences more precisely in biochemical and physicochemical terms.

Further analysis has shown that these self-differentiating units, called morphogenetic fields, do not always have distinct boundaries; some of these fields are larger than the area that actually forms the respective organ in normal development. For instance, Harrison found that if all the cells that normally form the fore limb are excised in the tail-bud stage of fig. 4, adjacent cells fill the gap and proceed with limb formation. In some instances, closely adjoining fields were found to overlap. It seems that these fields are gradient fields in the sense that the capacity for organ formation is highest in their centre and diminishes gradually toward the periphery. Fields of this type were also found in early chick embryos, and in other forms.

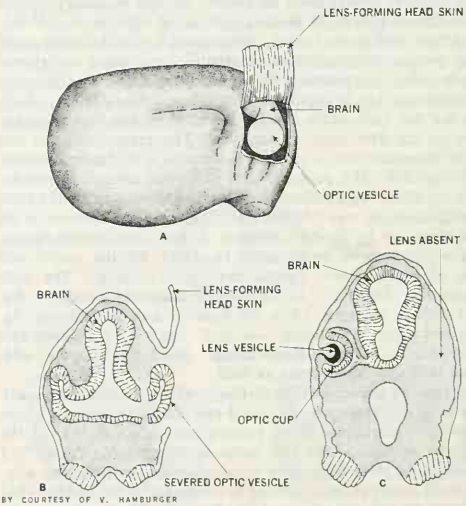
The most significant property of morphogenetic fields is their regulative capacity, which was analyzed by Harrison in a series of classical experiments. If one-half of a fore limb, or ear, or heart field is removed, the remaining half regulates to form a whole fore limb, ear or heart. It is therefore not surprising to find that by splitting or other suitable techniques two limbs can be obtained from one limb primordium, and two, or even three, hearts from a single primordium. The origin of supernumerary limbs, tails, hearts, toes, etc., which are occasionally found in monstrosities can be traced back to these early stages at which the organ primordia still possess the capacity for regulation. The same reasoning that was applied to the regulative capacity of the egg holds true here: the high regulative capacity of the morphogenetic fields rules out a rigid determination of structural details within the boundaries of the field. If half of a limb field can form a whole limb, then structures like individual muscles or skeletal elements cannot possibly have been preformed at the time of operation, otherwise a deficient limb would have resulted. In other words, both systems, the egg and the morphogenetic field, are capable of self-differentiation as a whole, but are devoid of detailed internal structuration. How they acquire this organization will be discussed below (see *Self-Organization*).

Embryonic Induction.—In a number of instances it was found that the fate of an embryonic area is determined by a stimulus from outside, or more specifically, by contact with adjacent embryonic tissue. Determinative interactions of parts comprise a variety of different processes that are collectively called embryonic induction. Inductions play an important role in epigenetic development and have been extensively analyzed.

The classical example of the determination of an embryonic structure by contact action is the induction of the crystalline lens by the optic vesicle. The fully formed eye can be considered as a double-layered cup, with a small opening, the pupil, in front. The inner layer of the cup is the retina, whose outer layer is a pigment epithelium. The lens is located behind the pupil and serves as a focusing device. Eye and lens originate from separate embryonic structures; the eye is formed as an evagination of the primitive brain vesicle (optic vesicle), whereas the lens originates as an invagination of the epidermis overlying the optic vesicle (fig. 5). The optic vesicle is transformed into the "optic cup" by the formation of a depression in its outer wall. At the same time the epidermis covering it folds in and forms the lens vesicle, which fits neatly in the optic cup. Eventually, the lens vesicle becomes detached from the epidermis and differentiates into the solid transparent crystalline lens, while the two layers of the optic cup differentiate into retina and pigment epithelium, respectively.

Several imaginative observers suspected that perhaps the optic vesicle might be instrumental in determining the lens formation by stimulating the overlying epidermis during the period when vesicle and epidermis are in contact. This idea of a causal relation was tested experimentally by extirpation (removal) of the optic vesicle before it had an opportunity to exert its influence. This experiment was first done on frog embryos. A flap of epidermis including the lens-forming area was carefully cut out with a glass needle and deflected (fig. 5[A]). The optic vesicle was then cut off at its base (fig. 5[B]) and the flap of epidermis healed back (fig. 5[C]). Great care was taken not to injure the cells that normally form the lens. In the majority of frog species no lens was formed in the absence of the optic vesicle, although the lens-forming area was left intact. It was concluded from this experiment that lens formation is induced by the optic vesicle.

The optic vesicle is called an inductor, and the lens area the reacting system. Inductions of this type, by contact, were found to be widespread among amphibians and other vertebrates and in some invertebrate embryos. For instance, the entire vertebrate nervous system originates as a plate, the neural plate, which comes into existence through induction by the directly underlying layer. Likewise, the inner ear, the nose, balancers and tail fin in salamanders originate by inductive interaction. In the formation of the wings and legs of the chick embryo, the solid (mesodermal) core of the limb bud induces the overlying epidermis to form the



BY COURTESY OF V. HAMBURGER

FIG. 5.—EXPERIMENT SHOWING DEPENDENCE OF LENS FORMATION ON PRESENCE OF OPTIC VESICLE: (A) WHOLE MOUNT; (B), (C) CROSS SECTIONS

typical pattern of feathers of the wing and of the scales and claws of the hind leg. In several instances, as in lens and ear induction, it was found that not one but two adjacent tissues are involved as inductors; they operate in succession and reinforce each other.

After the causal relation between the optic vesicle and lens formation was established, the question arose whether the inductor merely triggers off a chain of events in a group of cells that was already specifically predisposed for lens formation, or whether it could induce a lens in embryonic tissue that would normally form another structure. Two types of transplantation experiments gave the answer: the transplantation of the optic vesicle under flank epidermis and the transplantation of flank or head epidermis over the optic vesicle. In both instances, a typical lens was formed by the foreign epidermis, at least in some of the frog species in which these experiments were done. This shows that induction is more specific than a mere trigger action would imply. It can be stated generally that the inductor(s) not only determines the time and place of origin of a structure but also transmits to the reacting system specific "information" as to the type of differentiation that the system should follow.

The nature of the induction process becomes clear when another experiment is performed: a variation of the above-mentioned experiment of transplantation of flank epidermis over the optic vesicle. If the same piece of flank epidermis that was induced by the optic vesicle to form lens had been transplanted in a more posterior position in the head, where it would have been exposed to the ear inductors, it would have formed an inner ear. The embryonic epidermis thus proves to be relatively indifferent up to the time when it comes in contact with the inductor and potentially capable, or competent, of proceeding along any one of several lines of differentiation. However, there are definite limitations to the competence of the reacting system: it has to be in the proper physiological state of responsiveness. For instance, if it is too old or too young, the reacting system will not respond to an inductive stimulus.

The further analysis of the inductive process was greatly enhanced by another methodological advance. It was found feasible to exchange embryonic tissue between different species, orders or even classes. In adults it is not possible to make a graft even between two individuals of the same species, because of immunological reactions that destroy the graft, but these immunological incompatibilities do not exist in embryos. Transplantations between two different species are called heteroplastic, and transplantations between more distantly related forms, such as salamander and frog or mouse and chick, are called xenoplastic. The eye vesicle of a frog can induce a lens in the prospective head or flank epidermis of a salamander; the ear or brain inductors of the salamander can induce these structures in the frog, and so forth. Inductions of duck tissue by chick and vice versa have been obtained. These experiments prove that the inductive stimulation processes are mediated by agents (probably chemical in nature) that are not species-specific. However, the lens induced in salamander epidermis by the optic vesicle of the frog has all characteristics of a salamander lens, and the ear produced by the salamander's inductors in the frog epidermis is a typical inner ear of the frog. Invariably it is the hereditary equipment of the cells of the induced structure and not that of the cells of the inductor that determines the character of the induced organ.

The most striking illustration of this principle was obtained when flank skin of the frog embryo was transplanted to the head of an early salamander embryo. Head structures and mouth parts of salamander and frog larvae differ strikingly from each other in many respects. For instance, salamander larvae have true dentine teeth, whereas frog tadpoles have dark-brown horny jaws. As was expected, the flank skin of the frog embryo was induced to form head structures in the new position. In doing so, however, it followed its own genetic repertory, which provided for horny teeth only. The result was a curious combination: the embryo had a salamander head that was equipped with the horny jaws and other features of the frog!

The Organizer, a Complex Inductor.—A brief account of the gastrulation of Amphibia must be given to make the discussion of the discovery of the organizer comprehensible. After fertilization, the egg cell undergoes a series of cell divisions leading to the formation of a hollow vesicle, the blastula. The next phase of development is a series of cell movements, collectively called gastrulation. A considerable portion of the blastula, namely, the entire lower half, up to the equator, migrates into the inside of the blastula cavity over the edge of a slitlike opening, the blastopore. The invaginated cells arrange themselves in two sheets, the innermost endoderm, which forms the lining of the intestine and other organs, and a layer, the mesoderm or archenteron roof, which lies between the endoderm and the outer covering ectoderm. The latter gives rise to the nervous system, to the sense organs such as the ear and nose and to the epidermis of the entire body. The mesoderm forms musculature, skeleton and other tissues. At the beginning of gastrulation, the future mesoderm is located above the incipient blastopore, at its so-called dorsal lip. The blastopore marks the future tail end of the embryo.

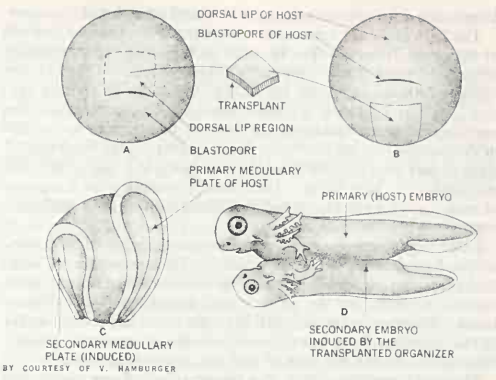


FIG. 6.—ORGANIZER EXPERIMENT (*see* TEXT)

The discovery of the organizer in the salamander embryo by Spemann and Hilde Mangold (1925) was one of the most significant events in experimental embryology. The upper lip of the blastopore, that is, the prospective mesoderm material at a stage when it is just about to move inward, was cut out with a fine glass needle and transplanted to the flank of another gastrula (fig. 6). The transplanted area proceeded with its invagination movements in its new position. Once inside, it underwent differentiation into mesodermal structures such as skeleton and musculature. In this respect, its behaviour was in no way unusual: the transplant self-differentiates into those structures that it would have formed in the normal course of its development. In addition, it performed extraordinary inductive activities, inducing in the overlying host flank ectoderm a more or less complete second nervous system, which later developed into brain parts and a spinal cord. Furthermore the transplant induced additional musculature, kidney and occasionally a second intestine. The most remarkable feature, however, was that all structures together, those that originated from the transplant by its self-differentiation and those that originated by induction, combined to form a whole, well-integrated, secondary embryo on the flank or belly region of the host embryo (fig. 6). These secondary embryos reached different degrees of perfection. In the best cases they were almost complete, with head, trunk and tail, heart, kidney, intestine and other organs. The capacity of the upper lip of the blastopore, or future mesoderm, to initiate and induce the formation of a whole secondary embryo has earned it the designation organizer.

As the foregoing discussion shows, organizer action is fundamentally a combination of self-differentiation and of complex inductions. Further transplantation experiments showed regional differences in the inductive capacity of the organizer. The part that comes to lie in the head region induces predominantly head structures, such as brain and eyes, whereas the part that comes to lie farther back induces predominantly trunk and tail structures.

Hence, one distinguishes between a head and a trunk-tail organizer.

The first organizer experiments were made on European newts of the genus *Triturus*. The research on the organizer was greatly enhanced by the use of differently coloured eggs. Those of one species, *Triturus taeniatus*, are brown, and those of another species, *Triturus cristatus*, are unpigmented. By transplanting the organizer from one species to the flank of the other, it is possible to distinguish precisely between those organs of the "secondary" embryo that were derived from the transplant and those that were contributed by host tissue. Occasionally, organs were composed of mixtures of both pigmented and unpigmented cells (chimeric organs).

Later, organizers were discovered in the embryos of lampreys, teleost fishes, frogs, toads and chicks. The occurrence of identical twins and of double monsters in mammals and in man gives indirect evidence for the existence of organizers in the highest forms. Embryonic regions that have inductive capacities similar to the organizer have been found in some invertebrate groups; for instance, in the sea urchins and planarians (flatworms).

The above-mentioned fact that the organizer of one species can operate in the embryo of another shows that organizers, like other inductors, are not species-specific, and the organizer of a frog or even of a fish can induce a secondary embryo in a salamander.

The Chemical Nature of Inductors.—In 1932, Holtfreter and others made the important discovery that the amphibian organizer retains its inductive capacity even after its component cells are killed in alcohol or by heat, freezing or other treatment. Soon afterward, Holtfreter found that a large number of tissues from vertebrate embryos and adults, and from some invertebrates, when alive or killed, can stimulate inductions in competent embryonic epithelium. It was thus established that embryonic inductions are mediated by chemical agents, apparently not of a highly specific nature, since they are contained in many different tissues. These chemical substances, so far hypothetical, have also been referred to as evocators.

J. Needham and his collaborators, in Cambridge, Eng., made the first attempts to characterize the evocators chemically. Although their contention that the evocator of the nervous system is a steroid was not supported by later evidence, their pioneer studies gave the incentive to numerous investigations along this line. The early hope that it might be possible to identify specific chemical substances as evocators of specific organs was soon dispelled. It was found that brain and other inductions could be elicited by a variety of chemically unrelated substances, such as fatty acids, steroids, nucleoproteins. These findings indicated that the specificity of the response (that is, brain formation, etc.) must be attributed, to a considerable extent, to the inherent properties of the reacting system. Moreover, an attempt to identify any one of these chemical agents with actual inductors of the living embryo meets with a serious methodological difficulty that has proved most difficult to overcome. All tests of chemical agents make use of the epidermal tissue of a very young embryo (the gastrular ectoderm) as the reacting system. From earlier experiments it has been postulated that the system used as a detector for inductivity contains in itself an inductor for nerve tissue and that the chemical substances that are to be tested may produce inductions in an indirect fashion, by merely releasing this evocator in the responding system. This argument is by no means farfetched. Some experimental data on record are hardly explicable in any other way. For example, differentiation of brain tissue in the ectoderm of the gastrula can be obtained in the absence of any inductor merely by raising or lowering the acidity of the culture medium to abnormal levels or by effecting similar, entirely nonspecific, changes. The most plausible explanation of these results is that the abnormal culture medium calls forth a slight injury to the epidermal cells, resulting in the release of the intrinsic evocator; that is, some sort of self-stimulation is involved. Some investigators even went so far as to postulate that all embryonic induction is nothing but an entirely nonspecific trigger mechanism.

This extreme viewpoint is hardly justified. In a number of laboratories several promising lines of investigation are being actively pursued that point in a different direction. Their results indicate that the living inductors release inductive substances into the culture medium. These substances are amenable to chemical analysis with microchemical methods; apparently nucleoproteins are involved.

Other findings indicate that there are at least two qualitatively different inductive agents, and efforts are being made to characterize their active components chemically by appropriate fractionation procedures. It was found that certain adult tissues induce exclusively forebrain and eyes; others induce predominantly trunk and tail structures. Inductively active fractions of such tissues differ chemically from each other; for instance, the trunk-tail evocator is more heat-labile than the forebrain evocator and the former can be converted into the latter by heat treatment. It seems that the most active fractions are proteins, but the problem cannot be considered as settled. (*See* also EMBRYOLOGY, CHEMICAL.)

Transmission of Inductive Agents.—Normal inductions in the embryo require an intimate contact between the inductor and the reacting tissue. Whether the inductive agents are transmitted by diffusion from inductor to reacting tissue; whether they penetrate into the reactive tissue or initiate a cell surface interaction are questions that will have to be resolved by radioisotopic tracer or "atom tagging" studies and other modern techniques. One new technique that promises to give further insight in the transmission of inductors may be mentioned. A tissue culture experiment was set up in which the inductor and the reacting tissue, though separated by a porous ("millipore") filter, still interacted. Apparently intimate contact is not always necessary; however, there is evidence that the transmission is not merely a simple diffusion process. The clarification of this basic problem remains for the future.

Self-Organization.—Autonomous segregation or self-organization are terms implying that some embryonic systems do not require "instruction" from outside, such as inductive stimuli, in order to proceed with their development. Differentiation in such systems is an autonomous transformation from simple to more complex organization, but the nature of the internal processes whereby regional differences gradually originate is very poorly understood. In a sense the initial steps in the development of the egg cell represent such a process, and the differentiation of morphogenetic fields is another example.

In eggs of annelid worms and mollusks the only sign of incipient organization is a simple stratification along the horizontal or equatorial plane, indicating qualitative differences between the upper (animal) and the lower (vegetal) pole region. During the first division of the egg, precursor materials for certain larval organs are sorted out and included in one of the two daughter cells; each subsequent cell division carries the segregation of specific organ-forming or organ-determining materials a step further. As a result, qualitatively different substances become separated from each other and are distributed to different cells or cell groups whose further differentiation they control. In these instances, then, the process of cleavage itself is instrumental in sundering and stabilizing certain qualitative differences, thus transforming a relatively homogeneous system into an increasingly complex system. Individual isolated cells, or small groups of them, undergo strict self-differentiation; they form exactly the same structures that they would have formed in the whole embryo; in other words, each blastomere has received a definite assignment from which it cannot deviate.

A striking example of visible segregation was found in the sea squirt *Cynthia* (*Styela*). The egg contains several cytoplasmic materials of different coloration. Fertilization initiates a flow and redistribution of these substances, which settle in different regions of the egg. Each of the differently coloured areas represents a specific organ-forming region. For instance, the yellow region will form musculature, and the light gray region will give rise to the nervous system.

In vertebrate embryos the morphogenetic fields (*see* above) apparently undergo comparable internal self-organization processes. The fields themselves are relatively homogeneous in structure, as shown by their regulative properties, but they gradually break

up into smaller, self-differentiating subunits. For instance, the limb field develops into a limb with its complex pattern of skeletal and muscle elements. It seems that this pattern gradually emerges from the embryonic limb bud by autonomous segregation processes, without apparent aid from structures outside the limb itself.

A very impressive demonstration of the capacity of embryonic cell groups for self-organization is given by cell dissociation and reaggregation experiments. It is possible to separate the cells that compose an organ primordium, such as a limb bud or a kidney primordium, by enzymatic action or by use of other agents that remove the substances that cement cells together. In this way suspensions of individual embryonic cells may be obtained. These can then be recombined mechanically or by gentle centrifugation and the aggregates can be grown in an appropriate culture medium. The cell clusters composed of embryonic kidney cells showed the capacity to form typical kidney tubules, and the aggregates of limb bud cells formed skeletal elements. Even more remarkable is the outcome of an experiment in which suspensions of these two cell types were intermingled and the mixture reaggregated. In the cluster thus formed, the cells, at first distributed at random, very soon begin to sort out: the embryonic kidney cells group together and form kidney tubules, and the embryonic limb bud cells aggregate and form nodules of skeletal tissue. In another experiment skin cells of chick embryo were dissociated and reaggregated; they reorganized into typical skin including feather buds. Apparently, embryonic cells have special affinities for other cells of their own kind, predilections that enable them to seek out each other, to adhere to each other and to build typical structures and organs together. These experiments may lead to a better understanding of self-organization in terms of the physiological properties of the individual cells that constitute an embryonic field.

Summary.—The application of the experimental method to embryological problems has greatly advanced our understanding of the developmental process. Although the time is not yet ripe for the formulation of a comprehensive theory of development, some insight has been gained into the processes whereby the egg manages to transform itself into a complex organism.

It has been seen that animal eggs are not entirely homogeneous but show the beginnings of qualitative regional differences. These incipient differentiations are elaborated and gradually stabilized in subsequent phases of development, and in this way the morphogenetic fields are established. The fields are transitional stages; they are organ primordia on their way to complete differentiation.

The same gradual stepwise segregation of subunits within a larger unit, by which the fields came into existence, repeats itself within the boundaries of the fields until eventually each cell group is irreversibly committed to its fate and visible differentiation begins. In other words, the larger units are blocked out first and the details filled in later. This concept of progressive differentiation is largely founded on experiments on vertebrate embryos; however, it applies also to invertebrates, with some modifications. In annelidan, molluskan and ascidian eggs the segregation processes begin in the cleavage stages or even earlier, and specific organ-determining agents are already sorted out in individual blastomeres. As a result, the blastomeres and the organ-forming areas derived from them become self-sufficient (self-differentiating units), and proceed with their development relatively independently of each other.

In vertebrate and sea urchin embryos, on the other hand, interactions between the parts of the embryo occur all the time, and they play an important role in the determination of the fate of the different regions. The best known type of such interactions is embryonic induction, in which organ formation is initiated by stimuli emanating from one or more adjacent regions.

The complexity of the embryonic organization thus increases stepwise, each advance being prepared in the preceding phase of development. It is in this sense that an epigenetic theory of development is adopted. It may be said that preformation of a sort is involved also (but not as it was earlier understood): the blueprint for the sequence of structural and physiological processes that constitute embryonic development is "preformed" in the hereditary equipment, the genes, contained in the nucleus of the fertilized egg.

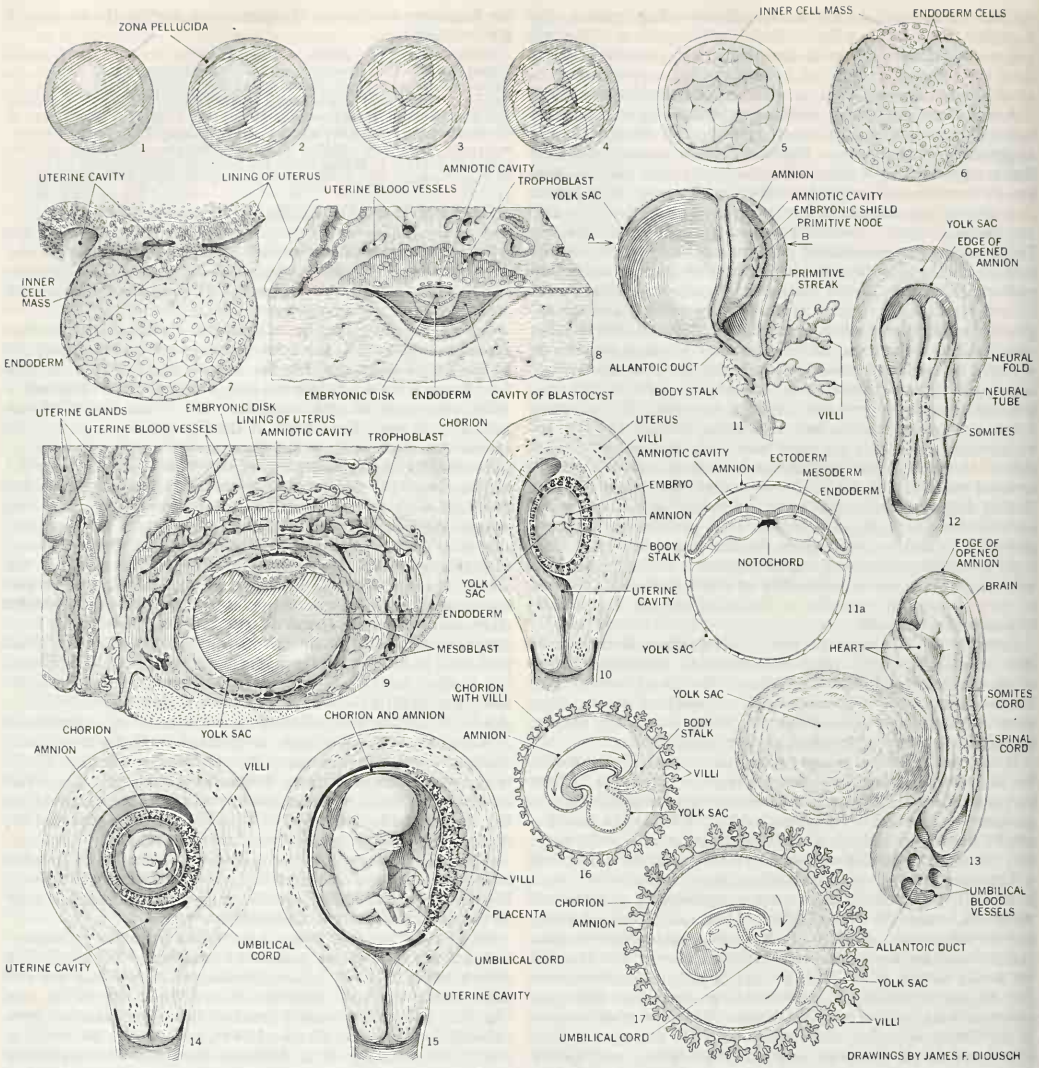
Bibliography.—J. Brachet, *The Biochemistry of Development* (1960); C. M. Child, *Patterns and Problems of Development* (1941); A. Kühn, *Vorlesungen über Entwicklungsphysiologie* (1955); V. Hamburger, *A Manual of Experimental Embryology* (1960); J. Needham, *Biochemistry and Morphogenesis* (1942); C. P. Raven, *An Outline of Developmental Physiology* (1959); H. Spemann, *Embryonic Development and Induction* (1938); C. H. Waddington, *Principles of Embryology* (1956); P. Weiss, *Principles of Development* (1939); *Analysis of Development*, ed. by B. H. Willier, P. Weiss and V. Hamburger (1955). (V. H.)

EMBRYOLOGY, HUMAN. The development of the human embryo follows in a general way the pattern described in the article Vertebrate Embryology. The special purpose of the present article is therefore to describe and illustrate those features of development that are peculiar to man or of special interest in connection with human biology.

The human body, like that of all higher animals, develops from a single cell, formed by the union of an egg cell or ovum from the mother and a sperm cell from the father. The human germ cells (*see* Fertilization: *Fertilization in Man and Other Mammals*) differ from those of other mammals only in fine details of size and of structure. The events of the first six or seven days of human development have not been as yet completely seen by embryologists and certain parts of the process can be described only conjecturally on the basis of observations made on other mammals more readily available for laboratory study. In such animals the fertilized ovum begins to undergo cell division (segmentation) within a few hours after fertilization and while moving through the oviduct (Fallopian tube) toward the cavity of the uterus. In most of those mammals in which transportation and segmentation of the ovum has been observed, the ovum enters the uterus on or about the fourth day after discharge of the egg from the ovary (ovulation). By this time it has divided into at least four cells; in some species (*e.g.*, the rabbit) division has proceeded still further, so that the embryo upon entering the uterus is already a solid little mass of cells. There is reason to believe that this is the case in man. Cell division of the ovum in mammals is characteristically holoblastic, as contrasted with the meroblastic division of the yolk-rich eggs of amphibians and birds; that is to say, the mammalian ovum divides completely into two equal blastomeres. Each of these divides again into two, and so on (fig. 1–4). Until the stage of 16 cells, the cells thus produced are approximately equal in size, but thereafter inequality develops, and the organism consists of an outer layer of small cells (trophoblast) surrounding a few relatively larger cells (inner cell mass). It is from the latter cells that the embryo proper (*i.e.*, the body of the new individual) is to develop.

On or about the fourth day, a little fluid begins to accumulate between the cells of the segmented ovum (fig. 5), which thus soon becomes a hollow sphere (blastocyst) with a wall one cell thick, except where the inner cell mass is heaped up at one side (fig. 6). Before this stage is reached, the clear protective zona pellucida has broken or dissolved away, and the embryo, which in man is less than $\frac{1}{50}$ in. in diameter, finds itself floating freely in the cavity of the uterus. It settles down upon the uterine lining and becomes fixed there, as shown in fig. 7, which is drawn from an early embryo of the rhesus monkey. The outer cells of the blastocyst on the side against the uterine lining now begin to enlarge and to increase rapidly in number. This thickened part of the trophoblast is able to erode and invade the cells of the uterus and thus to secure implantation of the embryo in the maternal tissue.

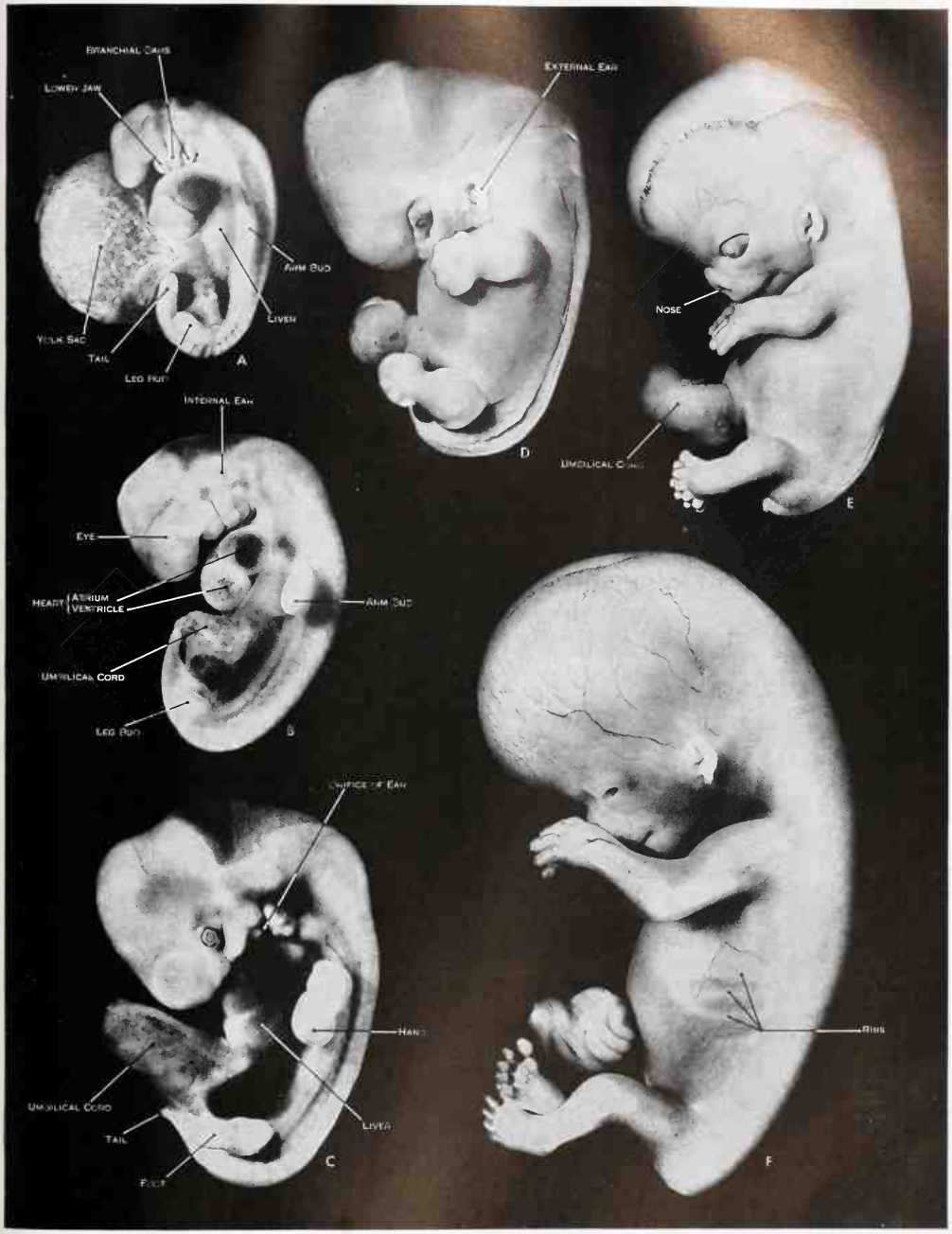
By the 8th day after ovulation, the human organism is still, practically speaking, a blastocyst, but, as shown in fig. 8, it is considerably modified from the simple form of the blastocyst (fig. 6). The trophoblastic attachment tissue is already quite extensive in comparison with the rest of the embryonic structures and is invading the maternal tissue. The blastocyst is no longer spherical; its thin free wall has collapsed against the attached side. The inner cell mass is now a somewhat flattened plate called the embryonic disc or shield. Between this embryonic shield and the trophoblast is a small new space, the beginning of the cavity



DEVELOPMENT OF THE MAMMALIAN EMBRYO

(Figs. 1–5, rabbit; figs. 6 and 7, rhesus monkey; figs. 8–17, human.) Figs. 1–4. Segmentation of ovum, × 150 (magnified 150 times). Fig. 5. Early blastocyst; half section, × 150. Fig. 6. Late blastocyst; half section, × 112. Fig. 7. Blastocyst fixed to uterine lining; half section, × 100. Fig. 8. Implanted embryo of 7½ days; half section, × 100. Fig. 9. Embryo of 13 days; half section, × 67. Fig. 10. Embryo of 23 days, embedded in uterus; ½ natural size. Fig. 11. Embryo of 18 days, in shield stage, with amnion, yolk sac and a part of trophoblast; ¾ view, × 16.5. Fig. 11a. Cross section of same embryo at level of arrows A–B, Fig. 11. Fig. 12. Embryo of 21 days; back view, × 14. Fig. 13. Embryo of 23 days with yolk sac, amnion opened; × 16. Fig. 14. Uterus with embryo of 6 weeks; ½ natural size. Fig. 15. Uterus with fetus of 3 months; ½ natural size. Figs. 16 and 17. Diagrams showing growth of amnion and formation of umbilical cord (stages correspond approximately to Figs. 10 and 14)

of the amnion. About the 13th day (fig. 9) the embryo is buried almost entirely in the uterine lining. The amniotic cavity has grown larger. On the underside of the embryonic shield, the endodermal cells (already seen at the blastocyst stage) now make a definite layer, destined to form the lining of the digestive tube of the embryo. At its margins the endoderm merges into the wall of a hollow structure, the yolk sac. This sac is all that the highest mammals possess in comparison with the relatively large and functional yolk sac of animals having yolk-rich eggs, *e.g.*, birds and reptiles. Outside these structures, but within the cavity of the ovum, there is a loosely arranged tissue (mesoblast) that forms a vaguely defined layer around the inside of the trophoblast and also over the outside of the amnion and yolk sac. The loose mesoblastic tissue forms a connection between the germ disc with its associated structures and the trophoblast. As the amniotic cavity enlarges, this connection persists in the region destined to be the caudal end of the embryo, forming the "body stalk" as indicated in fig. 10 and 11. A few days later, when the embryonic blood vessels develop, they will follow this mesoblastic route to the trophoblast in order to approach the maternal tissues to obtain nutriment for the embryo. Because of this direct connection via the body stalk, the human embryo does not require, for its nourish-

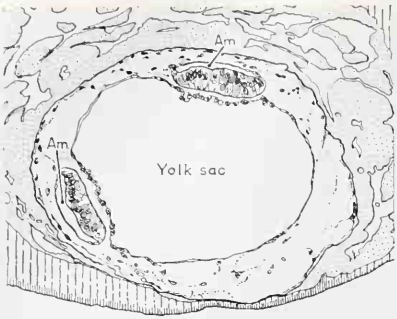


BY COURTESY OF THE CARNEGIE INSTITUTION OF WASHINGTON

HUMAN EMBRYOS

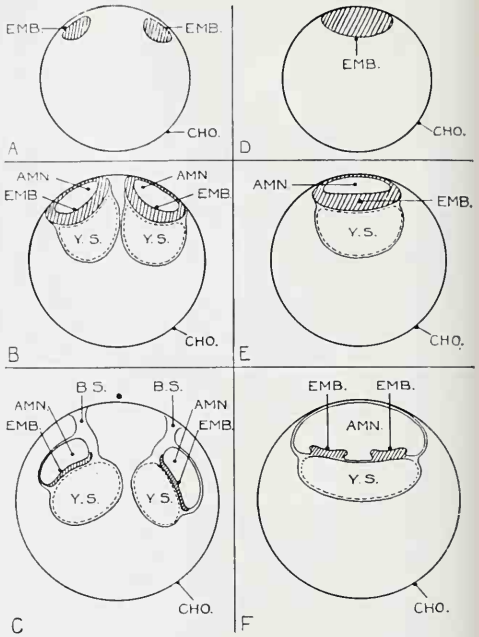
A, 27 days, x 10 (magnified 10 times). B, 29 days, x 8. C, 33 days, x 6.3. D, 37 days, x 5.4. E, 6 weeks, x 3.8. F, 8 weeks, x 3.3

EMBRYOLOGY, HUMAN

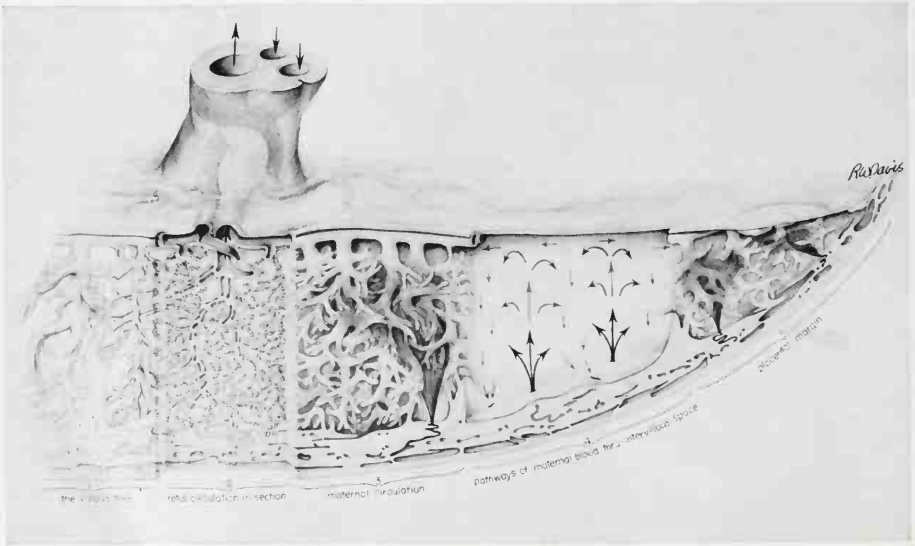


This diagram shows the probable developmental mechanism resulting in twin embryos with separate amnions enclosed in a common chorionic sac

Twin embryos, about 32 days old, within a single chorion but in separate amnions, each having its own yolk sac



These diagrams illustrate two types of single-ovum (identical) twinning in man: A, B, C, twinning by formation of two inner cell masses; D, E, F, twinning by formation of two embryos on a single germ disc



A composite drawing of the placenta to show its structure and circulation

MULTIPLE BIRTHS

BY COURTESY OF (TOP LEFT, TOP RIGHT) DR. G. W. CORNER, (TOP RIGHT A, B, C) "AMERICAN JOURNAL OF OBSTETRICS AND GYNECOLOGY," VOL. 70, P. 389, 1955, PUBLISHED BY THE C. V. MOSBY CO., (TOP RIGHT, D, E, F) BULLETIN OF JOHNS HOPKINS UNIVERSITY, (CENTRE LEFT) DR. W. R. M. MORTON, "JOURNAL OF ANATOMY," QUEEN'S UNIVERSITY, BELFAST, 1954, (BOTTOM) THE CARNEGIE INSTITUTION OF WASHINGTON

ment, participation of the yolk sac in its placental attachment nor the development of an allantois, as in many other mammals (*see* VERTEBRATE EMBRYOLOGY). The body stalk is ultimately modified, as the embryo grows, to form the umbilical cord.

Fig. 10, which shows the uterus and the embryo at the age of about 23 days, will serve to explain the next stage in the attachment of the embryo. The embryo, in its shell of trophoblast, is completely embedded in the lining of the uterus. At about the age of two weeks the trophoblast begins to put out rootlike processes, the villi, which grow until they form an extensive system of branching rootlets implanted in the uterine lining. The embryonic blood vessels of the embryo, which extend along the body stalk to the trophoblast, grow outward, always within the processes of the villi, and are thus brought into intimate relation with the maternal blood, which seeps through the uterine lining and among the villi. By this means nutritive substances are received by the fetus and waste products transferred to the maternal blood, as will be explained in greater detail below.

The embryonic disc of the 13th day (seen in section in fig. 9) is a round, featureless plate of cells. By 18 days it has become elongated (shield-shaped; fig. 11) and is marked by a shallow groove along the mid-line of its caudal end, overlying the primitive streak. The latter is a band of undifferentiated tissue generally considered comparable to the blastopore region of the amphibians (*see* VERTEBRATE EMBRYOLOGY). Like that region, it is a centre of growth from which more highly differentiated tissues arise, pushing forward as they grow. A cross section of the shield at this stage (fig. 11[A]) already shows the three germ layers of classical embryology: ectoderm on the surface, continuous with the amnion; endoderm lining the future gut and continuous with the yolk sac; and between them the mesoderm, continuous along the margins of the shield with the extra-embryonic connective tissue already referred to. Headward from the primitive streak is a small swelling, the primitive node. From this the notochord grows forward beneath the ectoderm, as in other vertebrates. As the shield grows larger, the primitive streak remains at the caudal end. In front of it, two long folds arise, forming a deep longitudinal groove in the mid-line of the shield. These neural folds progressively close together over the groove (fig. 12, 13) and ultimately become the spinal cord and brain.

It can be seen that the shieldlike embryonic body at the stage shown in fig. 11, if stripped of its membranes, *i.e.*, the amnion over its back and the trophoblast that envelops it, is directly comparable with the amphibian embryo at a similar stage (*see* VERTEBRATE EMBRYOLOGY). Disregarding for the moment the development of the membranes, which serve in mammals and man to facilitate intrauterine protection and nourishment of the embryo, it must be said that the development of the embryonic body is remarkably alike in all the five classes of vertebrates. For a general discussion of the development of the internal organs, *see* the article VERTEBRATE EMBRYOLOGY.

The changes of external form by which the embryo gradually acquires human appearance may be followed from the photographs on the plate. The ages cited in the following account are calculated from the time of discharge of the ovum from the ovary ("ovulation age"). At 27 days the embryo is about 5 mm. ($\frac{1}{5}$ in.) long, and under low magnification (*see* Pl. I [A]) exhibits a distinct head with barely visible optic and auditory vesicles (future eyes and ears) and a series of branchial bars. The heart, clearly seen through the transparent skin, is relatively large and already divided into the four chambers. The incipient liver makes a smaller prominence on each side at the level of the arm bud. The ventral body wall for the most part has closed in and the amnion (not shown in the photograph) now partly ensheathes the neck of the yolk sac and the body stalk, in which the umbilical arteries and veins run to and from the placenta. Thus the beginning of the umbilical cord has been made. Along the back, at the sides of the neural tube (spinal cord) and on into the tail, the row of somites is clearly seen, numbering at this time about 30 pairs. The first appearance of the arms is seen at 26 to 28 days as a little knob on each side of the body at the level of the 4th to 8th somites. At 33 days (*see* Pl. I [C]) the hand plate is distinguishable from the arm and soon bears ridges (*see* Pl. I [D]) that presage the separation into fingers. The leg bud appears about a day later than the arm (27 to 28 days) and follows the same plan of development. The rapidly growing brain causes a great enlargement of the head relative to the rest of the body. The eye, clearly seen at 29 days (*see* Pl. I [B]), becomes increasingly more conspicuous until the lids arise and gradually close over it. The lids fuse together in the ninth week but reopen during the seventh or eighth month. Rudiments of the external ear are visible toward the end of the fifth week (*see* Pl. I [C]) as a series of swellings or tubercles around the first branchial groove. The groove deepens to become the auditory orifice and the tubercles later blend to form the auricle. The nose is blunt and inconspicuous at first but the face with its bulging brow is unmistakably human early in the seventh week. The upper arm, forearm and hand may be distinguished before the embryo is five weeks old but as yet the leg shows only the foot and an undifferentiated segment uniting it to the body. At 5½ weeks (*see* Pl. I [D]) the liver, now large, swells out the abdomen. The heart no longer bulges prominently. The extreme caudal end of the embryo grows so slowly that the tail becomes less and less conspicuous and is finally concealed in the tissues of the rump, where only the coccyx at the tip of the spine remains as its relic. The external features of the human organism thus are established before the embryo is an inch long (*see* Pl. I [E]) and subsequent growth merely modifies the proportions without adding new surface structures.

The growth of the embryo is, at first, very rapid, but the rate of increase tapers off as pregnancy advances. The following table gives the average size of the embryo (crown-rump measurement or sitting height) and the average weight at the end of successive months, calculated from the beginning of the last menstrual period (for age from ovulation, deduct about two weeks).

Age in calendar months	Sitting height		Weight	
	Millimetres	Inches approximately	Grams	Pounds approximately
2	28	1.1	2.3	1/200
3	75	3.0	25	1/18
4	135	5.5	170	1/3
5	185	7.4	440	1
6	225	9.0	820	1¾
7	270	10.8	1,380	3
8	310	12.4	2,220	5
9	360	14.0	3,150	7

The heart begins to beat toward the end of the third week. The voluntary muscles are able to contract in response to external stimuli (touch or pin prick) after the eighth week and spontaneous movements may begin as early as 9½ weeks. The average duration of pregnancy is about 280 days (40 weeks), calculated from the first day of the last menstrual period, or about 265 days (38 weeks) from ovulation and fertilization of the ovum. It may extend to 300 days or even more, in which case the infant tends to be larger and heavier than the average. Infants born before 6½ months (27 weeks) are rarely viable; from 6¾ to 7 months they may survive with good care, and after 7 months (30 weeks) they are generally viable.

Membranes and Placenta.—As shown in fig. 9 to 15 the cavity of the amnion enlarges over the back of the embryo and finally spreads around it on all sides. This growth is illustrated diagrammatically in fig. 16 and 17. Thus the yolk sac is squeezed against the body stalk and the latter is narrowed into a cord (umbilical cord) connecting the belly of the embryo with the inside of the chorion (the outer layer of the embryonic mass called trophoblast in the earliest embryos). Ultimately (three months) the amnion unites with the chorion. Within the amnion the infant is supported and protected from physical shock by the amniotic fluid.

From the end of the second week, the embryonic mass, which is still quite small, lies embedded under the surface of the uterine lining, bulging toward the cavity of the uterus but separated from it by the surface of the lining. The chorionic villi now grow rapidly and burrow on all sides into the lining of the uterus (fig. 10, 14, 15). As previously explained, blood vessels extend to and from the embryo along the body stalk into these processes. As the chorion and its contents grow, the uterine cavity is practically obliterated. The chorion grows smooth over the bulging side of the mass, losing its villi. Only the basal area chorion remains (fig. 15); here the villi persist, grow larger and branch even more freely, forming a thick discoid organ of attachment, the placenta (*q.v.*). The nature of the placenta can be made clear by comparing it to the roots of a tree, for its villi ramify in the space it makes for itself by eroding the uterine lining. Blood vessels supplying the uterine wall are opened by the spread of the placental tissue and they exude blood into the space between the tightly packed placental roots or villi. Fine capillary blood vessels forming the terminal branches of the umbilical vessels of the embryo are thus caused to dip into a sluggish pool of maternal blood. Embryonic blood coursing in these capillaries is thus separated from the maternal blood only by the thin walls of the embryonic capillaries and the thin layer of covering cells of the villi. Oxygen, nutritive chemical substances (carbohydrates, nitrogenous foods and mineral salts) in the mother's blood can readily pass by filtration into the embryonic blood circulation, and the waste products of the embryo (carbon dioxide, urea, etc.) can pass outward into the mother's blood to be carried away and disposed of through the mother's lungs and kidneys. The placenta is thus an organ of respiration, nutrition and excretion for the developing infant. By virtue of this arrangement of the membranes and placenta, the developing organism is shut off from the outer world and lives in a dark, fluid-filled space, receiving from its mother only such water-soluble substances as can be transported in solution in the blood and can pass the barrier of the placenta under the influence of ordinary chemical and physical forces. There is no connection whatever between the mother's nervous system and that of the infant and no means by which ideas or impressions in the mother's mind can reach the child.

Multiple Births.—Twins occur about once in 80 pregnancies. They may originate from two ova shed at one time, or from one ovum that gives rise to two embryos. In the former case (which includes about two-thirds of the cases of twinning) the two offspring are no more closely related than any pair of children of the same parents. They may be of opposite sex and may not show close resemblance. Single-ovum twins are necessarily of the same sex and usually resemble each other very closely, even to colour of hair and to fingerprint patterns. These form about one-third of all cases of twinning. During gestation, single-ovum twins are attached by one placenta and occupy a single chorionic cavity; double-ovum twins have separate placentas and chorions, which may, however, fuse. Triplet births occur about once in 6,400 cases. (*See* MULTIPLE BIRTHS.)

Pathology.—The human embryo, like all other living things, is subject to disease and abnormality of growth. This statement need not cause undue alarm to expectant parents, for once the pregnancy is well enough established to be recognized, development proceeds normally in the majority of cases. Most of the embryonic mortality occurs very early, before or during the stage of implantation, *i.e.*, during the first 20 days, and therefore passes unnoticed in most cases. Probably fewer than two out of every three fertilized ova go on to birth at full term. Embryonic abnormalities are of various origins. Many of them are probably due to genetic causes. The combination of genetic factors derived from the parents is in part subject to chance and some embryos therefore do not begin their careers with an adequately balanced set of constitutional characteristics. Similar mischance in general does not occur to other offspring of the same parents. Defective health of the mother is in certain special circumstances a cause of death or malformation of the embryo. There are certain infectious diseases of the mother, notably syphilis, that may injure the child *in utero*. Malformations of the mother's internal reproductive organs, or inadequate functioning of the endocrine system, may provide an unfavourable environment for the embryo. Certain types of incompatibility of the blood groups of the parents are responsible for prenatal disease of the infant (*see* FETAL DISORDERS: *Erythroblastosis Fetalis*).

The scientific study of embryonic defects has abolished the mystery and superstition that formerly surrounded the birth of abnormal infants, especially those with obvious and extreme malformations, and has disproved the ancient belief that physical injury, repulsive sights or horror affecting the mother may cause deformity of the child.

See also references under "Embryology, Human" in the Index.

BIBLIOGRAPHY.—William John Hamilton *et al., Human Embryology* (1952); G. W. Corner, *Ourselves Unborn* (1944), a semipopular account; J. C. Brash, *Human Embryology* (1957); F. D. Allan, *Essentials of Human Embryology* (1960); G. L. Streeter, "Weight, Sitting Height . . . and Menstrual Age of the Human Embryo," Pub. Carnegie Inst. Washington, *Contrib. to Embryol.*, XI (1920), "Developmental Horizons in Human Embryos," Pub. Carnegie Inst. Washington, *Contrib. to Embryol.*, XXX (1942) and subsequent volumes. Many other valuable articles are found in the Carnegie *Contributions*.

(G. W. CR.; C. H. HR.)

EMBRYOLOGY, INVERTEBRATE: *see* INVERTEBRATE EMBRYOLOGY.

EMBRYOLOGY, VERTEBRATE: *see* VERTEBRATE EMBRYOLOGY.

EMBURY, PHILIP (1729–1775), one of the earliest of the chief founders of Methodism in America, who organized the first Methodist society in New York city, was born in 1729 probably in Ballingrane, County Limerick, Ire. His German Protestant parents had fled from the Palatinate under the persecutions of Louis XIV. Embury was converted through the instrumentality of John Wesley on Christmas day, 1752. His qualifications for religious leadership were soon recognized and he was appointed class leader and subsequently licensed as a local preacher. In company with a few others he emigrated in June 1760 to America, settling in New York city. For some time he seems not to have engaged in any religious activities, but in 1766 he began to preach. His first sermon was delivered to a company of five persons, whom he organized into a class. When his audience outgrew his living room he first rented a larger room and later a rigging loft. In two years' time the congregation had grown sufficiently to justify the building of a church, Wesley chapel, the first John Street church. On Oct. 30, 1768, Embury—who was pastor, a trustee and treasurer—preached the dedicatory sermon. In 1770 he organized at Ash Grove, near Camden, Washington county, the first Methodist society north of New York city. He died in Camden, N. Y., in Aug. 1775.

See Matthew Simpson (ed.), *Cyclopaedia of Methodism*, pp. 338 ff.; biography by Harris Elwood Starr in Allan Johnson and Dumas Malone (eds.), *Dictionary of American Biography*, vol. vi (1931).

(W. C. B.)

EMDEN, ROBERT (1862–1940), Swiss physicist and astrophysicist noted for his development of the theory of polytropic gas spheres and of its application to stellar structure. Born at St. Gallen on March 4, 1862, Emden studied mathematics and physics at Heidelberg, Berlin and Strasbourg, and was appointed in 1899 to the Technische Hochschule, Munich, where he became professor of physics and meteorology (1907) and honorary professor of astrophysics in the university (1928). He remained there until his retirement to Zürich in 1934.

Almost all of Emden's work concerns the application of thermodynamics to natural phenomena, mainly in geophysics and astrophysics. His famous book *Gaskugeln* (1907) was based on a course of lectures on the applications of thermodynamics. He maintained that stellar constitution and atmospheric circulation provided unusually helpful examples in teaching the theory of physics, and he also used them to encourage the use of numerical computation. Emden developed the theory of a self-gravitating gas sphere (first stated by J. H. Lane and A. Ritter and independently by Lord Kelvin) into a logical and systematic survey. He was thus able to give numerical solutions for a number of homologous series based on different values of the polytropic index n, of which the convective model ($n = 3/2$ for a monatomic gas) and

Eddington's standard model ($n=3$) are special cases. Emden took a leading part in the foundation of the *Zeitschrift für Astrophysik* (1930), which he edited until 1936. He died at Zürich on Oct. 8, 1940. (B. E. J. P.)

EMDEN, a seaport in the extreme northwest of Germany which after partition of the nation following World War II was included in the *Land* (state) of Lower Saxony, Federal Republic of Germany. It lies on the Ems estuary 112 km. (70 mi.) W.N.W. of Bremen. Pop. (1959 est.) 44,238. Despite much destruction during the war the town retains a historical character. Ramparts dating from 1616, with a moat, stand and also two ancient docks or "delfts." There are remains of the late Gothic "Great church" (1648). In the new town hall are the East Friesland district museum, the town armoury from the 16th–17th centuries, the city archives and library. The Reformed Great church also contains valuable printed documents of the Reformation period. The duty-free seaport of Emden is the third port of Germany and consists of three locked docks and a tidal harbour. It is the home port of large numbers of seagoing vessels and harbours also herring fleets and barges. It is connected by the Dortmund-Ems canal with the German inland waterway system, and by smaller canals with the surrounding district. Railways run to Bremen and southward to the Ruhr. There is an airport for services to the East Frisian Islands. Beside the important transshipping industry there are shipyards and dry docks, an oil refinery, a large power station and fish packing and canning plants. The main exports are coal, ore and grain. Emden was founded about A.D. 800 as a port and later developed trade with the Baltic countries and Russia. In the 15th century it became the capital of the county of East Friesland and was privileged with staple rights by which it dominated the Ems trade. During the Dutch wars of independence in the 16th century Emden gained much of Holland's trade and became for a time the most important northwestern European port. Fugitives from the Netherlands made Emden the centre of Calvinism in northern Germany. Their return, however, new Dutch policies and the development of a new course of the Ems farther from the town led to a decline in the 17th century, but the neutrality of Emden in the 18th-century naval wars brought prosperity once more. It became Prussian in 1744, was occupied by the French 1810–14 and passed to Hanover in 1815. After the port was connected to the Ruhr by the Dortmund-Ems canal in 1898, it became the German seaport of Westphalia. Canadian troops captured the city in April 1945. (Wo. S.)

EMERALD, a bright-green variety of beryl, highly valued as a gem stone since about 2000 B.C. It was intensively mined from Egyptian deposits during the times of Alexander the Great, Cleopatra and the Roman empire. Numerous excellent stones, probably originating from deposits in Colombia, were taken from the Peruvian Indians during the Spanish conquest of South America.

The formula ($Be_3Al_2Si_6O_{18}$) and properties of emerald are essentially those of common beryl. The green colour is due to the presence of a little chromium in substitution for aluminum. Well-faced crystals are common, generally in the simple form of stubby to slender prisms. Many can be split into plates along an imperfect basal cleavage.

Emerald is slightly harder than quartz, softer than topaz and considerably softer than corundum and diamond. The specific gravity is 2.64 to 2.72 for most specimens. The refractive indices and dispersion are low to moderate, so that cut stones show little brilliance or fire. Their high value is based instead upon colour, transparency and freedom from flaws. Wholly unflawed material of deep-green colour is essentially unobtainable in large pieces.

BY COURTESY OF AMERICAN MUSEUM OF NATURAL HISTORY

EMERALD, BRIGHT GREEN VARIETY OF BERYL: FROM THE CHIVOR MINE, COLOMBIA

An accessory constituent of some granites and pegmatites, emerald also occurs in veins and in masses of schist, shale, limestone and other rocks that have been invaded by pegmatitic fluids (*see* PEGMATITE). Associated minerals include quartz, alkali feldspars, muscovite, carbonates and sulfides, as well as chrysoberyl, phenakite, parisite and other uncommon species. Gem emerald of finest quality comes from mines at Muzo, Cosquez and Somondoco, in Colombia. Other important localities are in the Ural mountain region of the U.S.S.R., in Bahia, Braz., and in Australia, Austria, Egypt, Norway, South Africa and the United States (North Carolina).

Nearly all commercial emerald is facet-cut into gems that are square or rectangular in plan. Ideally, cuts are so oriented that the deepest possible colour is displayed uniformly, with imperfections positioned as inconspicuously as possible. In earlier times, the Romans and peoples of the far east used hexagonal cleavage wafers in their jewelry, often in pierced and strung form. Irregular pieces also were fitted into various mosaics.

Many virtues were formerly ascribed to the emerald. It was reputed to have medicinal value, and because of its colour was said to be good for the eyesight. When worn, it was held to be a protection against epilepsy; it cured dysentery; it assisted women in childbirth; it drove away evil spirits; and it preserved the chastity of the wearer. Because of the presumed religious significance of its colour as a symbol of immortality, it was widely sponsored by the church.

In 1930 emerald was synthesized in Germany, and small crystals were "grown" a few years later in the United States. The procedure involves very slow crystallization, probably through vapour transfer of material from a melt at high temperatures and pressures. The method employed in the United States has been much improved, and crystals weighing 200 g. or more can be grown. The material of best quality matches the properties of natural emerald almost perfectly. *See* GEM; AQUAMARINE; BERYL. (R. H. J.)

EMERGENCE, a term used in evolutionary theory to mean the rise of a system that cannot be predicted or explained from antecedent conditions. Conway Lloyd Morgan borrowed the term from G. H. Lewes, who, in his *Problems of Life and Mind*, 2nd series (1875), distinguished between "resultants" and "emergents"—phenomena that are predictable from their constituent parts, and those that are not (*e.g.*, a physical mixture and a chemical compound). The evolutionary account of life is a continuous history marked by stages at which fundamentally new forms have appeared: (1) the origin of life; (2) the origin of nucleus-bearing protozoa; (3) the origin of sexually reproducing forms, with an individual destiny that cells reproducing by fission lack; (4) the rise of sentient animals, with a nervous system and proto-brain; (5) the appearance of an animal knowing all this and more, namely man. Each of these new modes of life, though grounded in physicochemical and biochemical conditions of the previous and simpler stage, is intelligible only in terms of its own ordering principle. These are cases of emergence.

Lloyd Morgan emphasized the negative character of the principle: we should describe nothing as emergent unless we can show that it cannot be resultant (*cf. Emergent Evolution*, Gifford lectures, 1922). Like Lewes, he treated the distinction as inductive and empirical, not as "metempirical" or "metaphysical." He condemned Bergson's conception of "creative evolution" as speculative, while proclaiming "emergent evolution" as a scientific theory. Still, it has not been widely accepted by biologists; by the 1960s most works on evolution (especially those written in English) were avoiding the term. With genetics illuminating the mechanism of heredity (and hence the very conditions of evolution), and biochemistry elucidating the workings of the cell nucleus, many biologists are confirmed in their belief that "scientific" treatment admits only analysis into parts, not new kinds of wholes. Thus they tend to concentrate on the mechanisms of mutation and natural selection, effective in "micro-evolution," the change from variety to variety and species to species; and to extrapolate these to "macro-evolution," the origin of the great groups of living things.

Nevertheless, the concept of emergence still figures in evolutionary thinking. Some writers, such as the paleontologist P. Teilhard de Chardin or the zoologist A. Vandel, emphasize the series of levels of organization, moving toward higher forms of consciousness. A. N. Whitehead's philosophy of organism, with its doctrine of "creative advance," is a philosophy of emergence;

so is M. Polanyi's theory of personal knowledge, with its levels of being and of knowing, none intelligible wholly in terms of those below it. This analysis shows that even machines cannot be understood from their parts alone, leaving aside their operational principles; this holds all the more for living as against nonliving matter, or for higher as against lower forms of life. (M. G. Gr.)

EMERSON, RALPH WALDO (1803–1882), U.S. essayist, poet and philosopher; his philosophy is perennial because it arose from a desire to make all things new, to seize life freshly, experience it firsthand and use it to build beyond the old.

Education and Pastorate.—By ancestry, education and vocation, if not by wealth, Emerson, born in Boston, Mass., May 25, 1803, could have claimed his place in the Brahmin class of Boston and Cambridge; he could have settled securely among these guardians of the wealth and *status quo* of New England. Like all his ancestors before him—from Peter Buckeley, who helped to found Concord, Mass., to his father, William Emerson, minister of the First Church, the oldest church in Boston—he went into the ministry, having first prepared for this vocation at Harvard college. Though he early lost his father—a cosmopolitan, social and literary minister who had been drifting from orthodoxy to Unitarianism—Emerson had managed by means of scholarships, teaching, and the devoted economies of his mother, to take his degree in 1821 at the age of 18. Then by means of teaching and preaching he finished a protracted and desultory study of divinity at the Harvard divinity school. Dogged by illness and driven by ambition, he was finally ordained as the minister of the Second Church of Boston in 1829. Shortly afterward he married Ellen Tucker of New Hampshire, settled in the heart of New England and began to win fame as a preacher, second perhaps only to the celebrated William Ellery Channing, the spiritual leader of the Unitarians. His position was secure and his first struggle with adversity seemed to have been won.

He had, however, already attended to the new and stirring voices of "the times"—and when he later spoke what they prompted him to say his discourse was soon branded as "the latest form of infidelity." He had found Carlyle and Coleridge and had heard and read the local Swedenborgian druggist, Sampson Reed, whose *Observations on the Growth of the Mind* (1826) had challenged the reigning psychology of John Locke. Emerson's older brother William, who had gone to Germany to study theology and returned to study law, not only brought him news of Goethe, but of the new biblical criticism and the doubts that had been cast on the historicity of miracles. His own sermons, from the first, had been unusually free of traditional doctrine, a personal exploration rather of the uses of spirit, showing an idealistic tendency and announcing his personal doctrines of self-reliance and compensation. Indeed, his sermons had divested Christianity of all external or historical supports and made its basis one's private intuition of the moral law and its test a life of virtue and character. Slowly he was discovering what he later called "the famine of our churches," the failure of the churches to nurture the moral nature of man. Always an opponent of forms, he chose to make the rite of communion, which he was unwilling to administer, the test of his demand that he be free in his ministry. If he had been called to the ministry chiefly by his desire to wield the sceptre of eloquence and by a genuine spiritual ardour, he could not meet its routine pastoral demands. The weekly sermons and inevitable civic duties overtaxed him, and, "chilly" by nature, he did not succeed in consoling his parishioners on pastoral calls. Sceptical of the traditional faith, he wanted a freer and larger sphere of influence, and when Ellen Tucker died in 1831, he quickly—unnecessarily and impulsively as some thought—brought his ministry to an end by demanding more than his congregation would grant. Then he was free at last to follow an ambition that had been misdirected into the ministry—". . . in the uncouth solitude [to] unlock. . . . My stock of art. . . ."

Leader of the Transcendentalists.—The 1830s saw Emerson's fall from traditional greatness and his rise in a new form—a form that he made possible for others by his own daring example and heroic struggle, that of the independent literary man. During this decade his own personal doubts and difficulties were increasingly shared by other spiritual seekers. Before the decade was over his personal manifestoes—*Nature,* "The American Scholar" and the divinity school *Address*—had rallied together a group that came to be called the Transcendentalists, a loosely knit group, nothing like a party or a movement, of which he was popularly acknowledged the spokesman. The Transcendentalists, as he defined them, were simply "the party of the Future," the partisans of hope and of the "newness," fighting the immemorial battle of history against "the party of the Past." What distinguished these parties, however, was their philosophies; "mankind have ever divided into two sects, Materialists and Idealists," Emerson said. What was called transcendentalism in New England was only idealism reasserting itself. If he intended to mollify by these definitions, the philosophies were still unalterably opposed. Materialists founded on "experience," the idealists on "consciousness"; one began with the data of the senses and insisted "on facts, on history, on the force of circumstances and the animal wants of man"; the other began with the facts of consciousness and insisted "on the power of Thought and of Will, on inspiration, on miracle [intuitive perception], on individual culture." Brought closer home to the actual issues of the times, the party of the Past tended to unite in the belief in the innate depravity of man and championed the need for tradition and restraint, while the party of the Future championed the need for change and freedom. The party of the Future, accordingly, was the party of social reform, armed with a philosophy which would make all things new. Before its work was done, it had not only demolished the foundations of the social and institutional structure its fathers had built, and opened up the possibilities of a unique American culture, it had also relit the spiritual faith which its fathers had let die out. By replacing a defunct Christian myth with the myth of nature, it made primary religious experience available to all men; it took religion into the open air, naturalized it and democratized it.

When Emerson left the church in 1832, he was in search of a more certain conviction of God than that granted by the historical evidences of miracles. He wanted his own revelation, a direct and immediate experience of God—the presence of spirit within him that would strengthen his own conduct of life. He needed a new philosophy that would certify this experience and, undertaking a new life, he needed a model of the great man to hearten him in his endeavour. In 1829, at the beginning of his ministry, he had renewed his reading in Coleridge, and when he left his pulpit he went to Europe. There he saw Jussieu's collection of natural specimens arranged in a developmental order that confirmed his belief in man's spiritual relation to nature. In England the memorable events were his visits to Coleridge and Wordsworth, and especially to Carlyle. At home once more in 1833, he began to write *Nature* and established himself as a lecturer, beginning with courses in natural history and biography. By 1834 he had found a permanent dwelling place in Concord, and in the following year he married Lydia Jackson of Plymouth, Mass., and settled into the kind of quiet domestic life that was essential to his work. Travel and lecturing and an ever-enlarging tide of visitors were the only interruptions to the rural peace that he had wisely gathered as a defense around him; but from this sanctuary, he soon launched an attack that, for all of its seeming loftiness and serenity of utterance, taught his generation to "Beware when the great God lets loose a thinker on this planet." For then, he explained, "all things are at risk. It is as when a conflagration has broken out in a great city, and no man knows what is safe, or where it will end."

Publication of "Nature."—Emerson kindled that conflagration—or, more accurately, began a revolution in thought—by publishing anonymously in Boston in 1836 a little book of 95 pages entitled *Nature.* Having found the answers to his spiritual doubts, he formulated his essential philosophy, and everything he ever did was an extension, amplification or amendment of the ideas he first affirmed in *Nature.* The reasons for his doubts had been deeper than the Unitarian's retention of the beliefs in the historicity of miracles and in the personality of Jesus; for these only pointed to the philosophy that underlay the supposedly rational

Unitarian faith. With many of his generation, he could no longer find the living spirit in the "pale negations" of the church of his fathers. He had been unsettled and paralyzed by the mechanistic conception of the universe (Newtonian physics) and the psychology of "sensation" (the psychology of John Locke) that he had learned at Harvard and upon which the Unitarians had built their faith. There was no place for the piety and spirit of the Puritans in that machinelike world of matter in motion; one could not believe in free will in a world bound in the iron chains of cause and effect and known only passively by the sensations that came to him from without. Such a world, Emerson felt, was alien and external (known only in sensation, it was not real, but phenomenal) and determining men physically and psychologically, it made them victims of circumstance, men bereft of reality and the active powers of mind to seize it. Against this philosophy, their inheritance from the 18th century, and especially as it was embodied in the Unitarian Church, others in Emerson's generation had rebelled by 1836. In fact, in that memorable year Orestes Brownson published his *New Views of Christianity, Society, and the Church*, William Henry Furness his *Remarks on the Four Gospels*, George Ripley his *Discourses on the Philosophy of Religion Addressed to Doubters Who Wish to Believe* and Bronson Alcott his first report on the new methods of instruction that he was applying at the Temple school. What made Emerson's treatise more significant, however, was that it was truly more radical. It went to the roots, directed not immediately at the religious issues around which the conflict raged, but rather at the philosophy of man and nature, the deeper foundations of psychology and physics upon which faiths are reared.

Nature reclaimed an idealistic philosophy and put nature once more at the moral service of man. Still more important, it opened the channel through which the new romantic concept of an organic nature flooded American thought. Many men, like William James, John Jay Chapman and Oliver Wendell Holmes, were to testify in behalf of the liberation they found in Emersonian self-reliance and individualism—in the right to fashion oneself from the materials of life and to follow new vocations better fitting one's needs. Others, like John Dewey, who believed Emerson ranked with Plato, appreciated the democratic element in his thought and his faith in the power of ideas to reshape matter. Still others, like Henry James, saw in art a way to vindicate Emerson's great faith in perception. Yet the heart of Emerson's vision, as the title of his first book suggested, was the organic concept, a belief in a living, pulsing nature still in the process of becoming, a nature immanent with spirit, fulfilling the ideas of God and capable also of fulfilling the ideas of man. For in this Emersonian conception, man was not estranged from nature, but intimate with her, sharing the flow of her spiritual tides, and able, by means of his own creative powers of perception, to grasp her law and, thereby, with his ideas, to be himself a creator of the as-yet-unformed future. No philosophy ever spoke more grandly for the creative self-sufficiency, even the Godlike power, of man.

The essential and liberating idea of *Nature*, to be applied in Emerson's later work to all the departments of human life, was that by his own constitution and by the corresponding constitution of the universe, man was not the victim of his environment. Mind, not matter, was supreme. Ideas were sovereign, and with them as his instruments man could shape the universe to his needs. Emerson was intent on showing his generation how to see—how, as he said, to behold "God and nature face to face," how to "enjoy an original relation to the universe" and, accordingly, how to "have a poetry and philosophy of insight and not of tradition, and a religion by revelation to us, and not a history of theirs." In *Nature*, he showed men how they might conquer with perception, relax the despotism of the senses and unfix the world of matter; how they might be liberated and empowered if they translated matter into mind, if, by the process of symbolization, they turned everything external into an inner possession of consciousness. Moreover, he showed them how willing nature was to serve the moral ends of man, how bountifully and beneficently she supplied them on every level, from the physical to the spiritual. "As a plant upon the earth," he explained in a metaphor of organic dependence, "so a man rests upon the bosom of God; he is nurtured by unfailing fountains, and draws at his need inexhaustible power." "Who," he asked, "can set bounds to the possibilities of man?" For in the act of perception, when man is filled with the influx of spirit, "we learn," he said, "that man has access to the entire mind of God, is himself the creator in the finite."

Here, then, was a book expressly designed to teach men, as one reviewer appreciatively remarked, "to look on the world with new eyes." And this, of course, was its danger. The older generation and the orthodox quickly mounted the ramparts of the religious periodicals against this "newness," led by the Reverend Andrews Norton, professor of divinity at Harvard and the acknowledged "pope" of the Unitarians. At first he denounced these transcendental ideas as a German importation. Later, when they were brought to bear on the reform of the ministry itself in the divinity school *Address* (1838), his wrath grew boundless. Emerson, perhaps, had not fully measured the strength of the opposition. Galvanized by his own faith in the power of ideas, he had quickly applied the doctrines of *Nature* to the reigning institutions of New England, to the college and the church, the seminaries of leadership. In the previous year, he had delivered the Phi Beta Kappa address at Harvard, and in "The American Scholar" had described the duties and resources of the new, liberated intellectual that he had himself become. In effect, he put the "scholar" above the priest, made him the servant of truth whose duty was "the conversion of the world." Not only did he call men to a new individual vocation, he showed them how in nature, without the traditional institutions of culture, they might unfold their genius.

The divinity school *Address* of 1838, however, arrested Emerson's first attempt at reform, or rather it taught him that ideas, however provoking, did not reform things by themselves. Individualistic and constitutionally ill-equipped for public strife, Emerson did not, like Ripley, Brownson, Theodore Parker and other Transcendentalists, engage in the active reforms of the 1840s and 1850s. He remained detached, however, only to be free to speak the truth, and by his example to show the efficacy of the only kind of reform he valued, that of individual, moral regeneration. If not a doer, he could be a sayer, and during the 1840s he discharged the best of his thought.

The Essays.—In 1841, he published his *Essays, First Series*. These essays, consolidating his thoughts on moral individualism from previous lecture series, preached the ethics of self-reliance, the duty of self-culture and the need for the expression of self. Revolving around "Self-Reliance"—the most personal and challenging essay—the essays spoke for the authentic man, for the man who would live his own life, for the man "who has ventured," he said, "to trust himself for a taskmaster."

Essays, Second Series (1844) revealed Emerson's growing awareness of the limiting conditions of life. Treating the same themes, he was, however, more concerned with the relation of his doctrines to actual life and was faithful to his own experience of limitation. In "The Poet," for example, he confessed his failure to be the active great man he had defined in "The American Scholar." In "Experience," he let the reader into the recesses of his own scepticism. In "Nature," he described a titanic force that served its own ends rather than man's. During the lecture season 1845–46, he spoke on *Representative Men* (which appeared as a book late in 1849), and here again the modification of his defiant self-reliance was evident. Always a devoted reader of biographies, he had lectured on biography in 1835, and had extolled, among others, Martin Luther, John Milton and George Fox. If his *Representative Men*—Plato, Swedenborg, Goethe, Shakespeare, Napoleon and Montaigne—were more catholic choices and more representative of the sources to which his thought was tributary, there was nevertheless less insistence on the test of action and more on the test of perception. A poet from his youth, he published his remarkable *Poems* in 1846, and in 1847–48, as evidence of his growing fame, he was called on by radical groups to lecture in England.

At home, moreover, his reputation was secure, for time was wearing down the novelty of his rebellion. Slowly he had accommodated himself to society. Lecturing more frequently, he had

become an institution; and under the pressure of these engagements he began to draw on old materials and hammer them thin. If the drift of his thought had not been apparent in the 1840s, his lectures on *The Conduct of Life*, given as early as 1851 and published in 1860, at last made it clear that the greatest spokesman for the rights of spiritual manhood was not opposed to society. The growing crisis over slavery, in fact, had called him from his solitude. In time he was even called to serve Harvard as an overseer, to lecture on philosophy there and to mount the pulpit of the divinity school. Except for *May-Day and Other Pieces*, poems he published in 1867, the writing he did after 1860 was marked by waning powers. A new generation knew only the old Emerson and had absorbed his teaching without recalling the acrimony it had occasioned. Emerson upon his death April 27, 1882, was transformed into the Sage of Concord, shorn of his power as a liberator and enrolled among the worthies of the very tradition he had set out to destroy.

Essence of His Philosophy.—Emerson's philosophy was one of many attempts in the history of thought to destroy an ironclad, deterministic universe—to find again an open and novel universe, a field for new enterprises. Like so much of American thought, which it helped to shape and to which it first gave expression, Emerson's philosophy was radical, exploratory and optimistic, predicated not only on the power of man to change things but on the goodness of change itself. Based on an organic or vital rather than a mechanical conception of the universe—on growth and development rather than fixity—and based on a reassessment of man that restored his creative powers, it was eminently a philosophy of renewal which spoke immediately to Emerson's generation. For that generation, still living in the afterglow of the American Revolution and fired again by the War of 1812, was fervidly self-conscious about the relation of America to Europe and the past; it was determined to vindicate America by building a culture better than that of the old world. Wanting to redeem a tired and decadent world from the decay of centuries by starting afresh, that generation needed a philosophy that would not only help it overcome the servility to tradition but teach it how to use the ever-new materials of life—the resources of nature. This Emerson provided.

BIBLIOGRAPHY.—The definitive biography of Emerson is Ralph L. Rusk, *The Life of Ralph Waldo Emerson* (1949). Bibliographies are available in Frederic I. Carpenter, *Emerson Handbook* (1953); Stephen E. Whicher, *Freedom and Fate: An Inner Life of Ralph Waldo Emerson* (1953); and in the evaluative bibliographical essay on Emerson in Floyd Stovall (ed.), *Eight American Authors: A Review of Research and Criticism* (1956). Other biographical studies are James E. Cabot, *A Memoir of Ralph Waldo Emerson* (1887); George E. Woodberry, *Ralph Waldo Emerson* (1907); Oscar Firkins, *Ralph Waldo Emerson* (1915). Critical studies are F. B. Sanborn (ed), *The Genius and Character of Emerson* (1885); Henry D. Gray, *Emerson: A Statement of New England Transcendentalism as Expressed in the Philosophy of Its Chief Exponent* (1917); Paul Sakmann, *Emerson's Geisteswelt* (1927); F. I. Carpenter, *Emerson and Asia* (1930); Bliss Perry, *Emerson Today* (1931); Vivian C. Hopkins, *Spires of Form: A Study of Emerson's Aesthetic Theory* (1951); Sherman Paul, *Emerson's Angle of Vision: Man and Nature in American Experience* (1952). The standard editions of Emerson are Edward W. Emerson (ed.), *The Complete Works of Ralph Waldo Emerson* (1903–04); E. W. Emerson and W. E. Forbes (eds.), *The Journals of Ralph Waldo Emerson* (1909–14); Ralph L. Rusk (ed.), *The Letters of Ralph Waldo Emerson* (1939).

(S. PL.)

EMERY is an impure variety of corundum or aluminum oxide (alumina), much used as an abrasive or polishing substance, and occurring as a granular or massive, dark-coloured, dense substance, having much the appearance of an iron ore. Under the microscope it is seen to be a mechanical aggregate of corundum, usually in grains or minute crystals of a bluish colour, with magnetite, which also is granular and crystalline. Other iron oxides, like hematite and limonite, may be present as alteration products of the magnetite. Some of the corundum (Al_2O_3) and iron oxide may occasionally be chemically combined, so as to form an iron spinel, or hercynite. In addition to these minerals emery sometimes contains diaspore, gibbsite, margarite, chloritoid and sillimanite. Indeed, emery must be regarded as a rock rather than a definite mineral species.

The hardness of emery is about 8, whereas that of pure corundum is 9. Its abrasive power, or effective hardness, is not proportionate to the amount of alumina which it contains, but seems rather to depend on its physical condition. Its specific gravity varies with its composition from 3.7 to 4.3.

Emery has been worked from a very remote period in the island of Naxos, Greece, one of the Cyclades, whence the stone was called *naxium* by Pliny and other Roman writers. The mineral occurs as loose blocks and as lenticular or lens-shaped masses or irregular beds in granular limestone, associated with crystalline schists, and has been found to contain 52.4% of corundum, 32.1% of magnetite, 11.5% of tourmaline, 2% of muscovite and 2% of margarite.

Important deposits of corundum were discovered in Asia Minor by J. Lawrence Smith, when investigating Turkish mineral resources about 1847. The occurrence is similar to that in Naxos. The emery is found as detached blocks in a reddish soil, and as rounded masses embedded in a crystalline limestone associated with mica schist, gneiss and granite. It is believed to have originated by regional or contact metamorphism of bauxite deposits.

Emery has been worked at several localities in the United States, including one near Chester, Mass. In the second half of the 20th century the only commercial production, which increased rapidly after 1940, was at Peekskill, N.Y.

Long used principally as an abrasive, the largest application of emery became that of a nonskid material in floors, stair treads and pavements. A very fine emery dust is used by lens grinders, lapidaries and plate-glass manufacturers, although other abrasives often are preferred.

Emery wheels are made by consolidating the powdered mineral with a bonding medium, and are used not only by dentists and lapidaries but also in shops for grinding, shaping and polishing steel. Emery sticks, emery cloth and emery paper are made by coating the several materials with powdered emery mixed with an adhesive medium. *See* CORUNDUM; CORUNDUM, ARTIFICIAL; ABRASIVE.

(CL. F.)

EMETICS are substances given for the purpose of inducing vomiting, particularly in acute poisoning if washing out of the stomach is not practicable. It is customary to divide emetics into central, those which act on the vomiting centre in the medulla, and gastric, those which act directly on the stomach itself. Gastric emetics in common use are alum, ammonium carbonate, zinc sulfate, sodium chloride (common salt), mustard and warm water. Central emetics are apomorphine, tartar emetic, ipecacuanha, senega and squill.

Of these, tartar emetic and ipecacuanha come under both heads; when taken by the mouth they act as gastric emetics before absorption into the blood, and later produce a further and more vigorous effect by stimulation of the medullary centre. Their action is accompanied by intense depression.

Emetics find limited use in the removal of foreign bodies or of plugs of mucus from the bronchi or windpipe.

EMIGRATION, the movement of population out of one country into another. *See* MIGRATION.

EMILIA-ROMAGNA, a region of north central Italy, is bounded by Venetia and Lombardy on the north, Liguria on the west, Tuscany on the south, the Marches on the southeast and the Adriatic sea on the east. Pop. (1961) 3,646,507. Area 8,542 sq.mi. The region embraces the provinces of Bologna, Ferrara, Forlì, Modena, Parma, Piacenza, Ravenna and Reggio nell'Emilia.

The northern portion is a great plain from the Via Aemilia to the Po. Its highest point is not more than 200 ft. above sea level, while along the east coast are lagoons near the mouths of the Po and those called the Valli di Comacchio to the south of them. Farther south is the plain round Ravenna (10 ft.), which continues as far as Rimini where the mountains come down to the coast.

Immediately to the southeast of the Via Aemilia the mountains begin to rise, culminating in the central chain of the Ligurian and Tuscan Apennines. The boundary follows the summits of the chain in the provinces of Parma, Reggio and Modena, passing over the Monte Bue (5,840 ft.) and the Monte Cimone (7,103 ft.), while in the provinces of Bologna and Forlì it stays along the northeastern slopes of the chain.

With the exception of the Po the main rivers of Emilia-Romagna

descend from this portion of the Apennines, the majority of them being tributaries of the Po. The Trebbia (which rises in the province of Genoa), Taro, Secchia and Panaro are the most important. Even the Reno, Ronco and Montone, which now flow directly into the Adriatic, were, in Roman times, tributaries of the Po, and the Savio and Rubicon (Uso) seem to be the only streams of any importance from these slopes of the Tuscan Apennines which ran directly into the sea in Roman times. A considerable amount of electric power is derived from these rivers, and the stations are connected with the Alpine plants so that interchange at different seasons is possible. Emilia-Romagna is one of the leading regions of Italian agriculture. A large part of it is lowland, with water available both in the form of precipitation and from irrigation systems. Wheat and maize are the chief cereals, sugar beet, tobacco and potatoes the industrial crops. Vegetables are grown in the lowlands, while the best wines come from the Apennine slopes. There are large numbers of beef and dairy cattle. Cheese, meat and smoked meats of various kinds support a large food-packing industry and there are flour and sugar milling and the preserving of vegetables and fruit. Among the industries the manufacture of cars and trucks, farm machinery, chemicals and pharmaceuticals and clothing are particularly important. The discovery of large deposits of natural gas (at Cortemaggiore north of Piacenza and near Ravenna) and of oil (at Cortemaggiore) gives the region a vital role in the energy economy of Italy. The main rail artery is the Milan-Bologna-Rimini-Bari line. Bologna also has trunk railway lines connecting it with Verona and Venice to the north and Florence and Rome to the south. The region is well served by secondary railway lines and highways. Express highways connect Bologna with Milan and with Florence.

History.—The name Emilia comes from the Via Aemilia, the Roman road from Ariminum (Rimini) to Placentia (Piacenza) which traversed the entire district from southeast to northwest, its line being closely followed by the modern railway. In popular usage the name was transferred to the district (which formed the eighth Augustan region of Italy) as early as the time of Martial, and in the 2nd and 3rd centuries A.D. Aemilia was frequently named as a district under imperial judges, generally in combination with Liguria or Tuscia. From the 3rd to the 5th centuries the district of Ravenna was, as a rule, not treated as part of Aemilia, the chief town of the latter being Placentia. In the 4th century Aemilia and Liguria were joined to form a consular province; after that Aemilia stood alone, Ravenna being sometimes temporarily added to it. The boundaries of the ancient district correspond approximately with those of the modern.

In the 6th century A.D. Ravenna (*q.v.*) became the seat of a Byzantine exarch. After the Lombards had for two centuries attempted to subdue the Pentapolis (Rimini, Ancona, Fano, Pesaro and Senigallia), Pippin took these cities from Aistulf and in 755 gave them, with the March of Ancona, to the papacy to which, under the name of Romagna, they continued to belong. The other chief cities of Emilia—Ferrara, Modena, Reggio nell'Emilia, Parma, Piacenza (*qq.v.*)—were independent. Whether belonging to the Romagna or not, each had a history of its own and, notwithstanding the feuds of Guelphs and Ghibellines (*q.v.*), they prospered considerably. Pope Nicholas III obtained control of the Romagna in 1278, but the papal dominion during the Avignon period was only maintained by the efforts of Cardinal Albornoz, a Spaniard sent to Italy by Innocent VI in 1353. Even so, the papal supremacy was little more than nominal and this state of things only ceased when Cesare Borgia, the natural son of Alexander VI, crushed most of the petty princes of Romagna, intending to found there a dynasty of his own. On the death of Alexander VI, however, it was his successors in the papacy who profited by what Cesare Borgia had begun. The majority of the towns were thenceforth subject to the church and administered by cardinal legates. In 1796–1814 Emilia was first incorporated in the Italian republic and then in the Napoleonic Italian kingdom. After 1815 Romagna returned to the papacy and its ecclesiastical government. The duchy of Parma was given to Marie Louise, wife of the deposed Napoleon, and Modena to the archduke Francis of Austria, the heir of the last Este. In Romagna and Modena the government was oppressive, arbitrary, corrupt and unprogressive, but in Parma conditions were better. In 1821 and 1831 there were unsuccessful attempts at revolt in Emilia, which were sternly and cruelly repressed. Chronic discontent continued and the people joined again in the movement of 1848–49, which was crushed by Austrian troops. In 1859 the struggle for independence was finally successful, Emilia passing to the Italian kingdom almost without resistance. (G. Kh.)

EMINENT DOMAIN is the power of government to take private property for public use without the owner's consent. Constitutional provisions in the United States, and most civil-law codes, impose a duty to pay compensation to the property owner. In countries such as England, with unwritten constitutions, the supremacy of parliament makes it theoretically possible for property to be taken without compensation. Confiscation is the term most often used in contrast to eminent domain to describe a taking by the state without compensation, while expropriation, condemnation, compulsory acquisition and compulsory purchase are used as synonyms for eminent domain (*see* Confiscation and Expropriation).

English practice in paying compensation and the struggle in U.S. courts to distinguish the power of eminent domain from other powers of government such as the police and war powers, under which a sacrifice of property interests may be imposed without compensation, indicate that payment is a matter often determined by the moral, political and economic judgment of the community. Thus, the emancipation of slaves in England was accompanied by the payment of partial compensation to the owners, and in some European countries liquor-prohibition laws required that payment be made to the owners of distilleries. In neither situation was compensation paid in the United States. Restrictions placed on particular pieces of property because of their value as historic and artistic monuments are more often compensated for in England than in the United States.

Legislators in all countries, as well as courts in the United States, must determine what constitutes "property" (*e.g.*, air space for airplanes and radio waves have been taken in all countries without compensation); when property is "taken" (*e.g.*, U.S. courts have differed on whether the damaging of land adjacent to a public use constitutes a taking); and what constitutes "just" compensation (*e.g.*, most U.S. courts hold that the expense of moving to another site need not be included in compensation paid to an owner). Because of written constitutions, U.S. courts perhaps more than others have had to determine when a proposed use is "public" (*e.g.*, it is generally considered permissible to take land from A in order to sell it to B if B's use is one which serves a public end such as private low-cost housing, a public utility or a new industry for the community).

The history of the power of eminent domain is obscure. The *Institutes* of Justinian and other Roman legal literature are silent on such a state power. The idea of eminent domain as a power peculiar to sovereignty but coupled with a duty to pay compensation comes to us from 17th-century natural-law jurists such as Hugo Grotius (*De jure belli ac pacis*, book iii, ch. 20, sec. 7) and Samuel Pufendorf (*De jure naturae et gentium*, book i, ch. 1, sec. 19). The English practice at the time of the founding of Jamestown and Plymouth was for parliament to authorize the taking of property and to either prescribe the amount to be paid or to provide what looked like a judicial proceeding to determine the amount of the compensation, but this proceeding was held without the presence of the owner. The American colonies developed judicial procedures by which the owner could be heard on the question of compensation either in the proceeding commenced by him or in a proceeding brought by the government in which the amount of damage and the right to condemnation were determined.

United States.—The 5th amendment to the United States constitution provides: "nor shall private property be taken for public use without just compensation." While this provision applies only to the federal government, almost all the state constitutions have similar provisions. Constitutional provisions in many states also prescribe that certain purposes are public (*e.g.*, public housing in the New York constitution), making the "public" nature of the

purpose a judicial question; require compensation for damaging as well as taking property (*e.g.*, the Illinois constitution); determine the setting-off of benefits against damages; restrict the taking to an easement (*q.v.*) for some purposes; specify the tribunal for the award of compensation (jury or commissioners); and govern the time for compensation, stating whether it is to be paid before or after the government takes title or possession.

Although the determination of whether a use is public or not is said to be a judicial question, the courts have moved so far from the idea that actual use by an appreciable part of the public is a requisite to public purpose, toward a conclusion that if any public purpose is served public use is unnecessary, that it is almost correct to say that the question of whether a taking of land is for a public use is a legislative and not a judicial one (see *United States ex rel. T.V.A.* v. *Welch*, 327 U.S. 546 [1946]). Thus, use of the condemnation power has been upheld where the purpose was to clear slums, beautify an area, construct low-cost housing, provide off-street parking and promote industrialization.

While U.S. courts have relaxed control over the public purpose of condemnation, the question of just compensation has remained a judicial one. Unlike England, there have been very few legislative attempts to control or define just compensation. In general the judicial definition is that just compensation is a sum equal to the market value of the property at the time of the taking. Market value includes as a factor not only existing use value but also the best use to which the property may be put. Excluded from value is any value attributable to illegal use and any enhancement resulting from causes necessitating the taking. The judiciary has from time to time, however, excluded certain factors of market value. Thus, adaptability of the land for water-power purposes has been excluded when the private owner's right of water-power development is subject to control by the federal government (see *United States* v. *Twin City Power Co.*, 350 U.S. 222 [1956]). It has also been held that the cost of removal from the condemned property is not a necessary element of just compensation.

Condemnation may be authorized by legislative act for each particular enterprise or, as is usually the case, it may be authorized generally for specified purposes without reserving to the legislature a judgment regarding the merits of a particular undertaking. Many states and the federal government have "quick-taking" statutes providing that, upon the deposit of adequate security, the government may take title and possession before the proper price is determined judicially.

All kinds of objects of property and limited interests such as easements and leaseholds in objects of property may be taken as well as the whole title. While there is ordinarily no difficulty in distinguishing the power of eminent domain from other governmental powers under which the sacrifice of a property interest may be demanded, difficulty arises in attempting to draw a distinction between condemnation and a regulation which is permissible under the police power (*q.v.*) without compensation. In general, property rights may be restricted without compensation where the uses controlled are harmful to the public interest, but compensation is required where the use is beneficial to the public.

(A. Dm.)

Great Britain.—In the unwritten English constitution there is no general provision authorizing the compulsory purchase of the subject's land. Apart from a very doubtful survival of prerogative powers to seize property in wartime, the only authority for such compulsory acquisition is an act of parliament. The right to take land in this way is one of the attributes of sovereign power and can only be exercised by a statute which expressly authorizes it. The grant by parliament of such powers was known in Tudor times, but the great increase in their use during the first half of the 20th century was new. This came about by piecemeal enactment and without reference to any constitutional principle or theory of state rights, eminent domain or similar doctrine. The growth of compulsory acquisition, and the rather cumbrous and ill-conceived procedures by which the power is exercised, aroused resentment in the 1940s and 1950s.

There is no instance of a statute's authorizing such acquisition without also requiring the payment of compensation. Parliament alone can confer power to acquire land compulsorily and has conferred it on various government departments, local government bodies and public corporations. The power may come under a general public act, which authorizes the taking of land for particular purposes but does not specify the land to be taken and merely provides an "authorization procedure" when the authority concerned has decided what land it wants. There are many of these acts, and in general they authorize compulsory acquisition for all ordinary purposes. Where, however, these powers are inadequate, a local or private act may be passed empowering a particular town council or public corporation to acquire land for a specific purpose.

The usual "authorization procedure" is by compulsory purchase order (C.P.O.), and there is a general form which such procedure must follow. There are two other and not often used acquisition procedures: the provisional order confirmation bill and acquisition under the Statutory Orders (Special Procedure) act, 1945. Both these procedures differ from C.P.O. procedure in that they ultimately require the sanction of parliament for what is a legislative act, whereas a C.P.O. depends for its ultimate authority on the exercise by a "confirming authority" of what is in law an executive or administrative act.

Procedure by C.P.O. begins with an order made by the acquiring authority—the compulsory purchaser—subject to confirmation by some "confirming authority," usually a minister of the crown. The making of the C.P.O. must be advertised locally and notice of the order served on every owner, lessee or occupier of the land comprised in the order. Should any of these persons lodge objections with the confirming authority, the latter appoints somebody to hold a public inquiry into the proposal. (In exceptional cases this can be dispensed with and the order confirmed forthwith.) In general the person appointed is a ministry inspector, but modern practice (obligatory in many instances) is to appoint a lawyer or land agent. The inspector reports to the confirming authority who, in general, is not bound to act on the report. The report and the minister's decision are published.

If the order is confirmed, an objector may challenge its validity in the high court, but the court can only quash an order on purely legal grounds; *i.e.*, because of irregularity in form, failure to comply with statutory procedure or *ultra vires* (outside its own objects).

Assuming that the order is confirmed and not successfully challenged in the high court, the authority acquires the land, and the only matter outstanding is the amount of compensation to be paid. The basis of assessment is often said to be far too low and the method of assessment too much bound up with administrative procedure. In light of modern land values, a much more realistic approach to compensation was provided by the Town and Country Planning act, 1959. There is also provision for compensation to be assessed generally by an independent body—the Lands tribunal —and not by officials. The other principal statutory authorities for the acquisition of land and the assessment of compensation are the Lands Clauses acts, 1845–95; the Town and Country Planning acts, 1947–54; the Lands Tribunal act, 1949; and the Land Compensation act, 1961.

(G. A. F.)

Other Countries.—Nearly all other countries have constitutional or statutory provisions requiring that compensation be paid for property taken (*e.g.*, Mexico, const., ch. 1, art. 27; Soviet civil code, sec. 69). The French and German systems, unlike Anglo-American law, require that it be paid in advance of taking by the government. In the countries influenced by French and German law, the question of the public purpose to be served by the taking is an administrative one and is not determined in the regular courts where the issue of compensation may be disputed. Also, there are less general statutes providing for blanket authorization of condemnation for particular purposes such as highways than in the United States, and more often there is a requirement that expropriation of each parcel for each particular enterprise be authorized by the parliament of the country.

Other countries, particularly those influenced by Soviet law, differ substantially from the Anglo-American system with regard to what constitutes private property. In Mexico, for example, minerals in place in the subsurface are not capable of private

ownership and need not be expropriated. (A. Dm.)

EMINESCU, MIHAIL (Mihail Eminovici) (1849–1889), greatest Rumanian poet of the 19th century, was born Dec. 20, 1849, in Ipotesti near Botosani, and educated in the Germano-Rumanian cultural centre of Cernauti (then in the Habsburg empire) and at the universities of Vienna (1869–72) and Berlin (1872–74), where he was influenced by German philosophy and western literature. In 1874 he was appointed school inspector and librarian at the University of Jassy, but soon resigned to take up the post of editor in chief of the conservative paper *Timpul*. He died in an asylum in Bucharest, June 15, 1889, but his literary activity had come to an end in 1883, when he suffered his first attack of mental disorder.

His great talent was first revealed in 1870 by two poems published in *Convorbiri Literare*, the organ of the Junimea society in Jassy. Other poems followed and he became recognized as the foremost modern Rumanian poet. Mystically inclined and of a melancholy disposition, he lived in the glory of the Rumanian medieval past and in folklore, on which he based one of his most outstanding poems, *Luceafărul* ("The Evening Star"; 1883). Over all his poetry hangs a cloud of sadness. Schopenhauer's philosophy and the Indian conception of life influenced him.

Eminescu transformed Rumanian poetry both in form and in content. Simplicity of language, masterly handling of rhyme and verse form, profundity of thought and plasticity of expression made him the creator of a school of poetry that dominated Rumanian thought and the expression of Rumanian writers and poets at the end of the 19th and the beginning of the 20th centuries. His poems have been translated into several languages, but chiefly into German. Among his prose writings, apart from many studies and essays, the best known are the stories *Cezara* and *Sărmănul Dionis* (1872).

See *Poems*, trans. by E. Sylvia Pankhurst and I. O. Stefanovici, with introduction by N. Torga (1930). (G. Ns.)

EMIN PASHA, MEHMED (Eduard Schnitzer) (1840–1892), German traveler in Africa, administrator and brilliant linguist, was born in Opole (then Oppeln), Silesia, on March 28, 1840. He graduated in medicine at Berlin and then for nine years was in Turkish government service. He adopted a Turkish name but apparently he did not embrace Islam.

In 1875 he went to Khartoum and joined Gen. Charles Gordon as medical officer and in 1878 became governor of Equatoria province of the Sudan. He remained there after the Egyptian withdrawal, isolated by the Mahdist advance until relieved by H. M. Stanley in April 1888. He seems to have left the province with reluctance to go with Stanley to the coast. Emin Pasha was then asked by the German government to undertake an expedition in equatorial Africa to secure on behalf of Germany the territories south of and along Lake Victoria up to Laké Albert. Soon after the expedition started an Anglo-German agreement was signed (July 1, 1890) excluding Lake Albert from the German sphere of influence and the mission was modified. Emin had difficulties with the German authorities in Tanganyika and crossed into the Congo Free State in May 1891. The greater part of his caravan was sent down to the coast; he remained with the sick and was murdered by Arab slave traders at Kanema, on Oct. 23, 1892.

He never published a book but contributed many papers to journals in Germany and was a great collector of plants and animals.

See F. Stuhlmann, *Mit Emin Pascha ins Herz von Africa* (1894); A. Symons, *Emin Pasha* (1928). (R. M. P.)

EMIR: *see* Amir.

EMLYN, THOMAS (1663–1741), the first English minister of religion publicly to adopt the name Unitarian and the last Dissenter to be imprisoned for denying the Trinity, was born at Stamford, Lincolnshire, on May 27, 1663. In 1691 he became assistant minister of a Presbyterian church in Dublin. In 1702 he was condemned by the presbytery for heresy and, when he published *A Humble Inquiry Into the Scripture Account of Jesus Christ* in self-defense, was prosecuted for blasphemy and sentenced to a year's imprisonment, to be extended till he paid a fine of £1,000. In 1705 this was reduced to £70 and he was released. He died in London, July 30, 1741. His case and writings aroused sympathy in New England and the Unitarian controversy was renewed when his *Humble Inquiry* was reprinted at Boston (1756).

See E. M. Wilbur, *A History of Unitarianism* (1952).

EMMANUEL (Immanuel), a Hebrew word meaning "God is with us." It is of the same significant type as the names given by Isaiah to his sons (Isa. vii, 3 and viii, 3; Shearjashub="A remnant shall return," Mahershalalhashbaz="Speed spoil, hurry plunder") and the Puritan names of the 17th century A.D. This name was given in Isa. vii, 14 to the boy whose birth was predicted by Isaiah. Rezin of Damascus and Pekah of Israel were trying to force Ahaz of Judah into an alliance against the Assyrian invader. The birth of the boy was a guarantee that the menace would be removed before the boy was old enough to know the difference between right and wrong. The Hebrew word *Emmanuel* occurs again in Isa. viii, 8 and 10; in both cases it should be translated as "God is with us," not as the name "Emmanuel." In all three cases the teaching is that God can and will save. In Matt. i, 23 the birth of Jesus Christ is claimed as fulfillment of the prophecy of Isa. vii, 14. (N. H. S.)

EMMANUEL PHILIBERT (1528–1580), duke of Savoy from 1553, was born on July 8, 1528, at Chambéry, the son of Charles III of Savoy and Beatrice of Portugal, and became heir apparent to his father on his brother Louis's death (1536). With the state of Savoy in disorder because of the wars between the Holy Roman emperor Charles V and Francis I of France, he was sent in 1545 to serve the emperor, his mother's brother-in-law. Steadiness and military valour won him prestige: appointed commander on the northern front on June 27, 1553, he distinguished himself by the capture of Hesdin. On succeeding his father (Aug. 1553), he appointed his faithful Andrea Provana di Leyni to take the first measures to restore his devastated lands. After Charles V had passed his Burgundian and Italian inheritance to Philip II of Spain, Emmanuel Philibert's great victory at St. Quentin (Aug. 10, 1557) confirmed his moral right to Spain's support in the reconstruction of Savoy. At the peace of Cateau-Cambrésis (April 1559) he agreed to marry Margaret, sister of Henry II of France, an excellent woman but no longer young, and to hand over five fortresses as security to the French and two to the Spaniards. At Grünendal he signed secret agreements to allay Philip II's anxiety about the succession in case he should have no male heir.

Emmanuel Philibert governed Savoy as an absolute ruler, organized the country militia, adopted Italian as the official language and re-established the state's finances by strict administrative measures.

When the desired son, Charles Emmanuel, was born (1562), Emmanuel Philibert adroitly recovered Turin, Chieri, Chivasso and Villanova d'Asti from the French; he left them in possession of Pinerolo, with Savigliano and Perosa. As he valued the friendship of the Swiss highly, he came to terms with them, thereby securing Gex, Genevois and Chablais. He ably played the role of protector of the people of Montferrat (Monferrato), restive under the Gonzaga dukes of Mantua. He purchased Oneglia from the Genoese house of Doria and Tenda from some distant cousins of his. His small fleet was sent to take part in the battle of Lepanto.

Invited by Catherine de Médici to safeguard Henry III of France on his way back from Poland through Italy (1574), Emmanuel Philibert obtained from the new king the retrocession of the fortresses still in French hands. In Sept. 1575 he also recovered Asti and Santhià from the Spaniards. To further his designs on the marquisate of Saluzzo, which the French had held since 1548, he began in 1576 to exploit personal rivalries, the fear of Huguenot penetration and the ambitions of the governor of Milan. He was on the point of success when he died, at Turin, on Aug. 30, 1580.

Bibliography.—As well as the articles by various authors in the University of Turin's *Miscellanea storica* and in the special committee's volume celebrating the 4th centenary of Emmanuel Philibert's birth (1928), *see* A. Segre and P. Egidi, *Emanuele Filiberto*, 2 vol. (1928); P. Silva, *Emanuele Filiberto* (1928); C. Patrucco (ed.), *Lo stato sabaudo ai tempi di Emanuele Filiberto* (1928); R. Quazza, *Emanuele Filiberto e Guglielmo Gonzaga* (1929). (R. Q.)

EMMAUS, the name, in modern times, of the warm springs located 1 mi. S. of Tiberias, on the west side of the Sea of Galilee,

in northeastern Israel. It is also the name of at least two towns in ancient Palestine:

EMMAUS-NICOPOLIS (modern 'Imwas), on the eastern edge of the coastal plain, about 160 stadia (20 mi.) from Jerusalem on the way to Ramleh. There Judas Maccabaeus defeated Georgias (166 B.C.). The name Nicopolis was given to this Emmaus in the 3rd century A.D.

EMMAUS, 30 stadia (4 mi.) from Jerusalem, where Vespasian (according to Josephus) settled 800 army veterans; modern Qalôniyeh, 4 mi. W.N.W. from Jerusalem, would fit this description. Jesus Christ made his resurrection appearance to two of his disciples at an Emmaus 60 stadia (7½ mi.) from Jerusalem (Luke xxiv, 13–31). Some scholars would locate this Emmaus at Qalôniyeh. Others locate it at Al Qubaybah, a little over 60 stadia N.W. of Jerusalem. The most ancient tradition, however, favours 'Imwas, which fits best only if in Luke xxiv, 13 the early but less strongly attested reading of 160 stadia, rather than 60, is original. (F. V. F.)

EMMEN, a town of Drenthe province, Neth., is situated 54 km. (34 mi.) S.S.E. of Groningen by road and not far from the German frontier. Pop. (1957 est.) 63,170. It has a zoological garden, a museum and several cromlechs (prehistoric graves) dating from 1800 B.C. One of the new towns of the Netherlands, its population grew rapidly after 1947. It is on secondary railways from Zwolle and Enschede. The centre of the decaying industry of peat digging, Emmen had considerable unemployment from 1920. However, several factories have been opened after World War II, including one for nylon spinning, which have completely changed the character of the district and almost eliminated unemployment. In the neighbouring municipality of Schoonebeek rich oil fields were discovered during the war and have since been exploited. (K. H. GA.)

EMMET, ROBERT (1778–1803), Irish revolutionary hero who inspired the abortive rising of 1803, was the youngest son of Robert Emmet, physician to the lord lieutenant. Born in Dublin in 1778, he had a distinguished career as a student at Trinity college, Dublin, but before graduating he removed his name from the college books (April 1798) in protest against the inquisition into students' political views conducted by Lord Clare as chancellor of the university. His elder brother, Thomas Emmet (*q.v.*), was already prominent in the United Irishmen's activities, in which Robert also now became involved. A warrant for his arrest was issued in April 1799 but was not enforced, and he retired to France. There in 1800–02 he consorted with the leaders of the United Irishmen who had been exiled during the rebellion of 1798 and were planning a new insurrection with French support. An interview with Napoleon in Paris in Oct. 1802 convinced Emmet that the peace of Amiens would soon be broken and that the French would attempt an invasion of England in Aug. 1803. Divided counsels among the Irish leaders in France prevented any definite plan, but Emmet left Paris for Ireland in Oct. 1802 determined upon insurrection. In Dublin he was informed that 17 counties were ready to take up arms if Dublin gave a successful lead. Emmet began his preparations, while staying concealed in his father's house near Milltown. Pikes and other crude arms were collected and stored in Dublin early in 1803 but there were few firearms.

When war with France was renewed in May 1803, Thomas Emmet was in Paris conferring with Talleyrand and Napoleon. Meanwhile Robert's hand was forced in Dublin when the government discovered a concealed store of arms and there was an explosion at one of his secret depots. He decided quickly that a rising should take place on July 23. It was to include attacks on Dublin castle, on the Pigeon house and on the artillery barracks at Island bridge, while insurgents were to converge upon Dublin from other counties. The insurrection ended in utter confusion, with complete lack of discipline, distrust among the leaders and the treachery of unsuspected spies. The Wicklow contingent never arrived; the Kildare men retired on hearing that the rising had been postponed; while the contingent at Broadstairs waited vainly for the signal. Emmet, on receiving news that government troops were approaching, put on his green and white uniform and marched against Dublin castle with about 80 men, collecting a similar number along the route. During their march the excited crowd encountered the lord chief justice, Lord Kilwarden, on his way to an urgent meeting of the privy council. He and his nephew Richard Wolfe were dragged from their carriage and murdered. Emmet, recognizing that the rising was a fiasco, escaped and for several days subsequently lived in the Wicklow mountains. He then moved in to Harold's Cross to be within reach of John Philpot Curran's daughter Sarah, to whom he was secretly engaged. Hoping to escape with her to America, he remained in the neighbourhood till Aug. 25, 1803, when he was captured. He was tried for treason, found guilty, and hanged on Sept. 20, 1803.

High minded, with gallant courage, yet without discretion or common prudence, Emmet has been for generations one of the most romantic heroes of Irish lost causes. The poet Thomas Moore, who was his contemporary at Trinity college, recalled his passionately patriotic response when he heard Moore's melody "Let Erin Remember." Moore's famous songs "She is far from the land where her young hero sleeps" and "Oh breathe not the name" were directly inspired by Emmet's love affair with Sarah Curran, but she in fact soon found consolation with another suitor.

See R. R. Madden, *The United Irishmen, Their Lives and Times,* 4 vol., 2nd ed. (1858–60); Roy Jenkins, *Sarah Curran* (1958).
(D. G.)

EMMET, THOMAS ADDIS (1764–1827), Irish lawyer and a member of the supreme council of the United Irishmen, was the elder brother of Robert Emmet (*q.v.*). Born at Cork on April 24, 1764, Thomas was educated at Trinity college, Dublin, and later studied medicine at Edinburgh university and other medical schools in Europe. Returning to Ireland in 1788, he decided to change from medicine to the law, and after legal studies in London he was called to the Irish bar in 1790. His first brief was to defend Napper Tandy, and he soon became the foremost advocate in defending political prisoners. In 1795 he boldly took the United Irishmen's oath in open court, and was that year elected secretary of their supreme council. Before the rebellion of 1798 he deplored rash action and tried to insist upon waiting for armed aid from France. He was imprisoned with other leaders at Ft. George in Scotland until 1802, when he went to Brussels and became involved in urging Napoleon to promote an invasion of Ireland. He was in Paris when he heard of Robert Emmet's abortive rising and his execution in Dublin. He left France for the United States in 1804, and soon took a leading position at the New York bar. He died there suddenly while pleading in court on Nov. 14, 1827. (D. G.)

EMMETT, DANIEL DECATUR (1815–1904), U.S. composer of "Dixie," was born at Mount Vernon, O., Oct. 29, 1815. He joined Negro minstrel companies as a singer and composer and wrote both words and music for "Dixie." It was first performed on April 4, 1859, and immediately attained wide popularity. During the American Civil War it was adopted by the Confederate troops as a fighting song and became the unofficial "national anthem" of the south. Emmett died at Mount Vernon on June 28, 1904.

See H. Nathan, "Dixie," in *Musical Quarterly* (Jan. 1949).

EMOTION, as a human experience, is a commonly recognized psychological state, and many ordinary words describe different kinds of emotions—anger, fear, happiness, sorrow, etc. Behavioural scientists, in attempts to define psychological and biological attributes, have focused their attention on some of the more distinctive characteristics of emotional states: (1) the subjective feelings of pleasantness or unpleasantness; (2) the stereotyped external reactions, such as facial expression and bodily posture; (3) the internal reactions, such as changes in blood pressure, heart rate, sweating or concentration of certain hormones; and (4) the over-all effect of emotional states on a person's general behaviour. Though there has been much argument as to whether emotion is mostly "organizing" or mostly "disorganizing" to general behaviour, the argument is pointless. The question of organization or disorganization depends on the type of behaviour and the details of the situation.

Experimental Studies.—Experimental studies of normal emo-

tion have been carried out with animals and with human beings at all ages. Though it is possible to experiment with animals under situations and conditions that human beings cannot be subjected to—for example, intense electrical stimulation after removal of part of the brain—studies with animals provide no certainty about the exact emotional state experienced. A human being may be asked whether he is angry or afraid; such a report cannot be secured from an electrically stimulated rat. Hence in animal experimentation secondary evidence is resorted to; it is common, for example, to express the animal's degree of fear in terms of increase in frequency of the heart beat or defecation, or of how efficiently it learns to escape or avoid a noxious situation. In the primates, since to some extent they differentiate in their facial and vocal expressions, a better guess of the emotional state is permitted.

Psychological studies of human emotional states may be either direct or indirect. In the direct approach, a subject may be placed in an emotionally provocative situation and then asked to describe the experience; this can be correlated with physiological or photographic evidence obtained at the same time. In an indirect approach, a subject may be confronted with an unstructured situation (*e.g.*, an ink blot or a cloud picture) and asked for a report of his mental processes when looking at the material. This provides an opportunity for reporting affective experiences that are projected onto the stimulating material; such projections are part of the unconscious, or only partly conscious, emotional mental states.

Rating scales or questionnaires have been the commonest method of investigating emotional experience. In these procedures a number of people are asked to rate the emotionality or emotional expressiveness of one person in the group. Or one person may be asked to give a description of his own affective experiences in response to a series of questions dealing with life situations. Such methods give an indication of the degree and variety of emotional experience. Neither method is of particular value in prediction of behaviour or experience in any specific situation, though both provide clues predictive of the general style of reaction.

Expressive Emotional Reactions.—Since bodily reactions during emotion are fairly distinct and may be observed and recorded, they have been the objects of much experimental investigation. One of the most characteristic is the change of facial expression in emotion, about which there has long been speculation. In 1872 Charles Darwin formulated a systematic approach in his *Expression of the Emotions in Man and Animals,* which became basic for many experimental studies. Darwin postulated three principles for explaining most, if not all, facial reactions in emotion: (1) The principle of serviceable associated habits: "Certain complex actions are of direct or indirect service under certain states of mind, in order to relieve or gratify certain sensations, desires, etc.; and whenever the same state of mind is induced, however feebly, there is a tendency through the force of habit and association for the same movements to be performed, though they may not be of the least use." (2) The principle of antithesis: "Certain states of the mind lead to certain habitual actions, which are of service, as under our first principle. Now when a directly opposite state of mind is induced, there is a strong and involuntary tendency to the performance of movements of a directly opposite nature, though these are of no use; and such movements are in some cases highly expressive." (3) The principle of direct action of the nervous system: "When the sensorium is strongly excited, nerve-force is generated in excess, and is transmitted in certain definite directions, depending on the connection of the nerve cells, and partly on habit: or the supply of nerve-force may, as it appears, be interrupted. Effects are thus produced which we recognize as expressive." Darwin's principles were partly adopted and amended by Wilhelm Wundt and William James. All of these later formulations gave rise to experimentation.

Smiling, Laughing and Crying.—Clear and well-defined patterns of emotional expression are those of smiling, laughing and crying, types of behaviour that are uniformly accepted as indicative of emotion. Most children smile during the first weeks of life, suggesting, along with other considerations, that the ability to make the basic facial movements of expression is innate. Smiling is brought about in the young infant by a variety of experiences, and, according to one theory, by the end of the first year is "selected" as a socially appropriate response because it is greeted with pleasure by mother or nurse. This original usage is then generalized and blended into all the patterns of response that develop into the experience called amusement.

Laughing usually does not occur until after the 20th week of life, and during the first year it is a reflex form of behaviour. Individual differences among children are seen in the frequency rather than in the actual form of smiling and laughing. Human laughter has been classified as that derived from joy or from comedy, as a social phenomenon, as a release from tension and as a symptom of pathology. The joyful laugh, a bubbling over of good humour, commonly occurs in children and adults in a state of well-being; the comic laugh is directed at a joke or ridiculous situation. Laughter as a social phenomenon involves a wide range of social reactions; we laugh more easily in a group, for example, than when alone, and laughter sometimes is merely a social gesture. The laughter associated with relief of tension has sometimes been explained biologically; since the facial musculature is ordinarily not necessary for the active preservation of life, the excess energy set up by emotional stimulation spills over into activity of the facial and respiratory musculature in a way that does not interfere with other activities of the body. Laughter associated with pathological conditions can be caused by organic pathology of the brain or, it has been suggested, by the expression of unconscious tendencies in psychopathological states.

Crying has been observed experimentally in infants in a variety of situations. Crying induced by strangers increases in frequency up to about ten months of age. Crying caused by fear or strange situations can be distinguished from crying that occurs for other reasons. In the adult, crying, like laughing and smiling, becomes so bound up with individual social reactions that it is often impossible to be certain whether the response is truly emotional, only partly so, or devoid of emotion.

Anger and Fear.—Reactions occurring during offense, attack and positive movement are usually called anger; those occurring during defense or retreat are called fear. In the young there is little difference in facial reaction between the expressions of anger and fear. Both fear and anger are thought to be "preparedness states" of the young organism. Such preparedness, originally undifferentiated, provides the necessary energy to meet unknown situations. As the infant grows older and is able to differentiate more clearly among situations, anger and fear also become differentiated.

The condition that arouses anger in young children is a stoppage or an interference with activity, particularly with activity involving the so-called instinctive urges or drives. Anger responses in the child are outbursts of impulsive behaviour—kicking, stamping, slashing, holding the breath. With increasing age the reactions focus upon a given end. Along with the decrease in proportion of outbursts of undirected energy comes an increase in the frequency of retaliative behaviour.

The original adequate stimulation for fear is the sudden confronting of the individual with a situation requiring adjustive reactions he is unprepared to make. Fear is produced in the young infant by loud sounds or loss of bodily support. It may also occur in the presence of unexpected alterations in the appearance of the environment without any direct physical threat. The overt reaction to fear changes during the growth of a child. The early undifferentiated panic—screaming, crying or rigidity—changes into running away, withdrawing, tension and vocalization.

During the first half-second after a person is stimulated by a sudden, loud sound, flash of light or intense odour, there occurs a rapid response, known as the startle pattern, consisting of a sudden movement of the head, blinking of the eyes, a characteristic facial expression, raising and drawing forward of the shoulders, turning inward of the upper arms, bending of the elbows, turning downward of the arms, flexion of the fingers, forward movement of the trunk, contraction of the abdomen and bending of the knees. Not all of these movements occur in every individual every time he is so stimulated; the number of elements and the completeness of the pattern vary with the novelty and intensity of the stimulus.

This response is probably the initiating element of many, more general, emotional reactions, both motor expressions and conscious experiences.

Love.—Love is the term used to describe such diverse manifestations of behaviour and experience as the parent-child relationship, protecting reactions and appreciation of protection, friendship, attachment between the sexes, attachment to places, to things, to occurrences or to food. Most experimental studies on love reactions have been based upon mating and maternal behaviour of animals. Such studies have shown that this behaviour has a substantial innate component and becomes evident later in life through biological maturation rather than through experience and learning. On the other hand, the adult effectiveness of such instinctive behaviour can be considerably modified by very early experience. For example, infant female monkeys raised without their natural mothers show very deficient maternal behaviour when they become mothers themselves.

Visceral Changes.—A prominent criterion of emotion is the appearance of overt bodily or visceral changes, phenomena so marked that many attempts have been made to measure or to delimit emotion in terms of them. This type of measurement has been widely used in the detection of falsehood, the so-called lie detector being an instrument to record emotionally caused changes in blood pressure, respiration and electrical properties in the skin.

Theories of Emotion.—Until the 19th century various doctrines of pleasure-pain held a central place in philosophical psychology. With the advent of experimental psychology came a demand for theories that could be subjected to testing. The James-Lange theory (independently formulated by William James in 1884 and by Carl Lange in 1885) provided the first adequate basis for many experimental investigations. James made a distinction between emotion and instinct:

> Instinctive reactions and emotional expressions thus shade imperceptibly into each other. Every object that excites an instinct excites an emotion as well. . . . Common sense says, we lose our fortune, are sorry and weep; we meet a bear, are frightened and run; we are insulted by a rival, are angry and strike. . . . My theory, on the contrary, is that the bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur is the emotion . . . that we feel sorry because we cry, angry because we strike, afraid because we tremble, and not that we cry, strike, or tremble, because we are sorry, angry or fearful, as the case may be (William James, *Psychology,* Henry Holt and Co., Inc. [1890]).

In the end, of course, the subjective feelings of emotion must be dependent on activity in the brain. According to the James-Lange theory, this brain activity is dependent on the prior occurrence of some pattern of muscular or visceral reaction.

In 1895 John Dewey proposed what came to be known as the conflict theory of emotion. He suggested that emotions arise when behaviour is obstructed by competing response tendencies—that is, when a person is torn, so to speak, between alternative courses of action. It has come to be recognized that such conflict is one important source of emotional reactions, but it does not seem that this theory can account for the wide range of circumstances that are capable of inducing emotional behaviour.

When research workers began to study the brain directly in order to see how it mediated emotional behaviour, an alternative theory to that of James and Lange was proposed and a point of view was developed that came to be known as the Cannon-Bard theory (1928), after W. B. Cannon and Philip Bard, whose experiments were mainly responsible for it. According to this theory, the bodily responses usually associated with an emotional state are reflex by-products of the state which, though of adaptive, biological value, are not themselves causally related to the over-all nature of emotional behaviour. The behaviour, rather, is solely dependent on activity within the brain. (The hypothalamus received special attention from these early workers.)

Cannon's work on the visceral aspects of emotional states led him to propose that such bodily responses prepare the individual to withstand the possible damage resulting from a stressful situation. For example, the dilation of the pupils, the increased blood flow to the muscles and the shortened clotting time of the blood that accompany extreme emotional reactions were seen as changes that gave the individual a better chance of survival during an emergency. This part of Cannon's proposals has come to be known as the emergency theory of emotions.

Biological Basis of Emotion.—The basic theoretical argument —whether the unique characteristics of emotion have their origin in the brain itself or elsewhere in the body—has its counterpart in later physiological and anatomical research on the biological basis of emotion. With improved techniques of physiological measurement, it has been shown that there are different patterns of internal bodily responses for the basic emotions of fear and anger; for example, both heart function and blood pressure change in slightly different ways in these two emotional states. While it is known that such physiological changes can influence brain function and are thus capable of supplying the subjective component of the emotions, it is not known whether they in fact do so. Though, in a more general sense, there is ample evidence that many kinds of bodily change are capable of influencing brain function (and thus behaviour), this matter is not settled with regard to the origin of emotional feelings. That certain emotional moods, however, can be modified by drugs is evidenced in the effects of the tranquilizers and the excitants (*see* TRANQUILIZING DRUGS; NEUROPHARMACOLOGY AND PSYCHOPHARMACOLOGY).

While the causal relation of bodily changes to emotional feelings remains unsettled, a growing body of observations indicates that persistent or severe emotional reactions can have widespread effects on physiological functions. Such findings have been concerned almost exclusively with the so-called negative emotions, such as fear and grief, and the hormonal responses of the pituitary and adrenal glands have received the greatest amount of experimental attention in this regard. However, the influence of stressful emotional reactions on the pituitary, with the subsequent increased secretion of adrenal hormones, can affect many bodily systems. Through such a hormonal route, stressful psychological events can seemingly influence such bodily functions as resistance to disease, type of circulating blood cells and, possibly, even growth rate. Aside from hormonal effects, there is also clinical speculation as to whether specific types of chronic emotional reactions can lead to specific types of bodily disease (for example, asthma and stomach ulcer). (*See* PSYCHIATRY: *Modern Schools: Psychosomatic Medicine.*)

As pointed out above, the brain is responsible for controlling or mediating the various aspects of emotional behaviour—the subjective feelings, the body movements and the more primitive internal bodily changes. As early as 1928 Philip Bard demonstrated that part of the hypothalamus (*q.v.*) is crucially involved when an enraged animal attacks. Subsequent experiments suggest that this brain centre initiates the basic movement reflexes that go along with certain emotional states. Brain studies on the more complex characteristics of emotional behaviour began in earnest after J. W. Papez (1937) suggested that the so-called limbic system of the brain might be important in emotional arousal. His suggestion was based partly on the fact that this system is intimately connected with the hypothalamus. The limbic system is a collection of phylogenetically old, interconnected brain centres located deep inside the cerebral hemispheres (*see* BRAIN: *Anatomy of the Adult Human Brain: Cerebral Hemispheres*). It had been assumed that these more primitive portions of the brain were concerned almost exclusively with the sense of smell, and this misunderstanding prompted early anatomists to call this part of the brain by its alternative name, the rhinencephalon. Following Papez' important suggestion, many workers began to study emotional behaviour in animals after various parts of the limbic system had been damaged or were under electrical stimulation. While its role is still not completely understood, the limbic system certainly is involved in emotional behaviour.

The rhinencephalic brain centres seem to be concerned with more complex emotional responses, as contrasted with the more stereotyped reactions (such as hissing in the cat) which appear to be under hypothalamic control. A frightened animal can react adaptively to a threatening situation by crouching motionless or by running away; apparently two different centres in the limbic system initiate these alternative patterns of response. Some ex-

perimental results also suggest that fear and anger are mediated by separate parts of this system. Since in the laboratory it is easier to make an animal frightened or angry than it is to make it, for example, affectionate, animal studies have been concerned almost exclusively with the unpleasant emotions; thus it is not possible to say whether the limbic system plays any role in pleasant emotional experience.

All emotional states have one characteristic in common: the person is aroused. Arousal is most easily understood by comparing a waking with a sleeping person, the one being aroused, the other not. In the same sense, a person experiencing emotion is more aroused than one who is not emotional. Considerable information is available concerning that part of the central nervous system, called the reticular system, whose function is to keep the brain (and thus the person) aroused or activated. While there are two such systems with somewhat different characteristics (*i.e.,* the thalamic and the brain-stem reticular systems), they may be considered together. The chief characteristic of this combined system is its ability to be stimulated by all types of sensations, those arising outside the body (*e.g.,* light and sound) as well as those arising within (*e.g.,* pain). The reticular system also can be influenced by bodily changes of which we are not ordinarily aware—hormonal changes, changes in concentration of carbon dioxide in the blood, and the sensory effect of muscle movements. Once activated, it is capable, through its widespread connections with other parts of the brain, of arousing and maintaining the efficiency of the rest of the nervous system, both the higher and the lower centres. There is reason to think that when it is overstimulated the effect on the rest of the brain may be decreased efficiency, the electrical circuits of the brain becoming, so to speak, jammed. Emotion too may either augment or reduce efficiency of behaviour. Emotions of low or moderate intensity can alert the subject and improve over-all performance, but intense emotional reactions can impede it. It is thought, but not proved, that such effects of emotional states on behaviour are mediated through the reticular system. Certainly external stimuli may be coloured by emotion, and emotional reactions can have broad effects on internal bodily processes; thus, at the least, emotion is capable of influencing the factors that are known to influence the reticular system.

Emotional Development.—Most adult observers, confronted with photographs of faces bearing expressions of emotion, can identify the emotion fairly accurately. Investigation has shown that: (1) this ability increases with mental age; (2) some expressions are easily and accurately, others only with difficulty, interpreted; (3) individual differences in interpretive skill exist among observers; (4) facial expression may be considered as a method of communication or as a language mechanism, since the whole process of conveying and receiving information seems to be learned in much the same manner as spoken language.

Study of the development of emotion and emotional expressiveness in human beings has clarified several points. The first definite sign of emotional behaviour in the human infant is a general and undifferentiated excitement caused by changes (usually increases) in stimulation. This early undifferentiated excitement separates first into distress and delight. Distress in turn differentiates into expressions and feelings recognized as fear, disgust, anger and jealousy. Delight develops into elation, affection and joy. This differentiation takes place during the first two years of life. The extent to which the process is innately determined or dependent upon associative learning remains uncertain. The classical experiments of J. B. Watson seemed to demonstrate that the fear response could be attached to neutral objects by simultaneous presentation (conditioned emotional response; *see* CONDITIONING). For example, an 11-month-old infant was "natively" afraid of the sound of a loud gong; presenting the infant with a white rabbit (neutral object) when the gong was sounded soon led to crying and attempts to escape when the rabbit was presented, even though the gong was not sounded. Later experimentation failed to confirm this form of conditioned fear response in all infants. Some, but by no means all infants can be made fearful in this way, and the fearfulness so established does not remain as part of the emotional equipment throughout life. Where such conditioned emotional responses have been established in children, they have been removed in a variety of ways—through disuse, through familiarity, and by repression, distraction and social imitation (as a result of association with children who lack the fear).

Studies of development of emotional expressiveness in blind children give support to the idea that ability to adopt appropriate basic facial expressions is innate. Motion-picture records of the facial expressions of blind and sighted children in naturally occurring emotional situations showed that the facial responses of the blind exhibited a greater variety of patterns than did those of sighted children but that at a given age the blind children's patterns included those of the sighted children. Blind and sighted children were asked to exhibit the facial expressions they thought appropriate for various kinds of emotion. The blind children responded with some of the same facial patterns shown by the sighted children, though the latter differentiated their expressions more sharply and adequately than the former did. Seemingly, thus, the occurrence of the facial expressions of emotion, the interpretation of facial expression and the ability to produce relevant facial expressions depend on both innate factors and social experience.

Aside from such specific emotional expressions, it has been shown that early experience influences the individual's later emotional tone in a more general way. Studies with infant and young animals, as well as clinical observations of human babies and children in various types of institutions, have demonstrated that certain specific characteristics of the early general environment help to determine whether the individual will develop normal emotional patterns. In particular, if a young animal or child is not exposed to sufficient variation in environmental stimulation (sights, sounds, personal contacts, etc.) he is likely to develop a more or less lethargic temperament; as an adult he will tend to be less reactive to events around him, less outgoing and, so to speak, show less than normal excitability. Even monkeys, raised without their natural mothers, as adults are abnormal in their emotional relationships with other monkeys and with their environments in general. There is a growing suspicion that in both animals and human beings such early environmental deprivation can affect adversely even the biological resistance to physical and psychological stress and such basic functions as rate of body growth.

It should be emphasized, however, that all such learned or acquired influences on adult emotional status are only secondarily superimposed on a substantial number of innately determined characteristics. For example, the easily provoked ferocity of the lion and the contrasting timidity of the easily frightened deer are clearly based on genetic differences. Further, within a single species (for example, of rats) it has been possible to breed strains that differ markedly from the normal in emotional tone, or temperament. Different strains of domesticated dogs provide clear examples of intraspecies differences. While research of this type on human beings is rare, there is every reason to think that we are as open to genetic influence in this regard as are the lower animals.

Emotion in Pathological States.—Study of the disorders of emotion in patients suffering from mental abnormality sheds light on the whole topic of emotion. Although certain emotional experiences (*e.g.,* unreality feelings) seem to occur only in disease states, the majority of the affective responses in mental disorders are merely exaggerations of normal experience. A tendency in popular thought to identify any or all varieties of strong emotion as abnormal is the result of a literary use of the word emotion that implies unusualness. Normal emotion may be distinguished from abnormal in that it is usual or predictable, relevant or appropriate, correct in personal reference and of relatively short duration. (*See* also NEUROSES: *Types of Psychoneurotic Disorders;* PSYCHOSES: *Psychogenic Psychoses: Affective Reactions;* PSYCHOLOGY, ABNORMAL: *Psychotic Reactions: Affective Reactions.*)

Psychotic and neurotic depression offer examples of an exaggerated emotional state persisting over a long period of time (*see* DEPRESSION [IN PSYCHIATRY]). Its antithesis is mania, a state marked by elation, distractibility, overactivity and volatile expressiveness. The elation is experienced as a lifting or removal of all

feelings of inhibition; ideas, desires and wishes cease to be subject to the restraints of everyday life. The normal compulsion for logical thought is absent, the elated person feeling that his inspiration is of such high order that the usual plodding restraints of logic and reason are no longer necessary. The distractibility or distraught condition is a gross exaggeration of normal excitement. The emotional expressiveness is changeable and overactive so that the affected person screams with pleasure or sobs with joy over some trite incident.

The anxiety state includes vague fears, worries, guilt feelings, security seeking, apprehension and need for reassurance. Every human being has had this experience for greater or shorter periods of time and in varying degrees of intensity. In the mental patient the anxiety state is usually accompanied by such physical symptoms as sweating, pallor, flushing, sighing and irregularities of the heartbeat, and often it is unrelated to any adequate stimulus. The patient may complain of anxiety without knowing of anything about which he should be anxious, or he may attach his apprehension to some obviously inadequate or inappropriate situation. Phobia, or anxiety hysteria, differs from anxiety in degree and in that it is directed toward some object, person or idea. It differs from ordinary fear by being related to events and experiences that normally would not excite such a reaction, and it cannot be relieved by logic or reason.

Emotional dullness amounting to a complete loss of affective experience is termed apathy. Apathetic persons have no real joy or grief; nothing matters. A patient in the exaggerated apathetic state has no regard for his surroundings, his personal well-being or his comfort. He becomes untidy, dirty and insensitive to bodily discomfort and exhibits no feeling of disgust, shame or pleasure; he is without sympathy and equally indifferent to the ill or good fortunes of others. The outstanding characteristic of this emotional dullness is its apparent extreme uniformity—there is neither increase nor decrease in the experience of either pleasantness or unpleasantness, but rather a continued absence of such feelings. The condition seems to be approached by the normal person only in a state of extreme fatigue.

Exaggerated rage or anger (furor) occurs in a variety of psychopathological states, in senility and in patients suffering from organic damage to the brain. Such a patient may be incapable of making a decision, adjusting to a situation or changing a habitual mode of response. The continued pressure of everyday existence, acting as a stimulus to action, leads to frustration, as a result of which the patient gives way to bursts of undifferentiated excitement accompanied by destructive anger.

Conclusion.—Experimental study of the emotions since the beginning of the 20th century has yielded many isolated facts, but it has not been possible to develop a single, self-consistent theory covering all aspects of emotional experience and behaviour. With improvements in psychological research techniques, however, especially those that make possible direct experimentation on the brain, psychologists and physiologists are increasingly optimistic about the chances of solving the many remaining problems.

See FEELING, PSYCHOLOGY OF; SENTIMENT; *see* also references under "Emotion" in the Index volume.

BIBLIOGRAPHY.—P. T. Young, *Emotion in Man and Animal* (1943); D. O. Hebb, *Organization of Behavior*, pp. 171–275 (1949); C. T. Morgan and E. Stellar, *Physiological Psychology*, pp. 340–356 (1950); S. S. Stevens (ed.), *Handbook of Experimental Psychology*, pp. 474–516 (1951); J. Field, H. W. Magoun and V. E. Hall (eds.), *Handbook of Physiology*, sec. i: *Neurophysiology*, vol. iii, pp. 1529–53 (1960).

(C. LA.; R. A. MCC.)

EMPATHY, a term used as the equivalent of the German *Einfühlung*, which is very difficult to translate. It is modeled on the word sympathy. The term is used with special (but not exclusive) reference to aesthetic experience. The most obvious example perhaps is that of the actor or singer when he *feels* the part he is performing or (better perhaps) reading or studying. Similarly with other works of art: one may, by a kind of introjection, feel oneself into what one observes or contemplates. Whereas in empathy proper the observer assimilates himself to his aesthetic object, in animism he assimilates an inanimate object to himself. This latter process is much easier than the former, and no doubt it happens often enough in aesthetic experience that the subject assimilates the aesthetic object to himself—though he may mean to assimilate himself for the time being to his object —to be sympathetic with it. *See* AESTHETICS.

EMPEDOCLES (*c.* 490–430 B.C.), Greek philosopher, statesman, poet and religious teacher, founder of the Italian school of medicine, was a citizen of Akragas (Agrigentum) in Sicily. It seems that he carried on the democratic tradition of his distinguished family by helping to overthrow an oligarchic government which succeeded the tyranny in Akragas, but he refused the invitation of the citizens to become their king and later left the city. He died in the Peloponnese in 430.

Of his poem on nature (*Physis*) there are left about 400 lines, in unequal fragments, out of the original 5,000; of the *Purifications* (*Katharmoi*) less than 100 verses remain; of the other works, improbably assigned to him, nothing is known. His grand but obscure hexameters, after the example of Parmenides, delighted Lucretius. Aristotle, it is said, called him the inventor of rhetoric, and Galen regarded him as the founder of the Italian medical school. To his contemporaries he seemed more than a mere man, and even in modern times he has been celebrated by followers of Giuseppe Mazzini as the democrat of antiquity, par excellence.

Empedocles' physical theory was an attempt to provide an intelligible account of the physical world which would not be open to the criticisms brought by Parmenides against the monist theories (*see* PLURALISM AND MONISM: *Theories of Substance*). Empedocles had an alternative ready to hand in the medical theory, known to us from the Hippocratic writings, of innumerable substances composing the body and its foods. He chose a middle course between this theory and monism and assumed four primary substances or "roots," fire, air, water and earth, out of which all the structures in the world are made by their combination in different proportions (*e.g.*, bone consisted of two parts of earth, two of water and four of fire; in the importance attached to proportion we can trace Pythagorean influence). Nothing really comes into being or is destroyed, the apparent changes being due to alterations in the way in which the "roots" are mixed. The "roots" are brought together and separated by two agents, Love and Strife, which Empedocles describes in language which suggests that he thought of them as physical. Strife makes each of the elements withdraw itself from the others and "brings like to like"; Love makes them mingle together. When Strife is supreme, the four elements are arranged in concentric layers; when Love is supreme they are completely intermixed. The world as we know it shows the influence of both agents (*e.g.*, there is water and fire in the earth, as is shown by springs and volcanoes) and is therefore at a stage intermediate between the complete domination of either. The original condition was the "sphere," when Love was dominant and all the "roots" were mixed together; in the formation of the cosmos (*i.e.*, as Strife enters), first the air is separated, then the fire and then the earth, the compression of which forces the water out. The two motivating agents are active also in moral and physiological contexts; indeed, it seems to have been their physiological activity which suggested to Empedocles the extention of their functions to the whole cosmos.

It is in accordance with Empedocles' belief that the present world is a stage in which Love is being displaced by Strife that he asserts trees to have been the first living things to grow out of the earth: in them the sexes are still conjoined, and so, too, "whole-natured" forms of animals were produced without distinction of sex or species. Later, when Strife has gained ground, species and sexes are separate, as now. At the opposite arc of the cycle, when Love was gaining, separate limbs arose: hands without necks, arms without shoulders. After that incongruous combinations occurred: oxen with human heads and so forth. The fittest survived, the rest perished.

Empedocles was much interested in embryology and in the growth of plants, and there is a text in which he describes in detail the process of respiration. Growth and nutrition are on the principle of "like to like." The same principle applies to sensation and to thought: we think chiefly with our blood, especially with the blood round the heart, because in blood all the elements are most

completely mixed. Perception is due to the encounter of an element in us with the same element from outside, and this occurs when the passages of the sense organs are of just the right size to take the "effluences" which all things are always giving off. Thus our powers of perception (and of thought) depend upon the constitution of our bodies.

It is not easy to see how Empedocles' physical theories allow for the Orphic doctrine of the transmigration of souls which he appears to expound (*see* ORPHEUS). According to the decree of Necessity, Empedocles says, *daimones* who have sinned must wander 30,000 seasons from their home of bliss, being born into all manner of mortal bodies and tossed from one of the four elements to another. (Empedocles is such an exile from bliss, because he put his trust in senseless Strife.)

Escape from this condition is to be attained by purifications and by abstention from animal flesh, for animals are our kinsmen; indeed their souls may once have been in human bodies. At the top of the human scale are the prophets, the minstrels and the physicians; their next step upward is to the divine, sharing the hearth and the table of the gods.

BIBLIOGRAPHY.—For text *see* H. Diels and W. Kranz, *Fragmente der Vorsokratiker*, vol. i, 8th ed. (1956); also verse trans. by W. Leonard, *The Fragments of Empedocles* (1908). *See* further Clara E. Millerd, *On the Interpretation of Empedocles* (1908); J. Burnet, *Early Greek Philosophy* (1930); H. S. Long, "The Unity of Empedocles' Thought," *American Journal of Philology*, pp. 102–154 (1944); J. Schumacher, "Der Physis-Begriff bei Empedocles," *Sudhoffs Arch. Gesch. Med.*, vol. xxxiv, pp. 179–196 (1941); G. S. Kirk and J. E. Raven, *The Presocratic Philosophers* (1957). (A. L. PE.)

EMPEROR, the name, deriving from Lat. *imperator*, used to describe the sovereigns of the Roman empire and, since their time, a variety of other European sovereigns who either have held titles derived from Roman imperial titles or have adopted them in imitation. The word is also used in a loose sense for certain non-European rulers.

The Roman Emperors.—In ancient Rome *imperator* was only one of the titles by which the emperor was designated. The word was originally a military term: under the republic it was used of a magistrate who was acting abroad (*militiae*) and was thus in command of troops. His troops employed it in addressing such a magistrate, especially in saluting him after a victory; he could then use the title of *imperator* in public till the day of his triumph at Rome, after which it would lapse along with his power or *imperium*. In the later republic the senate itself on occasion invited victorious generals to assume the title; Julius Caesar was the first to use it continuously (from 58 B.C. to his death in 44 B.C. and, after 49, at home as well as abroad). Octavian (Gaius Octavius) on his adoption took the name Gaius Julius Caesar (*see* AUGUSTUS), and thus the name *Caesar* became part of the imperial titles. From 38 B.C. Octavian used *imperator* as a praenomen or forename, a practice which was to become regular; in 27 B.C. the senate gave him the cognomen of Augustus which, like Caesar, was borne by all the succeeding Roman emperors. Augustus was also addressed and referred to regularly as *princeps*—the first citizen of the state—but this was not an official title. The rule itself was referred to as *principatus* (and is often termed "the principate"), but Tacitus and other historians imply that this was a specious propaganda term for monarchy. In the early years of the principate the emperor wished to stress its dyarchic character, as the joint rule of senate and first citizen of the state; thus Tiberius claimed to be *dominus* of slaves, *imperator* of soldiers and *princeps* of everyone else. The most despotic of the emperors adopted arrogant titles which gave offense; Domitian, for instance, that of *dominus et deus* ("master and god"). Domitian's immediate successors Nerva and Trajan, however, reverted to the terminology of *principatus*.

Under the empire the title *imperator* was conferred by the senate, and an emperor normally dated his reign from the day of his salutation by the senate. But the troops were also regarded as still retaining the right of saluting an *imperator;* and there were emperors who regarded themselves as created by such salutation. This sometimes resulted (repeatedly during the second half of the 3rd century) in making the emperor the nominee of a turbulent soldiery. It was this evil which Diocletian strove to remove by reorganizing the principate so that he and his colleague Maximian were theoretically equal as *Augusti;* from A.D. 293, moreover, there were two junior colleagues with the title *Caesares*. The two *Augusti* were the joint emperors of a single empire; but from the time of the accession in 395 of Arcadius and Honorius there were in effect two empires as well as two emperors, one of the east with his capital at Constantinople and one of the west with his capital at Rome.

The practice of adopting divine titles was regular during the empire, having originated when Julius Caesar was pronounced a *divus Julius* after his death. Augustus when still Octavian had been *divi Juli filius* although as emperor he rejected the most formal divine honours.

The Byzantine Emperors.—Whereas the succession of emperors at Rome ended with the deposition of Romulus Augustulus in A.D. 476, the Byzantine emperors at Constantinople claimed to continue the Roman succession from Augustus. They were the supreme legislators, judges, military commanders, the protectors of the Orthodox Church and the guardians of Christian doctrine. From 629 the Greek title *basileus* (previously equivalent to "king") replaced the titles *imperator*, Caesar and Augustus as the emperor's official designation. In accordance with the tradition of Roman magistrates, however, the emperors were in theory elected to the empire, not emperors by hereditary right, and their sovereignty was personal, so that they could not arbitrarily alienate any territory of the empire. Even so, their formal election was regarded as the expression of God's will, and their coronation by the patriarch of Constantinople gave them a sacred character. An absolute monarch, the Byzantine emperor was venerated as God's representative on earth. (*See* BYZANTINE EMPIRE: *Government: The Church*).

The word "emperor" is also used occasionally instead of "tsar" (from Lat. *Caesar*) for the medieval Bulgarian and Serbian rulers of the Balkans.

The Emperor in Western Christendom.—Charlemagne, king of the Franks and of the Lombards, was unexpectedly crowned emperor by Pope Leo III in Rome on Christmas Day, 800. Thenceforward till the fall of Constantinople in 1453 there were two emperors in the Christian world, the Byzantine and the western. (For their relations and for the development and diminution of the imperial concept in western Europe, *see* HOLY ROMAN EMPIRE.) The term "Holy Roman emperor" is now generally used, for convenience, to designate the Western sovereigns, though the title was at first simply "emperor" (*imperator;* the German form *Kaiser* being derived from *Caesar*), then "august emperor," then, from 971, "Roman emperor." The addition of "holy" to the designation of the emperor, in historical writing, follows from its having been added to that of the empire (*sacrum imperium*, 1157): earlier references to the "holy" or "most holy" emperor appear to be occasional allusions to the emperor's consecration by the pope or to his divine right as king rather than formal titles. In western Europe, from 800 to the end of the empire in 1806, the conventional designation was merely "the emperor," which remains the best for historical purposes wherever there is no reasonable chance of ambiguity.

The dissolution of Frankish Europe into separate kingdoms under the later Carolingian sovereigns led eventually to the imperial title's passing in 962 to the East Frankish or German king Otto I, who was also king of Italy (the kingdom of Burgundy was further acquired by Conrad II in 1032). Thenceforward to 1806, though not all German kings were emperors (crowned by the pope), there were no emperors who were not German kings, so that election to the German kingship (*see* ELECTORS) came to be *de facto* necessary for attainment of the imperial title—with the final result that from 1508 to 1806 the style "emperor elect" or, more briefly, "emperor" was given to the German king in anticipation of his papal coronation (only one such coronation, that of Charles V in 1530, actually took place in the period).

In the west outside the Frankish and German sphere of influence the title emperor was sometimes assumed by princes supreme over more than one kingdom: thus Sancho III the Great of Navarre styled himself "emperor of Spain" on his annexation of Léon

(1034); Alfonso VI of Léon and Castile called himself "emperor of the Two Religions," to show his supremacy over Christians and Muslims alike; and Alfonso VII took the title "emperor of all Spain" (1135). The ascription of the title to certain Anglo-Saxon kings is thought to be an invention of the 12th-century writers.

The Later European Empires.—The Russian tsar Peter I the Great assumed the title *imperator* on Oct. 22, 1721. This not only suggested equality of status between the Russian sovereign and the Holy Roman emperor but also was a step toward promoting Moscow's claim to be "the Third Rome," the successor of Orthodox Constantinople against the Roman Catholic west. After the French Revolution had destroyed the kingdom of France, Napoleon Bonaparte in 1804, having been annointed by Pope Pius VII, crowned himself emperor of the French as Napoleon I. His claim to be the successor not of Louis XIV but of Charlemagne, together with his organization of the Confederation of the Rhine in Germany, was a threat to the Holy Roman empire of the Habsburg dynasty. Seeing this, Francis II, to retain an imperial title, took that of "hereditary emperor of Austria" before he dissolved the old empire. His successors retained it till 1918.

Napoleon III was emperor of the French from 1852 till his deposition in 1870–71 (the French second empire). Between 1871 and 1918 the kings of Prussia, William I, Frederick III and William II, were German emperors. Victoria of Great Britain took the title empress of India in 1876, but her great grandson George VI renounced the imperial title when India became independent.

Other Uses of the Title.—In the western hemisphere Jean Jacques Dessalines was emperor of Haiti from 1804 to 1806; princes of the house of Braganza were emperors of Brazil from 1822 to 1889; Augustín de Iturbide and the Austrian archduke Maximilian were emperors of Mexico from 1822 to 1823 and from 1864 to 1867 respectively. The title emperor is also generally used as the English designation for the sovereigns of Ethiopia and of Japan, for the Mogul rulers of India, for the former sovereigns of China, for the Inca rulers of Peru and for the Aztec rulers of Mexico.

EMPEROR FISH (*Pomacanthus imperator*), a large and brilliantly coloured coral fish of the family Chaetodontidae (butterfly fish and angelfish; *qq.v.*), reaching a length of 15 in. and found from east Africa to Malaya. It is an excellent table fish.

EMPETRACEAE, a family of low, evergreen shrubs, the best-known member of which, the crowberry (*q.v.*), gives the group its common name. There are three genera and eight species, confined to North and South America, extending to the arctic and the antarctic, and in the circumpolar regions of the northern hemisphere. The parts of the flowers are mainly in threes, and the fruit is a fleshy or dry berry or drupe.

EMPHYSEMA, a medical term used to describe an overdistention of the lungs because of dilatation and coalescence of the alveoli, or air sacs. This leads to the condition known as emphysematous lung and the barrel-shaped chest. It occurs in players of wind instruments and is caused by continued high pressure within the alveolar spaces. Sometimes there is rupture of lung tissue and air escapes into the interstitial tissues of the lung, thus producing an interstitial emphysema. This type of tissue emphysema may occur in other locations such as the space between the two lungs and the subcutaneous tissues, as in the case of fractured ribs which tear a lung and allow air to escape into surrounding tissues.

Air or gases may also accumulate in the body cavities, causing conditions described as pneumothorax (in the chest) or pneumoperitoneum (in the abdomen). There is no specific cause for these conditions; they may be the result of trauma or disease processes. Infection with gas-producing organisms may produce interstitial emphysema; the commonest infectious organism of this type is the bacillus of gas gangrene (*Bacillus aerogenes capsulatus*). *See* also Lung, Diseases of; Respiratory System, Diseases of.

(F. L. A.)

EMPIRE, a term usually applied to a state of large size exercising political dominion over others, with or without the latter's consent. An empire is usually highly centralized, but it may be a federation like the German empire from 1871 to 1918, a unitary state like the Russian empire of the tsars, or variegated like the British empire. The Commonwealth of Nations, it should be noted, consists of the coequal dominions and is now differentiated from the more centralized empire consisting of the United Kingdom and the crown colonies and protectorates. An empire is usually governed by an emperor, though this is not the prime prerequisite. The empire-building process differs from that of nation-building in that, ostensibly, the latter involves the bringing together of people of some homogeneity and usually in contiguous territory (*see* Nationalism), but it is often difficult to separate the two processes.

The motivation behind imperial expansion has been multiple and complex. At different times and in varying amounts economic, strategic and psychological factors have caused one state to extend its suzerainty over others. The question of which is primary has been endlessly debated, but it would seem that they are usually so closely interwoven that it is impossible to separate them.

The history of empires is intimately bound up with the spread of civilization. Stronger civilizations either imposed their culture upon weaker peoples, or, when the conqueror was more powerful but culturally inferior, absorbed knowledge from the defeated. "Conquered Greece made Rome her captive." This has sometimes been used speciously as a rationalization for empire. Knowledge may well spread through peaceful contact and trade. That the imperial process has not always been beneficial for mankind may be seen by the sterile and militaristic Assyrian conquest in the near east during the first millennium B.C.

THE ANCIENT WORLD

The two centres of western civilization of ancient times, the Tigris-Euphrates valley and the Nile valley, saw several conquests over the years. More vulnerable to attack than partially sheltered Egypt, Babylonian civilization was invaded and ruled in turn by the Akkadians (*c.* 2600 B.C.), Amorites (*c.* 2100 B.C.), Kassites (1750 B.C.), Assyrians (8th and 7th century B.C.), Chaldeans (7th century B.C.), Medes (7th century B.C.) and Persians (6th century B.C.). (*See* Babylonia and Assyria.) Of these, the Assyrian empire in the "fertile crescent" and the Persian empire extending from the Indus river to Greece were the most impressive. The Assyrian empire collapsed, as it lived, by the sword. The Persian empire seemed to culminate the process of cultural diffusion going on for three millenniums. Covering the widest area, 3,000 mi., it was divided into 21 satrapies ruled by civil and military officials. A measure of beneficence was to be found in the tolerating of local customs as long as tribute was paid. Fine roads facilitated military and commercial traffic and helped to knit the empire together. And while the great monarch was absolute, Persian civilization made a virtue of cultural heterogeneity.

In fact, the military conquests of Alexander the Great in the 4th century B.C. appear more understandable if one sees his exploits as reversing the direction of expansion, made easier by the unity of the area provided by the Persians. Alexander, though at the head of Greco-Macedonian forces, sought to fuse Hellenic and near eastern culture. He also adopted the quasi-divine accoutrements of the Persian monarchs, later to be taken on by the Roman Caesars. Though perhaps the greatest and best-known of the ancient conquests, Alexander's empire, covering very much the same territory as that of Darius, the Persian, collapsed with his death in 323 B.C., and was divided among his generals.

The Roman Empire.—The classic expression of empire and the greatest empire of the ancient world was the Roman. Indeed, the etymological derivation of the term empire is to be traced to the Latin word *imperium*. In its earlier use it signified generally the sovereignty of the state over the individual. It was delegated in different degree to each of the magistrates. Under the Roman empire it was gradually concentrated in the hands of the *princeps* or emperor. Augustus (27 B.C.–A.D. 14), with whom, it is generally believed, the Roman empire came into being, while avoiding the semblance of monarchy, held the chief magistracies and, in fact, concentrated all power in his hands.

What had motivated and permitted Rome to expand? The details are obscure in spots, but one explanation is that Rome expanded at first in search of defensible frontiers. Its citizens' hardi-

ness of character, its military tactics, its policy of *divide et impera* (divide and conquer), its often wise treatment of former enemies (giving them various privileges and making them allies) all help to explain Rome's conquest of the Italian peninsula by the middle of the 3rd century B.C. Victory in the Punic Wars with Carthage gave Rome control over the western Mediterranean. By 146 B.C. Greece was taken, and by the time Julius Caesar died the Anatolian plateau succumbed, Gaul was absorbed and Egypt was soon to be defeated. Central and southeastern Europe except for Germania were later secured, Britain was invaded and the Mesopotamian valley taken.

Both under the republic and the empire, Roman administration of the provinces conformed to no set pattern. Under Augustus there were imperial provinces under the emperor's control (through the proconsuls), senatorial provinces and client states. Many municipalities were permitted powers of self-government, and many provinces exercised it. They were taxed for Rome's benefit, but after a while, the money, instead of going for tribute, went to build public facilities in the provinces, to pay for the upkeep of the proconsul and his army, or was taken in graft. This contributed to Rome's downfall, insofar as the capital had adjusted itself to receiving taxes from the provinces, but these were difficult to collect in later years.

Rome's imperialism brought with it the *Pax Romana* which for two centuries beginning with Augustus conferred peace, equitable law and good administration over a wide area. Peace, good roads, two common languages (Latin and Greek), and good law fostered the accumulation and dissemination of culture. And among Rome's greatest accomplishments was that it acted as a vehicle for the absorption, digestion and transmission of Greek and eastern culture to western Europe.

The empire, which lasted for 500 years dating from Augustus' reign, did not die in a day any more than it was built in one. The reasons for the fall of the Roman empire have continued to attract scholarly attention. Apparently they were legion. Perhaps an interaction of factors provided a momentum that became cumulative. The burden of an essentially nonproductive capital that drained the empire's wealth, the growth of a landless proletariat unable to compete with large slave-worked plantations, the cost of public buildings and games, the debasement of the coinage, the spread of malaria, the fighting over the succession to the emperorship, the barbarian pressures, graft, the loss of civic pride, the Christian concentration on the eternal afterlife rather than on "eternal" Rome—all these and others intertwined to bring about the decline of the empire. And when Diocletian at the end of the 3rd century and Constantine at the beginning of the 4th divided the administration between two Augusti and set up a capital at Constantinople, the attempt to improve government resulted in further division. Roman civilization had begun to collapse. (*See* ROMAN HISTORY.) But while in 476 a barbarian chief who deposed the boy emperor did not even bother to take on the title, the eastern Roman empire, the Byzantine, lived on for 1,000 years. Its centre, Constantinople, was the largest city in Europe during the middle ages.

The idea of imperial Rome, however, persisted in western Europe and was resurrected in the Carolingian empire. In the year 800 Charlemagne, the Frankish king, somewhat to his discomfiture, was crowned emperor of the Romans. This gesture of gratitude on the part of the pope confirmed Frankish support for him as against the Lombards and the Byzantine empire. Charlemagne's long rule (768–814) did not sufficiently solidify his domains, which were divided among his three grandsons. The Carolingian interlude serves only to highlight the political anarchy before and after him. Feudalism became more deeply entrenched.

The concept of empire, nevertheless, did not die. The Christian church modeled its structure in part upon the Roman empire with its supreme governor, the *Pontifex Maximus*, and its provinces and dioceses. More important, the church conceived of itself as the inheritor of a universal domain. While unable to exert secular control over western Europe, the church, by having crowned the successors of Charlemagne, became intimately associated with the Holy Roman empire. In 962, Otto I, a Saxon prince, was crowned by the pope as first Holy Roman emperor. The association was far from friendly during a good part of the middle ages. Disputes over the anointing of bishops, the control of landed wealth and the role of religion beset the relationship. It has been commonly charged that the structure was "neither holy, nor Roman, nor an empire," but it survived until dissolved by Napoleon in 1806. Though it originally included most of Italy within its boundaries, it was basically a loose association of German states.

Arabic Expansion.—During the aftermath of Rome's decline a great expansion of Arabic power took place under the religious inspiration of the prophet Mohammed. Without the catalyst of religious motivation it is hard to see how the spread of Muslim civilization could have been so rapid. Within a little more than a century after Mohammed's death in 632, Arabic control extended from Spain to the Indus river. Later, the Muslim culture and religion penetrated India, the Indies and central Asia. But religious fervour is very far from explaining the whole story. A search for more fertile lands, familiarity with the terrain, the weakness of existing empires, the absorption of allies, a large measure of tolerance toward conquered peoples so long as tribute was paid—all these factors aided in the spread of the Muslims. Constantinople was not captured, though Byzantine lands were. The Omayyad dynasty transferred the capital to Damascus, but by 750 another dynasty, the Abbasid, moved it to Baghdad. The caliphate of Cordova in Spain did not acknowledge Abbasid supremacy. In the 10th century north Africa and Egypt were united in the Fatimid caliphate, and still later, other vassal states broke away from Baghdad's rule. Both the invasion of the Seljuk Turks and the Christian counterattack of the Crusades helped to stem the tide of Muslim advance, though, as indicated above, internal disagreement also weakened it. The Arabs themselves often attribute the slowing down of their cultural advance to the invasion of the Asian Turks.

Far East.—Empires have not by any means been confined to the western world. Not only have modern empires such as that of the Ottoman Turks been created, but long before, several large Asian dominions were established in India, China, in central Asia, and on the Iranian plateau. The essence of empire existed in Africa on occasion, and in the pre-Columbian Americas.

In ancient India, in 322 B.C., the year after Alexander the Great died, Chandragupta Maurya inaugurated an empire that covered northern India and lasted until 184 B.C. His grandson, Asoka, extended his domain to the south and became an effective patron of Buddhism. Six centuries later as Rome experienced its troubles, another Indian empire, the Gupta, flourished in the Ganges valley and across to the Indian ocean. The Mogul empire founded by Baber (Babur) in the 16th century was extended to most of India by Akbar in the 17th century.

In China, after several hundred years of political turbulence, a king of Ch'in, a northern state, took the conquest title, Shih Huang Ti, meaning First Universal Ruler. Though a powerful man, one who, for example, had most of the Great Wall finished, his rule was considered harsh and after his death the people accepted a new dynasty known as the Han. It endured for over 400 years (202 B.C.–A.D. 221) and covered most of what is now the heart of China. A feudal age followed it and then another empire, the T'ang, lasting nearly 300 years (618–906). It was highly centralized and covered much of western China as well as the area governed by the earlier Han empire. When it fell another 400-year period of smaller dynasties followed. In the 13th century the Tatar conqueror Genghis Khan united the Mongolians, overran China, swept across the Asian steppes, subjected Russia to tribute and went into Poland as well as Asia Minor. A rude people, but ferocious, the Golden Horde held sway in Russia for 200 years. One of the Mongol emperors of China, Kuhlai Khan, was a benevolent ruler who established peace and order in his kingdom. But this too was succeeded by other dynasties as war lords came back to plague China. During the 17th and 18th centuries the foreign Manchu dynasty of China received tribute from Mongolia, Manchuria, Turkistan, Burma, Nepal, Annam and Korea.

European Maritime Empires.—In their essentials, Asian and European empires bore remarkable resemblances to each other and occasionally came into contact. By the end of the middle ages,

Europe and China and India were generally in comparable stages of development. Western Europe was soon to burst from its confinement, however, and inaugurate an amazing world-wide expansion that was unique in its magnitude and operation. The economic accumulation of the previous few centuries, the growth of nation-states along the Atlantic, the increasing technical knowledge of navigation and the search for an all-water route to trade with the east that would circumvent the Muslim and Italian middlemen, aroused an outburst of maritime activity scarcely seen before. Western Europe's age of exploration had begun in earnest. The alliterative "gold, glory and the gospel" sum it up. But though Europe's energy was widespread, it did not focus politically in one universal empire. Individual countries, Portugal, Spain, England, France and Holland participated as separate nations and, usually, as rivals. From the 15th through the 18th centuries western Europe reared what has become known as "the old, colonial system." Little Portugal, with incredible energy and audacity, led the parade. Following the contour of Africa, she set up stations on the coast, defeated the Arabs along the way, reached India and pushed farther on to the source of the spices, the Indies. Remnants of this empire still existed in the early 1960s in Angola and Mozambique in Africa. In the process Portugal managed to lay claim to Brazil; Cabral having touched there in 1500. Not until 1822 did Brazil separate from Portugal. As with the other maritime empires, the economies of these colonies were joined to the mother country under the prevailing mercantilist philosophy of the period. Illustrative of much that happened in Latin America (and North America, too) was the subsequent miscegenation of Europeans, native Indians and African slaves. This biological fusion, it must be noted, has been characteristic of empires throughout the ages.

The Spanish domains included the rest of Latin America except for some West Indian islands and indeed were impressive. The pope was sufficiently impressed in 1494, to divide the world between the two powers, a decision that was ratified by the treaty of Tordesillas. The centuries that followed witnessed a struggle for empire that saw the Dutch track down the Portuguese source of supply of spices and seize the Indies. The Spanish empire, meanwhile, had been built on the taking of gold from the Incas and silver from the Mexican Indians. Conquistadors were accompanied by Roman Catholic priests as they daringly crossed the jungles to get the precious metals and convert the heathen. Spanish mercantilist policy also restricted trade with other powers, though the Spanish dependence upon Dutch, French and English goods meant that bullion was exported to pay for them. Spanish ships were prey for marauding English and other buccaneers, and ultimately English policy, which favoured and benefited by the independence of the Spanish American colonies, saw it achieved by the early 19th century. The Monroe Doctrine of 1823 which opposed the reimposition of Spanish or French control of the former possessions was given teeth by the British fleet throughout the century.

Explorers of England and France penetrated the North American interior during the 17th century. The French concentrated on the fur trade with the Indians while the English sent a stream of settlers to New England and the southern colonies. Both also went to the rich West Indies. Rivalry between the two ended in a series of wars 1689–1713, 1740–48, 1756–63 for the control of North America and Bengal. Coupled with the earlier defeat of the Dutch during the 17th century commercial wars, this victory gave to the British a commanding position in empire founded upon settlement and upon commercial and naval supremacy. If their control over the colonies was not as rigid as that of other empires, it was annoying enough to the American colonists who chafed under the now more rigorously enforced navigation laws. At the apex of imperial power, the British empire was soon to be rocked by the breaking away of the 13 American colonies. This first great rebellion against a modern empire is continually harked back to by those opposed to empire, but one must remember that the American colonists originally wanted, not separation, but rather to enjoy their rights as Englishmen.

Indifference to Imperialism.—In the aftermath of the American Revolution, followed by the revolutions in Latin America, the mercantilist philosophy was dealt a body blow by advocates of laissez faire. It was remarked that despite their having broken away, the Americans continued to trade with England. Why then ought one to go to the expense and trouble of securing and maintaining colonies? These two factors were instrumental in bringing about a relative indifference to the acquisition of colonies during the first two-thirds of the 19th century. It is true that the British extended their dominion over much of India, annexed Natal and colonized New Zealand and that the French marched into Algeria. But the European powers, except perhaps in the case of India, made no intensive effort to develop their domains. A most significant constitutional change of the period, foreshadowing the evolution of the dominion system, was contained in Lord Durham's report (1839) that called for Canadian responsible government. During the 19th century when one thought of the term empire one could just as reasonably have thought of the first and second empires of the two Napoleons in France, the empire of the tsars of Russia, the Austro-Hungarian empire, the German empire after 1871 or the Turkish empire as think of empires beyond the seas.

Revival of Imperialism.—The 1870s and 1880s saw the beginning of a scramble for possessions that brought modern imperialism into being. In the short space of two generations the African continent was carved up, and Japan had joined the western powers in subjecting China and southeast Asia to control. A bewildering mosaic of protectorates, spheres of influence, colonies, dominions and extraterritorial concessions indicated imperial control through diverse techniques. Once again, no single cause can adequately explain why a movement of such magnitude took place. To say that the British empire was acquired "in a fit of absence of mind" is misleading, though illuminating. While it does provide an antidote to the view that there was always a consciously diabolical plot for gaining territory, it gives too, a distorted perspective by implying that expansion was innocent and haphazard. The forces behind the new imperialism were numerous and interacting. A desire to sell surplus goods abroad in a world of increasing tariff barriers, a desire for raw materials, a wish to invest abroad where profits (and risks) were greater, a claim that European countries needed a place for surplus population, a search for strategic bases, a missionary spirit, a superabundant nationalism, a sense of obligation toward people considered backward, an improvement in the means of communication and transportation, a greater knowledge of tropical medicine, a desire for adventure in exotic lands, a wish to divert public attention from home problems and a rush to take over territory, whether valuable or not, in order to forestall others —all these in diverse combinations explain the revival of imperialism. There were outspoken critics of empire during the process, but despite their arguments, apostles of empire, to be found in all classes, were not self-conscious in their defense of the movement.

Not until World War I did the spread of nationalism in the colonies gather momentum. The swelling tide toward national independence was hastened by Pres. Woodrow Wilson's call for self-determination of peoples and the Bolshevik call for the elimination of "imperialism, the last stage of monopoly capitalism." Ireland (except for Northern Ireland) and Egypt became independent in 1922 and India gained greater self-government (1919, 1935). The mandates system of the League of Nations, to a small degree, curbed indifferent colonial administration. The trusteeship system of the United Nations was more effective. A measure of truculence and even embarrassment crept into the defense of empire as it began to crumble. There was, too, a more sophisticated rationalization for its being as defenders pointed to the momentous Statute of Westminster (1931) as an example of enlightened empire. This epoch-making document, anticipated in the imperial conference of 1926, stated *de jure* what had been practically *de facto* during the 20th century—the independence of the dominions. They were to be "autonomous communities within the British empire, equal in status, in no way subordinate to one another, in any aspect of their domestic or external affairs, though united by a common allegiance to the crown, and freely associated as members of the British Commonwealth of Nations." They were given the right to declare war or peace and to secede from the group. Three have done so, Eire, Burma and South Africa. In 1950 India became a republic within the commonwealth, thus beginning a new phase of

its structure. The English sovereign is recognized as the head of the commonwealth but not as the head of India. The statute is looked upon with pride by supporters of empire and is a highlight of the inter-war period.

In a frank and boastful manner German, Italian and Japanese militarists exhorted their people, as a law of nature, to expand. Their aggressions of the 1930s and 1940s were amazingly successful and almost won World War II for them. But blatant fascist slogans were countered by the Allied call for the defense of liberal western civilization, and in the name of these ideals colonial peoples appealed for independence from imperial rule. Just as World War I had led to the end of the Romanov, Habsburg, Hohenzollern and Ottoman empires, so too did World War II hasten the weakening of the Dutch, French and English imperial structures. In the early postwar period the following countries attained independence or self-government; the Philippines, Indonesia, Indochina, Korea, Malaya, Burma, India, Pakistan, Ceylon, Syria, Lebanon, Israel, Libya, Tunisia, Morocco, Ghana, and the Sudan. Outright imperial control continued for the most part in Africa but by the early 1960s nearly the whole continent had achieved independence. In short, a climax of anti-imperialism had been reached.

There is no guarantee that self-government will be good government, but the attitude of these former colonials is that the perils of independence are to be preferred to their previous status. It has often been said, especially in the case of Great Britain, that the object of the empire was to act as a "trustee" for those parts unable to stand on their own feet until they could govern themselves. This concept undoubtedly was sincerely held by many government officials and private citizens, and it is true that India, Pakistan, Burma, Ceylon, the Sudan, and Ghana celebrated their independence without war. It is also true that in all cases pressure from within these lands forced the British to make their decision. Done with greater tact and grace than the French in Indochina, Tunisia or Morocco or the Dutch in Indonesia, it nevertheless was not done simply through altruism. It may be that "granting freedom" hides a virtue born of necessity.

As one attempts to put it into perspective, it may readily be acknowledged that no movement of such complexity and scope can ever be seen to be completely good or bad. It brought the benefits of western industrialism, administration and liberalism to lands which may have resented them at first but have learned to respect and admire them. In addition, it brought racial discrimination, exploitation and psychological tensions of a new sort. Partisans may stress one or the other of these consequences of imperialism, but both have resulted from it. *See* also COLONY; IMPERIALISM; BYZANTINE EMPIRE; and articles on individual countries.

BIBLIOGRAPHY.—A. Toynbee, *A Study of History*, abridged by D. C. Somervell (1946); T. Frank, *Roman Imperialism* (1914); E. Gibbon, *The Decline and Fall of the Roman Empire* (1776–88); W. C. Abbott, *The Expansion of Europe*, 3rd ed. (1938); R. L. Schuyler, *The Fall of the Old Colonial System* (1945); C. A. Bodelsen, *Studies in Mid-Victorian Imperialism* (1924); G. Bennett, *The Concept of Empire, Burke to Attlee, 1774–1947* (1953); J. A. Hobson, *Imperialism*, 3rd ed. (1938); N. Lenin, *Imperialism*, rev. trans. (1933); P. T. Moon, *Imperialism and World Politics* (1926); A. L. Burt, *The Evolution of the British Empire and Commonwealth from the American Revolution* (1956); E. Walker, *Colonies* (1944); E. Barker, *The Ideas and Ideals of the British Empire* (1941); Sir I. Jennings, *The British Commonwealth of Nations* (1948); P. Knaplund, *Britain, Commonwealth and Empire, 1901–1955* (1956). (W. SY.)

EMPIRE STYLE was a phase of neoclassical art (*q.v.*) that flourished during the time of the first empire in France (1804–15). It was encouraged by Napoleon's desire for a style inspired by the grandeur of imperial Rome. In architecture it was exemplified by Pierre Alexandre Vignon's La Madeleine (originally the Temple of Glory), Jean François Chalgrin's Arc de Triomphe, Charles Percier and Pierre Fontaine's Arc du Carrousel and Vendôme Column; in painting by J. L. David's "Le Sacre" and Baron Antoine Gros's battle scenes; in sculpture by Antonio Canova's heroic statues of Napoleon and his family. The furniture and decorative arts showed both Pompeiian and Egyptian influences with motifs of winged victory and the laurel wreath for symbols of triumph; bees, sheaves and cornucopias for prosperity; fasces and sphinxes for conquest (*see* FURNITURE: *Empire Style*).

See L. Hautecoeur, *L'Art sous le Revolution, le Directoire, et l'Empire* (1953); Emile Bourgeois, *Le Style Empire* (1930). (WM. F.)

EMPIRICISM, a term signifying reliance on experience (Greek *empeiria;* cf. *empeiros* "skilled"), used with special implications in medicine and in philosophy.

In Medicine.—In Greco-Roman times the Empirics constituted one of the three main schools of medicine. They confined themselves to observation and accepted remedies which were discovered to work, being skeptical of all attempts to provide general theoretical explanations. Hence, from the 16th century, "empirical" came to be contrasted with "scientific" and finally degenerated into an equivalent of "charlatan" or "quack." In nontechnical contexts the word retains this meaning. An empiricist is one who trusts in experience and is skeptical of methods which, while they seem theoretically sound, have not been proved in practice.

In Philosophy.—Empiricism is the attitude of mind that emphasizes the part played by experience in knowledge against that played by reasoning. In its extreme form the empiricist doctrine is (1) that we have no ideas other than those derived from sense experience; and (2) that statements, other than those of logic, can be known to be true only from experience. From the time of Francis Bacon this attitude has been opposed to rationalism, the belief that some ideas are independent of experience and that some nonlogical truths can be known in a manner which precludes their being falsified in experience. This controversy dominated philosophy until Kant attempted a reconciliation in his "critical" philosophy. It was generally agreed that experience plays some part in knowledge, and philosophers are classed as empiricists or rationalists according to the extent to which they emphasize or minimize this part. While rationalism flourished mainly on the continent, empiricism was the dominant trend in Great Britain, and the phrase "British empiricists" is commonly used to refer to Locke, Berkeley, Hume and their lesser followers. The two main points at issue concern (1) the origin of ideas and (2) the possibility of necessary truths.

Locke denied the rationalist theory that some ideas are "innate" and likened the mind to a wax tablet on which ideas are imprinted in sensation or to an empty room furnished with ideas which can come only through the senses. All thought consists in combining, comparing and analyzing these ideas. Like his successors, Locke wrote in a style that suggests that he is writing the natural history of the human mind. But, taken as psychology, his doctrine is demonstrably false where it is not trivially true. It is no doubt true that babies are incapable of thinking before they have any sense experience; but it is not true, as Locke seems to suggest, that we apprehend the ideas of sensory qualities, such as "red," "hard" and "round," before we understand the more complex ideas of material things. Locke's thesis is, however, better understood as a philosophical thesis about the logical relationships between certain general types of concept. It is the thesis that all the words that we use to describe the world can be defined in terms of words which stand for simple qualities known to us only by experience. The controversy was mainly concerned with the question of whether the ideas of substance, cause, reality, likeness and difference can be treated in this way.

All philosophers have regarded the truths of logic as necessary; that is to say, as capable of being known with a certainty beyond the possibility of refutation in experience. Rationalists, such as Plato and Descartes, had also claimed this status for some metaphysical, moral and even scientific truths. Typical examples are that God exists, that every event has a cause and that space is Euclidean. It was held both that no possible experience could falsify these statements and also that their truth is discovered not by experience but by the natural light of reason.

This was denied by Locke, who drew a sharp distinction between statements which assert or deny the existence of something and statements about the meanings of words. Only the latter could be necessary, and that only because they are "trifling": they are the logical consequences of arbitrary decisions to use words in a certain way, and they convey no information about the world. This view was developed by Hume into the doctrine that all significant

statements can be divided into statements about "matters of fact" which are contingent (*i.e.*, not necessary) and must refer solely to what can be experienced and statements about "relations between ideas" which may be necessary. Hume examined all the more important rationalist principles and showed that they turn out to be contingent and therefore incapable of proof. He also showed that a thoroughgoing application of empiricist principles must lead to skepticism; for the generalizations, theories and laws used in science and in ordinary life cannot be validly inferred from any number of observed instances without the introduction of extralogical principles (such as the law of universal causation) which cannot themselves be proved without circularity.

The empiricist doctrine that only the truths of logic can be necessary leads to problems about the status of mathematical truths. While it would be agreed that statements of applied mathematics contain empirical elements and can therefore be classed as contingent, the truths of pure mathematics seem both to be necessary and to convey information which is not about the meanings of words. Alone among empiricists J. S. Mill adopted the course of denying their necessity; he considered them to be based, like natural laws, on the uniform experience of all mankind over many centuries and therefore held that, however unlikely this might be, experience might lead us to revise them. Largely as a result of Gottlob Frege's criticisms, this view came to be abandoned by empiricists, some of whom followed Frege and Bertrand Russell in holding that mathematics is actually a branch of formal logic, containing no symbols that cannot be defined in terms of logical symbols. Others take the more cautious view that truths of mathematics are like truths of logic.

If the word "experience" is understood in its usual sense empiricism is almost trivially true as a theory of knowledge underlying both common sense and scientific beliefs. For, however great the part allowed to theory construction in science, science is distinguished from metaphysics by the fact that it admits no statements which cannot, directly or indirectly, be tested in experience. But empiricist philosophers have seldom used "experience" in this sense, and the sense that they have given it involves them in skeptical and antimetaphysical views which are much more controversial. Locke sometimes uses "experience" as the equivalent of his phrase "idea of sensation or reflection"; and in this usage whenever a man sees or touches anything he is said to "have a visual or tactual experience." Knowledge is now limited to those qualities of which we are aware in sensation. Locke is then driven to the conclusion that we are not aware of material things, and if he had been a consistent empiricist he would have adopted a skeptical attitude to their existence. It was left to Berkeley to draw and to embrace the inevitable conclusion that material things just are collections of perceived ideas, a view which ultimately leads to phenomenalism (*q.v.*). Berkeley himself accounted for the continuity and orderliness of the world by calling in the perceptions of an unsleeping God. Hume saw that this hypothesis is as unhelpful and, on empiricist principles, as unwarranted as Locke's hypothesis of unperceivable material things. He also extended the Berkeleian analysis to the mind, which he described as a "congeries of perceptions."

In the early years of the 20th century the doctrines of Berkeley and Hume were revived by the logical positivists in opposition to the speculative metaphysics of the 19th century. They substituted the technical term "sense-datum" for the ambiguous 18th-century term "idea" and were forthright in condemning as meaningless all statements that were neither truths of logic nor statements about sense-data (*see* Logical Positivism). This extreme position was later largely abandoned, partly because of the difficulty of specifying the language which may be used to characterize sense-data and partly because it was discovered that the concept of a sense-datum turned out to be just as metaphysical as the concepts which the logical positivists hoped to eliminate. There remained, however, a strong empiricist element in British philosophy, and in the U.S. the empiricist tradition was continued by the pragmatists (*see* Idealism; Pragmatism). *See* Knowledge, Theory of; *see* also references under "Empiricism" in the Index.

Bibliography.—*The British Empirical Philosophers,* ed. by A. J. Ayer and Raymond Winch (1952), is a comprehensive selection of standard works from Locke to J. S. Mill. A. J. Ayer, *Language, Truth and Logic,* 2nd ed. (1946), provides short exposition of the extreme empiricist position. For original contributions to the empiricist tradition *see* H. H. Price, *Perception* (1932) and *Thinking and Experience* (1953); A. J. Ayer, *The Foundations of Empirical Knowledge* (1940); and Bertrand Russell, *An Inquiry Into Meaning and Truth* (1940), *Human Knowledge* (1948) and other works. (P. H. N.-S.)

EMPLOYEE ASSOCIATIONS. In private industry in the United States employee associations are organizations of employees that are concerned primarily with welfare and recreational activities. In public employment they are also active in advocating legislative and administrative action in matters of compensation and working conditions. The associations sometimes serve as agencies for providing worker participation in systems of employee representation for the discussion of mutual problems with management.

Employee associations in public employment are often difficult to distinguish from labour unions, but in general the associations may be identified by the limitation of their membership to the employees of a single employing unit and their failure to engage in collective bargaining for the negotiation of contracts of employment. Both employee associations and labour unions may be found among the employees of a single employer, in which case their relationship is typically a competitive one. In the United States employee associations or company unions developed in the late 19th century and flourished after World War I, when many were actively antilabour union. With the increasing unionization of labour beginning about 1930 company unions virtually ceased to exist and employee associations took on their present character.

Membership in employee associations may be automatic upon employment or may be voluntary. Automatic membership indicates some form of employer sponsorship in order to provide a channel of communication with employees or to furnish an administrative agency for employee benefit programs. Voluntary membership indicates less direct employer support and may mean that the organization was initiated entirely by employees and receives no official encouragement. In either case the principal activities include some or all of the following: the provision of various forms of group insurance programs (*e.g.*, life, health, automobile, etc.); the operation of a credit union; the supervision of recreational centres or vacation resorts; the operation of medical centres, co-operative stores or buying clubs; and the sponsorship of athletic teams or cultural events. Particularly in public employment, employee associations often act as representatives of individual employees in proceedings involving hiring policies, disciplinary practices or other administrative rulings.

Employee associations sometimes are converted into labour unions through affiliation with an existing union or through becoming an independent union (*i.e.*, one not affiliated with a larger body of organized labour) by entering into a bona fide collective bargaining relationship with the employer.

In Great Britain, employee associations are limited mainly to associations of professional (administrative, technical and clerical) workers, formed primarily for the preservation and improvement of professional standards. In the second half of the 20th century they had come to exercise many of the functions of trade unions, including bargaining about salary and fee scales. The largest such association in Great Britain is the National Association of Local Government Officers (NALGO), which is organized into branches and subdivided into staff committees. In Sweden, Germany and other European countries, it is common for nonmanual workers to belong to associations which are similar in many respects to trade unions. These associations do not, however, affiliate with the central organization which caters for manual workers' unions; they have their own separate national body.

See also Labour [Trade] Union. (J. W. Go.)

EMPLOYEE STOCK OWNERSHIP, a type of employee benefit plan that permits employees to purchase stock of the company which employs them generally at a lower net cost than if stock were purchased in the open market. Employee stock purchase plans first appeared in the United States in the beginning of the 20th century. During World War I and during the brief period of

prosperity that followed, the number of such plans grew rapidly, but the growth slackened considerably during the recession of 1921–22. Another period of rapid growth continued until the 1929 depression. The pattern of growth during prosperous times and decline during depressed times continued thereafter. Employee stock ownership plans are more prevalent among growth companies and companies that exhibit some defensive tendencies, such as utilities and food companies. Companies whose earnings are highly volatile have been more reluctant to introduce such plans.

Employers have adopted employee stock ownership plans for a variety of reasons. They believe the work of employees will improve if they have a financial interest in the company, and that employee turnover will be reduced. Employers generally consider the stock of their own company a good investment, particularly for employees who have knowledge of their trade and are in a good position to watch their investment.

Eligibility requirements vary. Time of employment required may vary from one month to five years, with the more liberal plans applicable to those with the longer period of service. Where salary is a factor, the plan is frequently more liberal for those at the lower levels.

The stock made available for purchase is generally common stock but sometimes preferred stock is offered. It may come from treasury stock that has been bought and held by the company, unissued stock, stock bought in the open market or some form of employee's special stock. The latter is often used by closely held companies where owners do not wish to give up control. In this case the stock is usually nonvoting and carries a fixed dividend rate, although the rate may never be below the rate on the voting common.

Several methods are used in determining how much each employee may buy. A common method fixes the amount as a percentage of salary, with the percentage declining as the salary increases. Another method allows purchase of a fixed percentage of annual salary, with the percentage rising with length of service.

The price the employee pays varies among plans, but all have some advantage over purchase in the open market. Where the employee must pay the market price, he is generally given a bonus for such purchase, and the bonus may vary with his length of service. Other plans provide for purchase at a discount from the market price. This discount may vary with salary or length of service.

The most common method of payment is on a monthly basis with interest payable on the balance due. The right to cancel generally rests with both the employee and the employer. Employers generally restrict their right to cancel to severance of employment or to nonpayment by employee.

Managements generally favour employee stock ownership because they believe that by this means they may obtain the greater loyalty of their employees. Trade unions were for many years strongly opposed to employee shareholding. Their opposition stemmed from the attempts made in 1889 by the South West Metropolitan Gas company in London to persuade its employees, by offering them shares, to have nothing to do with the union that had compelled the company to reduce its daily hours of work from 12 to 8. In modern times unions have become less hostile to schemes of employee shareholding, since they have learned from experience that they have made little difference to their membership or their bargaining strength. *See* also PROFIT SHARING. (F. GR.)

EMPLOYEE TRAINING. Until the advent of the Industrial Revolution the training of employees was essentially a personal relationship between the master craftsman and the apprentice (*q.v.*). The rise of modern industry changed this relationship and brought with it both improvements and problems.

By the beginning of the 20th century many U.S. corporations had established their own schools to train employees. They did so on the assumption that specialized training departments could do the job better than the individual foreman. The foreman usually had many other responsibilities and frequently lacked interest in the teaching process. The formation of such schools put the spotlight upon the need for improved training techniques. At the same time, however, the principle that the prime responsibility for training must rest with the master craftsman (today's foreman or supervisor) was temporarily lost sight of. Unsuccessful attempts to turn over training to private or public industrial schools provided further evidence that the line administrator, because he has the real responsibility, is the only one who can successfully train the employee in the particular skills needed on the job in the particular company. This is not to say that he should not receive assistance from staff departments specializing in education and training, as well as from public or private industrial schools, from universities and from outside consultants. In the provision of such assistance lies the great advance in employee training in this century.

Training-Within-Industry Programs.—Much progress was made in employee training after the beginning of World War II. After the English-speaking countries became the arsenal for the Allied cause they were forced in limited time to expand their industrial capacity to a prodigious degree. They undertook the greatest employee training program in all history. There was no time to set up additional corporation schools or to expand private or public industrial schools. Instead, carefully planned training-within-industry programs had to be instituted with every foreman and every supervisor pressed into service as trainer. This produced the attitude which still prevailed long after the war ended, that the core of good employee training is a program to train the trainer.

In the United States the federal Training Within Industry service was initiated and enjoyed marked success during World War II in teaching foremen and supervisors to be good trainers. So successful was the program that after the war some of its directors, outstanding former executives in personnel and training, incorporated the Training Within Industry foundation as a nonprofit organization to assist business firms with their training programs on a fee basis. The foundation offered such programs as "Job Instructor Training," "Job Methods Training" "Job Relations Training" and "Job Economics Training."

Job economics training represented an important new development in employee training. This program was designed to help foremen and supervisors teach their employees how the economic system operates. This moved over from the strict teaching of "how," which is usually called training, to the "why," which is thought of as education. In employee training it became difficult to draw the line between education and training. In fact, in many companies the former training department is called the education department.

Most employers are aware that good training saves time and reduces expense and that productivity is strongly influenced by the employee's attitude and his understanding of company problems and objectives. Education as well as training is needed, and it has become common for a company to precede its job training with a brief but comprehensive orientation of the new employee to the company's organization, policies and business philosophy. The induction programs of many firms also include a careful explanation of the free-enterprise system, with emphasis upon the individual's responsibilities as a citizen.

Management Training.—During the depression of the 1930s many companies felt they could not afford to hire their normal complement of new employees; during World War II young men were not available for hire. As a result, there was a lack of middle management personnel after the war. Young men who returned from military service had to be trained and educated quickly for management responsibilities. This gave great impetus to management development programs. Such programs had been adopted by some of the larger corporations after World War I but not on so large a scale as after World War II.

This particular type of training went through an interesting cyclical phase. At the turn of the century the entire emphasis was upon self-development. It was up to each individual to prepare himself as best he could for management responsibilities. There was a laissez-faire attitude in most companies which presumed that, left alone to fend for themselves, the best men would eventually come to the top prepared to assume management responsibilities successfully. Unfortunately, all too frequently the best men were sidetracked or defeated by internal politics and did not reach the top. Moreover, those who assumed managerial po-

sitions were more successful in meeting ruthless competition than in understanding management, which was becoming increasingly complex. Some of these men were obviously unfit for management responsibilities and yet somehow had reached high positions.

At the close of World War I some of the larger U.S. corporations had instituted formalized recruiting programs, particularly on college campuses, in a search for young men with management potential. These men, hired in groups, were put through planned training programs. At the outset these programs consisted of little more than an orderly and logical sequence of job rotation. Later, formal classes were added and there developed skilful blends of orientation, correspondence courses, formal classes, apprenticeship, job rotation and supplementary higher education.

Most of these programs proved to be worthwhile as a basis for management development. Their formal utilization of group training was necessitated by the numbers involved and the need to control cost. It led, however, to attempts to carry such formalized training on up through the higher echelons of management and most of these attempts failed dismally.

Failure was the result of the pendulum's having swung too far from the concept that the best management development must rest upon self-development and that the boss is in the best position to guide and motivate such self-development. As a man matures in a management position the more his developmental needs become peculiar to his own special situation and the more difficult it is to fit him into a formal program. Employee training for management has at least two phases—formalized group training of young men deemed to have management potential and personalized individual training with accent upon self-development of men for managerial positions.

Tuition Refund System.—A common policy in employee training at almost all levels of job difficulty is the tuition refund system. As a means of encouraging self-development through further education, many concerns offer their employees the opportunity to take courses at local schools or colleges at company expense. In order to provide the supervision needed to avoid abuse the education department or its equivalent usually guides, directs or controls the courses an employee takes under such a system. The majority of companies reserve the right to restrict the subsidy to courses that have a direct bearing upon the job the employee is doing or for which he is preparing himself. A few companies go so far as to assume that any form of education or training will improve the employee's value to the organization.

Companies have learned from experience the value of requiring the employee to participate in the payment of his tuition expense. Refund of his share is made contingent upon successful completion of the course. Some corporations, in order to stimulate the employee to gain maximum benefit from the course, give bonuses for outstanding achievement and sometimes relate salary increases to the student's performance.

It is safe to say that business management in the English-speaking countries has come to realize that experience alone is not the best teacher because, in addition to being slow and expensive, it can also be deficient. Corporation executives have learned that it is economically unsound to permit employees' talents to remain undeveloped or be late to blossom. The interests of the employee and the company are best served by planned training under skilled trainers. Finally, the improvement in employee morale and the increase in individual productivity are vital to corporations as manpower costs continue to rise.

BIBLIOGRAPHY.—R. B. Hersey, *Better Foremanship: Key to Profitable Management*, 2nd ed. (1955); D. H. Fryer *et al., Developing People in Industry* (1956); B. von Bleicken, *Employe Training Handbook* (1953); H. B. Haas and C. H. Ewing, *Tested Training Techniques* (1950); E. G. Planty *et al., Training Employees and Managers* (1948). Papers of American Management Association: L. F. Urwick, *Management Education in American Business* (1954); *Practical Methods of Management Development* (1951). Papers of National Industrial Conference Board: *Company Programs of Executive Development* (1950); *Developments in Supervisory Training* (1952); *Employee Education* (1951). (H. P. A.)

EMPLOYERS' ASSOCIATIONS: *see* TRADE ORGANIZATION.

EMPLOYERS' LIABILITY: *see* WORKMEN'S COMPENSATION; CASUALTY INSURANCE: *Great Britain and Europe: Employers' Liability.*

EMPLOYMENT AGENCY (EXCHANGE), an organization to help workers find employment and employers to find workers. Employment agencies may be either privately owned or publicly provided or managed. Their services are available to the unemployed, to those who seek different or better jobs, and to employers. An employment agency may charge fees of the employer, the worker or both, or it may defray the expenses of administration from public or private funds. Its scope may be limited to certain trades and occupations or to classes of workers (skilled or unskilled, male or female). An agency may be local, national or, in some exceptional instances, international in its scope. In some countries or under certain circumstances the notification of vacancies may be obligatory upon employers, while refusal of suitable work offered may entail suspension or disallowance of an applicant's unemployment benefits. (*See* UNEMPLOYMENT INSURANCE.)

In addition to private employment agencies that charge fees for their services, there have been attempts by trade unions, employers' associations and philanthropic societies to organize the placing of labour without charge. These have generally failed for various reasons. An exception is the successful placement of painters in Great Britain by the trade unions in the painting industry.

Both in periods of national emergency, when labour shortages are widespread, and in normal times, public employment offices play a vital role in uniting the "jobless man and the manless job." They are relied upon to provide comprehensive information about opportunities for employment and to disseminate this information to those who need it. The auspices must be absolutely impartial. Employers and wage earners must have faith in the accuracy and reliability of the information.

UNITED STATES

Early History.—The first publicly financed employment offices in the United States were established by individual municipalities. The first move in that direction was made by the city of New York in 1834. Many years later, in 1886, San Francisco appropriated $5,000 for the support of a free employment office. Several other cities established free employment exchanges during the depression of the 1890s, and gradually the number operated by municipalities increased. The early free employment offices usually catered to unskilled and casual labour.

The recurrence of periodic unemployment, the abuses of private employment agencies, and the lack of farm labour in many states were largely responsible for the entry of several states into the field of employment services. In 1890 the state of Ohio established a state-directed system and a number of other states quickly followed suit. By 1923, public employment office laws had been enacted in 32 states. Municipal employment offices continued to exist in some cities. These offices were quite inadequate and failed to provide an effective organization of the labour market in their communities. Little attention was devoted to problems of administration; there was no uniformity in their records or standards, no co-operation between offices, and no definite policies of office management. Moreover, with few exceptions, their operations were on a small scale, their methods unbusinesslike, and their statistics valueless.

The state services gradually became aware of the need for administrative, program and policy ties with the state-administered offices in other states. This clearly called for a federal institution capable of tying the state agencies together into a cohesive, integrated system with comparable programs, policies, standards and operating practices in all states.

Nationwide Federal-State System.—As early as 1919 the Kenyon-Nolan bill was introduced in congress to establish a federal-state employment service system, but it failed to pass. In 1931, Sen. Robert F. Wagner of New York introduced a series of bills to provide such a federal co-ordinating agency. Congress passed such a bill in 1931, but Pres. Herbert Hoover vetoed it because he believed a straight-line federal agency was needed to meet the overwhelming problems then disturbing the whole economy.

The Wagner-Peyser act, providing for a federal-state system of public employment offices, was passed in June, 1933. It abolished the former U.S. employment service, which had until then operated the federal offices, and created in its place a new bureau in the department of labour. The new U.S. employment service (USES) was charged with encouraging the establishment of state-administered employment offices throughout the nation and providing for federal grants to defray operating costs. The law made it the duty of USES to ". . . promote and develop a national system of employment offices for men, women and juniors, . . . to maintain a veterans' service to be devoted to securing employment for veterans, to maintain a farm placement service, to maintain a public employment service for the District of Columbia, . . . to assist in establishing and maintaining systems of public employment offices in the several states. . . ." The federal employment service was to prescribe minimum standards, develop uniform administration and statistical procedures, publish employment information, and promote a system of clearing labour between states.

Creation of a federal-state employment service in the midst of a serious economic depression inevitably involved the new agency in developing emergency programs for public works and work relief projects. As a result, the national re-employment service was set up under the direction of the U.S. employment service to refer workers to public works and relief projects. The USES and the national re-employment service were promptly involved in mass registration and referrals of several million jobless wage earners; it began the accumulation of occupational information which later provided the basis for the *Dictionary of Occupational Titles*. The *Dictionary of Occupational Titles* was introduced to the states in 1939, giving local offices, for the first time, a comprehensive and uniform system of classifying skills and job orders. This indispensable tool for selecting workers in accordance with employers' specifications has been in continuous use ever since.

Following enactment of the Social Security act in 1935, all states which sought to participate in the federal-state unemployment insurance program were required to provide that such benefits could be paid only through a public employment office. This new responsibility required complicated administrative procedures and forced a major expansion in both federal and state services. While it was not intended that the public employment service become a subsidiary of unemployment insurance or that it limit its services to finding jobs for the unemployed who sought unemployment benefits, by 1936, however, within a few years after the enactment of the enabling legislation, a state employment service had been established in all states operating in collaboration with the U.S. employment service.

During World War II and the Korean War the nation relied heavily upon the public employment services for the allocation of manpower in accordance with over-all national objectives. The federal-state employment service gained experience and achieved a steady and substantial growth in both the variety and the effectiveness of the services provided through its 1800 employment offices. It proved flexible and adaptable to constantly changing conditions. As a result, the U.S. employment service was given broader responsibilities.

Appraisal of U.S. Placement Institutions.—The day-to-day job choices that workers make affect not only their own occupational futures but, in the aggregate, these decisions determine the number of workers available in each major skill category and the productivity of the national economy. Thus there must be active, effective institutional arrangements for getting to each worker basic facts and information about the world of work and his relation to employment opportunities each time he is about to make a major decision affecting his vocational future. It is the rule of the principal placement institutions to supply these workers' needs.

A prospective job holder, therefore, requires three types of placement information: information as to available openings, forecasts as to labour market trends, and advice as to his own capabilities and their relation to openings. An appraisal of existing placement institutions as to whether they are equipped to render such assistance necessitates an objective answer to each of these questions.

Employers and employees in the U.S. do not depend entirely on outside employment agencies, either public or private. Most large business organizations maintain extensive employment or personnel departments. And, in addition, other institutions in our society now perform placement functions of a noteworthy character. These are the school or university, the professional society, the trade union and newspaper classified advertising.

Studies indicate that the limited use of the USES does not result from a bias against government agencies as such, nor does it reflect an adverse experience with the USES during World War II. Furthermore, most employers who realize that unemployment compensation and the employment service are administered jointly do not have a bias against the employment service on this account. Many employers do not use the USES because they feel that they, better than anyone else, can choose the kind of people best suited to work in their organizations. Others do not use the USES because they are not familiar with the services of the public employment offices. Still others have tried the employment service and found it wanting. Worker reactions to the employment service usually reflect the attitudes and policies of employers. Workers who leave an employer voluntarily are not required to register with the employment service, and most of them apparently do not do so. Studies indicate that the USES has been most successful in placing unskilled and semi-skilled workers. In spite of their shortcomings, it is evident that the local and national public employment offices represent the best means of bringing job information and jobless workers together. Their association with the unemployment insurance program guarantees them access to most of the available workers in the community. Because it operates in all occupational fields, the USES has the widest range of job openings available in the community. Through the several state systems, each local office is in contact with employment opportunities in other communities that cannot be filled through local sources. It provides a vast and well-ordered supply of information about the labour market and supply and demand prospects for each occupation. Many local offices can offer testing and counseling facilities to aid the job seeker in appraising his own capabilities.

GREAT BRITAIN

Since its inception under powers conferred upon local authorities by the Unemployed Workmen act of 1905, the employment service in Great Britain has had a checkered career. It was at first the subject of much criticism but later came to be universally accepted. The modern machinery of the national employment service was created by the Labour Exchanges act of 1909, embodying most of the recommendations of the Royal Commission on Poor Laws, 1905–1909. The British ministry of labour was created in 1916. Legislation prescribing its activities was passed by the government of the day and not a year passed without one or more acts amending the administration of unemployment insurance or placing new burdens on the ministry.

By the 1960s the employment service of the ministry of labour discharged its responsibilities principally in four ways: (1) It maintained a placement service and clearing house system, made transfer grants and loans, and provided training facilities. (2) It offered special services for the disabled. The 1944 act stipulated that employers must include 3% of registered disabled workers in their work force. (3) The employment service conducted a constant review of the employment situation, with monthly surveys of the general manpower position, showing the estimated changes in the size of the total working population, the pattern of employment in the main industry groups, and the level and industrial distribution of unemployment. (4) The employment service played a leading part in the government's policy of distribution of industry. The availability of labour of the right type determined the most suitable areas for industrial development. Development of land for industrial purposes was controlled by the Town and Country Planning acts of 1947, and had to be consistent with the proper distribution of industry.

OTHER COUNTRIES

The municipal and state systems of Germany, established before

the turn of the century, served as a model for the British agencies. Since the beginning of the 20th century, other European countries have established employment agencies similar to those in Great Britain. From the proceedings of the first session of the International Labour organization in 1919 came the adoption of an international uniform system. A draft convention on employment adopted by the conference provided in its second article as follows:

"Each member (that is, each state) which ratifies this Convention shall establish a system of free public employment agencies under the control of a central authority. Committees, which shall include representatives of employers and workers, shall be appointed to advise on matters concerning the carrying on of these agencies.

"Where both public and private free employment agencies exist, steps shall be taken to co-ordinate the operation of such agencies on a national scale.

"The operations of the various national systems shall be coordinated by the International Labour Office in agreement with the countries concerned."

In addition to Great Britain, most of the members of the International Labour organization ratified the employment convention, and, where their legislation was not previously in accord, took steps, or adjusted their legislation to give effect to it.

Certain states which were not members of the International Labour organization or had not ratified the convention had within their territories employment exchanges under the control of the government, and satisfied some or all of the conditions laid down by the conference. The New Zealand government, for example, established exchanges as long ago as 1891.

BIBLIOGRAPHY.—Leonard P. Adams, *Wartime Manpower Mobilization* (1951); R. C. Atkinson, L. C. Odencrantz, B. Deming, *Public Employment Service in the United States* (1938); K. Bergstrom *et al., Administration of Public Employment Offices and Unemployment Insurance* (1935); A. Gilbert, *Administration of Public Employment Offices and Unemployment Insurance* (1936); Shelby M. Harrison *et al., Public Employment Offices: Their Purpose, Structure and Methods* (1924); Carl G. Uhr, "Employment Security in Sweden," vol. 26 (1959); T. S. Chegwidden and G. Myrddin-Evans, *The Employment Exchange Service of Great Britain* (1934); Oscar Wiegert, *Administration of Placement and Unemployment Insurance in Germany* (1934); W. Haber, "The U.S. Employment Service in a Changing Economy," in *Studies in Unemployment*, ch. 11, pp. 283–310, U.S. Senate Special Committee on Unemployment Problems (1960). (W. HR.)

EMPORIA, a city of eastern Kansas, U.S., on the Neosho river, 70 mi. S.W. of Topeka; the seat of Lyon county. The city has railroad shops, large sheep barns and some manufacturing industries, mostly processing farm products; but it is primarily an educational centre and a commercial metropolis for a diversified agricultural and stock-raising region. The original Kansas State Normal school, established at Emporia in 1863, became the Kansas State Teachers College of Emporia in 1923. The College of Emporia is a Presbyterian coeducational institution, established in 1882. Emporia was settled in 1856 and incorporated as a city in 1870; it was named for an ancient city in northern Africa. It was the home of William Allen White (*q.v.*), owner and editor of the *Emporia Gazette*. For comparative population figures *see* table in KANSAS: *Population*.

EMPSON, SIR RICHARD (d. 1510), English lawyer and minister of Henry VII, remembered, with Edmund Dudley (*q.v.*), for his unpopular administration of the crown revenues, was born at Towcester, Northamptonshire. He studied law in the Middle Temple and became a bencher there, retaining his chamber in that Inn until his death. From 1475 he served on local commissions in Northamptonshire and two years later was acting as attorney for the duchy of Lancaster. Although he had worked for Edward IV and Richard III, he remained in the royal service after Henry VII's accession. From March 1486 the king began to reward him with grants of stewardships and wardships. In 1491 Empson, one of the members of parliament for Northamptonshire, was chosen speaker of the house of commons. From 1493 to 1495 he was recorder of Coventry and ten years later was elected high steward of the University of Cambridge. From 1494 Empson was sometimes styled "king's councilor" and was working with Sir Reginald Bray whom he succeeded as chancellor of the duchy of Lancaster in 1504. He was then knighted and Henry VII joined him and Edmund Dudley by act of parliament to the feoffees responsible for carrying out his will. From that time these men were closely associated in carrying out the king's legal and financial policy, which made them so unpopular. The death of Henry VII left them without a protector and they were arrested in April 1509, on Henry VIII's accession. In August, Empson was sent to Northampton, where he was tried on a charge of constructive treason and convicted. He was brought back to London and executed on Aug. 17, 1510.

There are few personal records. Empson was called "the king of Northamptonshire" where, in 1499, he was given licence to empark 430 ac. of his manor of Easton-Neston and to rebuild and crenellate the manor house. His elder son, Thomas, was restored to his estates in 1513 and another son, John, became a barrister of the Middle Temple. His daughter, Elizabeth, married Sir Thomas Lucy of Charlecote, Warwickshire.

See articles by G. R. Elton and J. P. Cooper in *The Historical Journal*, vol. i, ii and iv (1958–61). *See* also Conyers Read (ed.), *Bibliography of British History, Tudor Period*, 2nd ed. (1959). (D. M. BE.)

EMPYEMA, a term in medicine applied to an accumulation of purulent fluid within a cavity of the body, especially the chest. *See* PLEURA, DISEASES OF: *Empyema of the Thorax*.

EMPYREAN, the name used in later Greek cosmology and pre-Copernican astronomy for the outermost and highest sphere. It was pictured as consisting of pure fire, which was supposed to have a natural tendency to fly upward. By the later pagan philosophers it was regarded as the abode of God. The term empyrean was adopted by Christian authors to designate the highest heaven and continued to be used in this way until the 18th century. Subsequently it became a vague synonym for cosmic space. (H. W. PA.)

EMS, a German town and spa, which after partition of the nation following World War II was located in the *Land* of Rhineland-Palatinate, Federal Republic of Germany. It is situated on the Lahn river, 11 mi. E. of Coblenz. Pop. (1959 est.) 9,672. In 1786 the German archbishops issued there the Punctation of Ems. Provoked by a papal move which had made the church in Bavaria subject only to Rome (1784), this denounced the monarchical structure of the church and claimed a status equal to Rome for all metropolitan sees. The Punctation, strongly influenced by Febronianism, envisaged a German national church independent of Rome.

On July 14, 1870, Bismarck published what came to be called the Ems Telegram. This was an edited version of the official report of exchanges between King William I of Prussia and the French ambassador, Vincent Benedetti (*q.v.*), at Ems on the previous day. In it the king appeared to have insulted the ambassador, which he had not. The telegram precipitated the long-prepared Franco-German War.

EMSER, HIERONYMUS (1478–1527), German editor and essayist, remembered chiefly because of a long and public controversy with Martin Luther, was born at Ulm on March 16 (26?), 1478. He studied Greek and jurisprudence, and lectured on classics at Erfurt (1504), where Luther may have been among his audience. In the same year he became secretary to Duke George of Albertine Saxony. At first Emser sided with the reformers, but he desired a practical reformation of the clergy without any doctrinal breach with the past or the church, and his liberal sympathies were mainly humanistic. The disputation at Leipzig in 1519 completed the breach between Emser and Luther. Emser warned his Bohemian friends against Luther, and Luther retorted with an attack on Emser. In the violent controversy that followed, and that lasted until his death, Emser wrote some eight tracts. He also entered into a controversy with Huldreich Zwingli, took an active part in organizing a reformed Roman Catholic Church in Germany, and in 1527 published a German version of the New Testament as a counterblast to Luther's. He died on Nov. 8, 1527.

EMSLAND, an area on the lower Ems in northwest Germany, forms a narrow strip running north to south, north of the town of Lingen between the Ems river and the Dutch frontier. It runs

Painted enamel beaker known as the "Monkey cup"; Flemish, 15th century

Cloisonné and bronze double vase with monster figures as supports. Chinese, Ch'ing dynasty, reign of Yung-cheng (1723–35). Height 28⅞ in.

Martyrdom of St. Thomas à Becket. Two copper gilt plaques with champlevé enamel, fragments of a chasse. French, Limoges, early 13th century. 5⅛ by 6⅛ in.

Top, part of an arch, a copper gilt plaque of champlevé and cloisonné enamel; 5 13/16 by 1 3/16 in. Bottom, fragment of a shrine, a plaque of the same materials and workmanship, 1 7/16 by 3 3/16 in. Mosan or Rhenish, after 1150

Copper gilt shaft with champlevé enamel. German, Cologne, c. 1180. 7⅛ by 1 1/16 in.

The Crucifixion, a copper gilt plaque with champlevé enamel. French, Limoges, 1st half of the 13th century

PAINTED, CLOISONNÉ AND CHAMPLEVÉ ENAMEL

BY COURTESY OF (TOP LEFT) THE METROPOLITAN MUSEUM OF ART, THE CLOISTERS COLLECTION, PURCHASE, 1552, (TOP RIGHT) THE BROOKLYN MUSEUM COLLECTION, (ALL OTHERS) THE LUCY MAUD BUCKINGHAM MEDIEVAL COLLECTION, THE ART INSTITUTE OF CHICAGO

ENAMEL

Early far-eastern enamel. Cloisonné-decorated back of a silver mirror made in China before the 9th century and deposited in the Shoso-in at Nara, Japan

Lineless cloisonné vase made by Sosuke Namikawa at Tokyo in the late 19th century

Lineless cloisonné enamel made by the Ando Cloisonné company, Nagoya, Japan, in the first half of the 20th century

Another fine example of lineless cloisonné, also made by the Ando company

CLOISONNÉ ENAMELS OF CHINA AND JAPAN

BY COURTESY OF (TOP LEFT) IMPERIAL HOUSEHOLD OF JAPAN. PHOTOGRAPHS, (TOP RIGHT, BOTTOM LEFT, BOTTOM RIGHT) NAOSHI TOMOBE

Cloisonné snuff bottle with coral stopper; height 2⅜ in. Ch'ing dynasty, reign of Ch'ien-lung (1736–96)

Beaker-shaped imperial vase with gilt bronze dragon and phoenix handles. Turquoise blue ground, flowers and vines in typical Ming colours; height 41½ in. Reign of Ching-t'ai (1449–57)

Incense burner in cloisonné with stand and cover of same material and design. Turquoise blue ground; 24½ by 12½ in. Ch'ing, reign of Ch'ien-lung

Imperial teapot of painted enamel. Ch'ing, reign of Ch'ien-lung

Incense burner of cloisonné on copper, in three parts, bowl, ring and cover. It is decorated with floral designs and openwork on turquoise blue ground; height 16 in. Ch'ing dynasty

Jui sceptre (repaired); cloisonné. Length 15¾ in. Ch'ing, reign of K'ang-hsi (1661–1722)

Cloisonné trumpet-shaped vase decorated with eyes and cloud forms on turquoise blue ground; height, 14½ in. Ming, reign of Ching-t'ai

Cloisonné dove-shaped altar wine vessel on wheels, patterned after a type of ancient ritual vessel. Besides the conventional ornamentation, the decoration consists of scrolls and various designs such as monsters' heads, copied from very old bronzes; height 22 in. Ch'ing, reign of K'ang-hsi

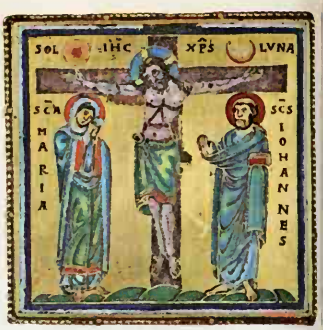
CHINESE ENAMELS OF THE MING AND CH'ING DYNASTIES

BY COURTESY OF (TOP LEFT, TOP RIGHT, SECOND ROW RIGHT, BOTTOM CENTRE) THE METROPOLITAN MUSEUM OF ART, NEW YORK CITY, (TOP CENTRE, BOTTOM LEFT, BOTTOM RIGHT) THE DEPARTMENT OF FINE ARTS, BROOKLYN MUSEUM; (SECOND ROW LEFT) S. AND G. GUMP COMPANY

ENAMEL

CLOISONNÉ AND CHAMPLEVÉ ENAMEL

The baptism of Christ, a champlevé enamel plaque attributed to Godefroid de Claire. Mosan, 12th century



Champlevé plaque of the crucifixion attributed to Godefroid de Claire. Mosan, 12th century

Pendant in the form of a monkey, in gray-blue enamel ornamented with gold and rubies. German, 16th century

Christ driving the money-changers from the Temple, by Pierre Reymond; painted enamel. Limoges, 16th century

Adoration of the shepherds, enamel painted on copper. Limoges, 16th century

Medallion of St. George; cloisonné enamel on gold. Byzantine, 11th century

Square chasse with scenes from the life of Christ; champlevé. Limoges, 13th century

Plaque showing the Virgin with St. Simeon holding the Christ Child; cloisonné enamel on gold. Byzantine, 12th century

BY COURTESY OF THE METROPOLITAN MUSEUM OF ART (TOP LEFT, TOP CENTRE, TOP RIGHT, BOTTOM LEFT) GIFT OF J. PIERPONT MORGAN, 1917, (SECOND ROW LEFT) FLETCHER FUND, 1945, (SECOND ROW RIGHT) BEQUEST OF BENJAMIN ALTMAN, 1913, (BOTTOM CENTRE) GIFT OF GEORGE BLUMENTHAL, 1941, (BOTTOM RIGHT) GIFT OF ANN PAYNE BLUMENTHAL, 1938

through Bourtanger moor and is a belt about 10 to 15 km. (6–9 mi.) wide and about 100 km. (62 mi.) long. Emsland includes three strips of land running north to south—the bog of the Bourtanger moor that stretches across the frontier; a narrow strip alongside it that has better drained sandy soil and contains the main villages; and the flood plain of the Ems that is covered by drained meadowland. Peat bog still covers large areas, but much of it has been drained and brought under cultivation, sometimes by cutting the peat, sometimes by the deep plowing of the peat and the underlying sands. Small holdings with rye and potato crops are found in this area. The Ems valley has much milk production. The district is predominantly Roman Catholic. East of the river are higher sandy geest lands that were divided into several political units with connections with Hanover, whereas Emsland had close ties with the bishopric of Münster before the secularization of its territory in 1815. There was a big increase in oil production in the 1950s. The oil-bearing beds are associated with anticlines of Cretaceous sandstones that are a continuation of the strike of the Teutoburger Wald. The beds lie at depths of 1,300 to 2,800 ft. below the surface. German oil production in 1957 reached 4,000,000 tons, almost all of it from Emsland. (R. E. Di.)

EMS RIVER, Germany, rises on the south slope of the Teutoburger Wald at an altitude of 358 ft. and flows generally northwest and north through Westphalia and Hanover to the east side of the Dollart, immediately south of Emden. After passing through the Dollart the navigable stream bifurcates, the eastern Ems going to the east and the western Ems to the west of the island of Borkum to the North sea. Length 230 mi. The Ems has a marked winter maximum and summer minimum flow.

Between 1892 and 1899 the river was canalized for a distance of 43 mi. At the same time, and as part of the plan to provide an entirely German outlet for the Westphalian industrial district, the Dortmund-Ems canal was constructed. It runs from Dortmund to Meppen (94 mi.) where it joins the canalized Ems (55 mi.) and then follows the open river to Emden. Altogether, the canal is 167 mi. in length, has a surface width of 99 ft. and a depth of 10 ft. The total fall is 230 ft., overcome by 19 locks and a ship lift. It will take boats of 750 and even 800 tons. From Henrichenburg a branch runs to Herne, which is connected with the Rhine near Ruhrost by the Rhine-Herne canal. There is a difference in level of 46 ft. between the main canal and the Herne branch, and this necessitated a huge lift at Henrichenburg. The canal is connected with the Mittelland canal system at Bevergern. Northward traffic is mostly in coke and coal from the Westphalian coal field, to which are brought imported iron ore and other raw material and foodstuffs. (R. E. Di.)

EMU (Emeu), the name given to ostrichlike birds of the family Dromiceiidae. With the cassowaries (*q.v.*), the emus form the order Casuariiformes, peculiar to the Australian region and characterized by the fact that the aftershaft of each feather is as long as the main shaft.

BY COURTESY OF NEW YORK ZOOLOGICAL SOCIETY

FLIGHTLESS AUSTRALIAN EMU AND CHICKS

The emus differ from the cassowaries in having a duller colour, feathers on the head, no helmet or cervical fleshy outgrowths, and in bearing on their inner toes a peculiar claw. The only living species is *Dromiceius novaehollandiae* of Australia. The emu, second largest of extant birds (the largest being the ostrich), has been much diminished in numbers by constant persecution. It inhabits open country, feeding on fruits, roots and herbage and keeping in small companies. The nest, a shallow pit in the ground, contains 9 to 13 blue-green eggs, which are incubated by the cock, who curiously enough chases his mate away. The young at birth are clad in striped down.

The peculiar structure of the trachea is correlated with the loud booming note of the bird during the breeding season. Three subspecies are recognized, inhabiting northern, southeastern and southwestern Australia, whereas a fourth race, now extinct, formerly lived on Tasmania. Other species, also extinct, once inhabited nearby King and Kangaroo islands. (G. F. Ss.)

EMULSION, in physical chemistry, a mixture of two or more liquids in which one is present in droplets of microscopic or ultramicroscopic size. Emulsions are formed from the component liquids either spontaneously or, more often, by mechanical means, such as agitation, provided that the liquids that are mixed have no, or a very limited, mutual solubility. Emulsions are, or can be made, stable by the presence or addition, respectively, of emulsifying agents which form films at the surface of the droplets (*e.g.*, soap molecules), or which impart to them a mechanical stability (*e.g.*, colloidal carbon or bentonite). Unstable emulsions separate with time into two liquid layers. Stable emulsions can be destroyed by inactivating or destroying the emulsifying agent; *e.g.*, by adding appropriate third substances or also by freezing or heating. Some familiar emulsions are milk (*i.e.*, fat droplets in an aqueous solution) or butter (*i.e.*, droplets of an aqueous solution in fat).

Emulsions are important in many fields; *e.g.*, in the dyeing and tanning industries, in the manufacture of metals from low-grade ores by flotation (*see* Metallurgy), in the manufacture of synthetic rubber and plastics, in the preparation of cosmetics, such as shampoos, and of salves and therapeutic products.

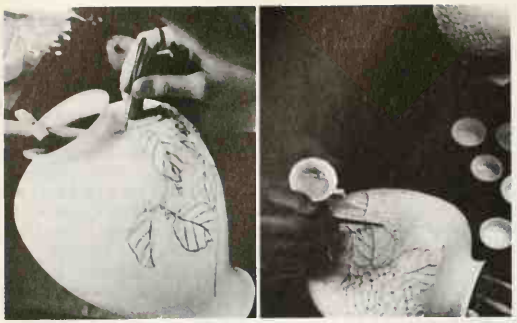
The term emulsion is often applied to mixed systems which should better be characterized as solutions, suspensions or gels. For example, the so-called photographic emulsion is actually a gelatine gel in which tiny crystals (*e.g.*, of silver bromide) are dispersed. *See* also references under "Emulsion" in the Index volume. (Wi. H.)

ENAMEL, in art, is a vitreous glaze or combination of vitreous glazes fused on a metallic surface. The general term enamels is conveniently applied to objects which have this material as the principal decoration. (For a discussion of the hard, glossy paint known as enamel, *see* Paint.) The origin of the art is obscure, and until all the available early material has been analyzed and proved to be true enamel, as distinct from cut stone (lapis) or glass applied cold, it is impossible to do much more than repeat the conjectures of the writers of the past on its early history. (H. Tt.)

ENAMELING PROCESSES

The base of enamel is a clear, colourless, transparent vitreous compound called flux, which is composed of silica, minium and potash. This flux or base—termed *fondant* in France—is coloured by the addition of oxides of metals while in a state of fusion, which stain the flux throughout its mass. Enamels are either hard or soft, according to the proportion of the silica to the other parts in its composition. They are termed hard when the temperature required to fuse them is very high. The harder the enamel the less liable is it to be affected by atmospheric agencies, which in soft enamels produce a decomposition of the surface first and ultimately of the whole enamel. Pure—or almost pure—metals are in most respects the best to receive and retain the enamel. Enamels composed of a great amount of soda or potash, as compared with those wherein red lead is in greater proportion, are more liable to crack and are less cohesive. It is better not to use silver as a base, although it is capable of reflecting a higher and more brilliant white light than any other metal. Fine gold and pure copper as thin as possible are the best metals upon which to enamel. If silver is to be used, it should be fine silver, treated in the methods called champlevé and cloisonné.

The brilliancy of the substance enamel depends upon the perfect combination and proportion of its component parts. The intimacy of the combination depends upon maintaining an equal temperature throughout its fusion in the crucible. For this purpose it is better to obtain a flux which has been already fused and most carefully prepared and afterward to add the colouring oxides, which stain it dark or light according to the amount of oxide



CHISABUROH F. YAMADA

CLOISONNÉ PROCESS OF DECORATING WITH ENAMEL

(Left) Artisan solders narrow bands of metal (cloisons) edgewise onto the surface of china to outline the design; (right) a paste of enamel powder, water and paste is pressed tightly into the cells formed by the cloisons to colour the design

introduced. Many of the enamels are changed in colour by the difference of the proportion of the parts composing the flux rather than by the change of the oxides. For instance, turquoise blue is obtained from the black oxide of copper by using a comparatively large proportion of carbonate of soda, and a yellow green from the same oxide by increasing the proportionate amount of the red lead. All transparent enamels are made opaque by the addition of calx, which is a mixture of tin and lead calcined. White enamel is made by the addition of stannic and arsenious acids to the flux. The amount of acid regulates the density or opacity of the enamel.

After the enamel has been procured in the lump, the next stage in the process, common to all methods of enameling, is to pulverize it. To do this properly the enamel must first be placed in an agate mortar and covered with water; next, with a wooden mallet a number of sharp blows must be given to a pestle held vertically over the enamel, to break it; then, with the mortar held firmly in the left hand, the pestle must be rotated with the right, with as much pressure as possible on the enamel, grinding it until the particles are reduced to a fine grain. The powder is then subjected to a series of washings in distilled water, until all the floury particles are removed. After this the metal is cleaned by immersion in acid and water. For copper, nitric acid is used; for silver, sulfuric, and for gold, hydrochloric acid. All trace of acid is then removed, first by scratching with a brush and water, and finally by drying in warm oak sawdust. After this the pulverized enamel is carefully and evenly spread over those parts of the metal designed to receive it, in thickness sufficient to cover them and no more. The piece is then dried in front of the furnace, and when dry is placed gently on a fire-clay or iron plank and introduced carefully into the muffle of the furnace, which is heated to a bright pale red. It is now attentively watched until the enamel shines all over, when it is withdrawn from the furnace. The firing of enamel, unlike that of glass or pottery, takes only a few minutes, and in nearly all processes no annealing is required.

The following are the different modes of enameling: (1) champlevé; (2) cloisonné; (3) *basse-taille;* (4) *plique-à-jour;* (5) painted enamel; (6) incrusted; and (7) miniature-painted. These processes were known at successive periods of ancient art. Today they are known in their entirety. Each has been largely developed and improved. No new method has been discovered, although variations have been introduced into all.

1. Champlevé enameling is done by cutting away troughs or cells in the plate, leaving a metal line raised between them, which forms the outline of the design. In these cells the pulverized enamel is laid and then fused; afterward it is filed with a corundum file, then smoothed with a pumice stone and polished by means of crocus powder and rouge or charcoal.

2. In cloisonné enamel, upon a metal plate or shape, thin metal strips are bent to the outline of the pattern, then fixed by silver solder or by the enamel itself. These strips form a raised outline, giving cells as in the case of champlevé. The rest of the process is identical with that of champlevé enameling.

3. The *basse-taille,* or translucent, process is also a combination of metal work in the form of engraving, carving and enameling. The metal, either silver or gold, is engraved with a design, and then carved into a bas-relief (below the general surface of the metal like an Egyptian bas-relief) so that when the enamel is fused it is level with the uncarved parts of the metals, and the design shows through the transparent enamel.

4. *Plique-à-jour* enameling is done in the same way as cloisonné enameling, except that the wires or strips of metal which enclose the enamel are not soldered to the metal base but are soldered to each other only. Then these are simply placed upon a sheet of platinum, copper, silver, gold or hard brass, which, after the enamel is fused and sufficiently annealed and cooled, is easily removed.

5. Painted enamels are different from all these processes both in method and in result. The metal in this case is either copper, silver or gold, but usually copper. It is cut with shears into a plate of the size required, and slightly domed with a burnisher or hammer, after which it is cleaned by acid and water. Then the enamel is laid equally over the whole surface both back and front, and afterward "fired." The first coat of enamel being fixed, the design is carried out, first by laying it in white enamel or any other which is opaque and most advantageous for subsequent coloration.

In the case of a grisaille-painted enamel the white is mixed with water or turpentine, or spike oil of lavender, or essential oil of petroleum (according to the taste of the artist) and the white is painted thickly in the light parts and thinly in the gray ones over a dark ground, whereby a slight sense of relief is obtained and a great degree of light and shade.

In coloured painted enamels the white is coloured by transparent enamels spread over the grisaille treatment, parts of which when fired are heightened by touches of gold, usually painted in lines. Other parts can be made more brilliant by the use of foil, over which the transparent enamels are placed and then fired.

6. Incrusted enameling involves the application of enamel to the surface of jewelry or other small objects, either in the round or in deep relief.

Miniature-enamel painting is not true enameling, for after the white enamel is fired upon the gold plate, the colours used are not vitreous compounds—not enamels in fact—as is the case in any other form of metal enameling; they are either raw oxides or other forms of metal, with a little flux added, not combined. These colours are painted on the white enamel, and afterward made to adhere to the surface by partially fusing the enamel, which when in a state of partial fusion becomes viscous. (A. Fr.; X.)

ORIGINS AND EUROPEAN DEVELOPMENT

Though a vitreous alkaline glaze was applied to earthenware and stone by the Egyptians from the 11th dynasty, there is no evidence to indicate that true enameling on metal was practised in Egypt until after the time of Alexander the Great. Gold ornaments have been found with inlaid coloured stones in Egyptian tombs, but in the days of ancient Egypt coloured enamels do not seem to have taken the place of the cut stones. The explanation may be that the Egyptian glass was a soda-lime glass which fused only at a very high temperature and so could not be used on metal because the latter would melt first. It would seem that to apply enamel to gold, silver or copper it was necessary to have a glass with a high percentage of lead.

No evidence has come to light to indicate that true enameling was practised in the Assyrian kingdom or by the Phoenicians, by whom it is possible in theory that the craft was introduced into Greece. Certain jewels of the Mycenaean period are ornamented with a vitreous blue substance, of which a sceptre in Cyprus dating from about 1000 B.C. is a most important example. Six gold rings, discovered in a Mycenaean tomb of the 13th century B.C. at Kouklia, Cyprus, are decorated with true cloisonné enameling. In these examples, the earliest yet discovered, the different coloured enamels were not in the form of powder when they were fired and

fused together, but were shaped, opaque glass fragments, which suggests an intermediary stage between inlay and true enameling. The Aegina treasure of about the 11th century B.C., now in the British museum, London, has been analyzed and the blue substance proved to be lapis, and so the results of the analysis of the sceptre and other Mycenaean ornament of this kind will be of the utmost significance. If it is established that true enameling existed in Mycenaean work, then it would be reasonable to expect the art to have been inherited by the Greeks and transmitted by them to the rest of Europe, perhaps by way of the colonies on the north shore of the Black sea and in the south of Italy.

The most ancient examples of enameling outside Mycenaean art are on the bronze ornaments found in a cemetery at Koban in the Caucasus, which are claimed to date from the 8th to the 9th century B.C.—between two and three centuries earlier than the Greek colonies north of the Euxine. Again the results of an analysis would be valuable, but certainly Greek and Etruscan jewelry belonging to the 6th to 3rd century B.C. has been found, which, though not analyzed, is claimed to be enameled—admittedly sparingly and on very small areas. Celtic enamelers of central and western Europe from the 3rd century B.C. substituted for coral, which was becoming scarce, a red enamel to ornament their bronze objects. Northern French examples of the La Tène period and British finds such as the Witham shield (British museum) are among the finest surviving examples of these undoubted enamels.

During the Roman period, it is significant that enameling, both cloisonné and champlevé, was carried on almost entirely in those old Celtic areas which had become the northern provinces of the Roman empire, and in north Africa. It may well be to these provincial works that the often-quoted passage in Philostratus (*Icones*, i, 28), written in the time of Septimius Severus (A.D. 146–211), refers. The author, describing a boar hunt at which the riders appear with horse trappings ornamented in bright colours, writes: "It is said that the barbarians in the ocean pour these colours into bronze moulds, that the colours become as hard as stone, preserving the designs." This is a fair description of the process of champlevé enamel (see *Enameling Processes* above). Although he does not state that these people were the only exponents of the craft known to him, it does suggest that they were exceptionally skillful and indicates that the technique was not practised in Greece or Italy at that time. The phrase describing these people is singularly applicable to the inhabitants of the British Isles, where just these objects—enameled horse trappings—have been found in many places, such as Polden hill, Somerset, which date from the 1st century A.D. Examples are to be seen in the British museum.

After the collapse of the Roman empire in the west, there is no evidence of the continued practice of enameling until the 5th or 6th century, when Anglo-Saxon bronze ornaments are found with champlevé enamels. Among the finest examples of this period are the small enameled panels or escutcheons found on the hanging bowls in the Sutton Hoo ship burial now in the British museum. On the continent throughout the earlier centuries of the Christian era the popular inlaid jewelry seems to take the place of enamel decorations, though the survival of a few 8th-century examples of western European origin, such as the cloisonné-enameled Ender Purse reliquary, now in Berlin, has been argued as evidence of the continuous practice of the craft since Roman times. The extant examples are neither firmly dated nor documented, and there is always the possibility that the enameled parts are later additions. However, more recent writers have claimed that they are due to the influence of the Sassanian Persians, who most probably produced cloisonné enamels, though the evidence is not conclusive. The remarkable cloisonné-enameled ewer in the Treasury of St. Maurice d'Agaune, Switz., is decorated with subjects of purely Persian character. The enameling has been claimed to be late Sassanian work, but the possibility that it is Islamic work of the 11th century cannot be ruled out.

The great period of Byzantine cloisonné enameling commences in the 10th century in Constantinople, continuing throughout the 11th century, during the first half of which the best work was done. The masterpiece of this period is the altar screen known as the Pala d'Oro in St. Mark's, Venice, believed to have been brought from Constantinople to Venice about 1105, though outstanding pieces exist in the Reiche Kapelle, Munich, and the Marcian library, Venice. Cloisonné enameling first appears in Russia with Christianity in the 10th and 11th centuries, during which time it also was practised in the Rhineland, in Sicily and in the Islamic area, as is magnificently exemplified by the Rhenish cross of the emperor Otto and the abbess Mathilde in Essen, Ger. (973–982), and by the enameled copper basin at Innsbruck, Aus., bearing an Arabic inscription with the name of an Urtukide prince reigning around 1144.

The triumph of the champlevé process of enameling in the second half of the 12th century ousted the cloisonné technique. It is a matter of dispute, through lack of sufficient dated evidence, whether the revival of this technique took place first in the valleys of the Rhine and the Meuse at Cologne, Aachen, Trier, Liège and Hildersheim, or in the great southern centre of Limoges in France. W. L. Hildburgh has claimed a Spanish origin for the earliest champlevé enamels of the Romanesque period, the outstanding example of which is the altar front of Silos (Burgos museum, Spain). The Mosan and Rhenish enamels differ in colour scheme from those of Limoges: green and yellow are dominant and the blues are paler and almost icy compared with the richness of the Limoges school. One of the earliest datable pieces is a curious enamel in the British museum representing Henry of Blois, brother of King Stephen of England and bishop of Winchester, which must have been executed between 1139 and 1146, perhaps in England, as it is unlike the Mosan or Rhenish work and even less like the Limoges enamels. Almost as problematical is the plaque with the enameled figure of Geoffrey Plantagenet, count of Maine, who died in 1151, which is generally regarded as a work of Limoges, though the colour scheme is not yet typical of the later Limoges school. In the Mosan school the famous enamelers Godefroid de Claire (fl. 1150–69), working in Liège, and Nicholas of Verdun (fl. 1181–1205) brought variety of form and design to their works, which are of a very high artistic standard. The more than 50 panels of the pulpit at Klosterneuburg near Vienna by Nicholas of Verdun are without parallel in this field of art. The best work from Limoges was executed in the 12th and early 13th centuries, but thereafter the output was commercialized and the standards fell throughout the late 13th and 14th centuries.

It was in 13th-century Italy that the style of enameling which was to supersede champlevé in popular taste was first evolved. In France it is known as *émail de basse-taille* and in England is called translucent enamel on sunk relief. Vasari says that the famous Giovanni Pisano made a retable embellished with enamels on silver for Arezzo cathedral, and the chalice made for the convent of San Francesco d'Assisi in 1290 is signed by the great Duccio of Siena. Throughout the 14th century splendid reliquaries, chalices, processional crosses, etc., were made in Italy, especially in Tuscany, many of which survive *in situ*, richly decorated in this manner.

The technique quickly spread to Spain, where the enameled objects, particularly processional crosses, of Barcelona, Tarragona, Gerona and Valencia form an impressive array. During the second quarter of the 14th century it was well established in France, as illustrated by the enameled plinth of the silver figure of the Virgin given by Jeanne d'Evreux to St. Denis in 1339, and now in the Louvre, Paris. The most splendid French production in this technique is the royal gold cup of the kings of France and England, now in the British museum, made to the order of John, duke of Berry, as a gift to his brother, Charles V of France, in 1381. The sides and the cover have scenes depicting the life and martyrdom of St. Agnes in the most glowing rich colours and elegant draftsmanship of the period. The stem has a band of white enameled roses added when it belonged to the Tudor kings of England, before it was given by James I to Philip III of Spain's representative in 1604. Though there is documentary evidence that *basse-taille* enameling was made in England in the 14th century, there are no indisputably English works extant, though several small examples are said to be English on grounds of style. The silver cup at King's Lynn, Norfolk, of the second quarter of the 14th century, is claimed to be English, but the attribution is con-

jectural. The great era of *basse-taille* enameling ended with the Renaissance, though it remained popular in Spain and southern Germany, chiefly in Augsburg, to the middle of the 16th century.

Enameling in an openwork frame, producing an effect of a stained glass window in miniature, is known as *plique-à-jour*. Because of its extreme fragility and the difficulties of production, few examples survive. One of the loveliest pieces is in the Victoria and Albert museum, London, a silver-gilt Merode beaker of Flemish or Burgundian origin, probably early 15th-century, decorated with two bands of enamels set in tiny windows with Gothic tracery. It was a technique popular in Transylvania in the 16th and 17th centuries, examples of which were in the National museum, Budapest, Hung.

The process of enameling in the round, often called incrusted enameling, in which the enamels are applied to the convex surface of the object, is thought to have been used by the Greeks to decorate diminutive jewelry. It was revived on a large scale in France, probably in Paris, during the second half of the 14th century. Among the finest and earliest surviving examples is the Reliquary of the Holy Thorn in the Waddesdon bequest in the British museum. Bearing the arms of John, duke of Berry, the prince who commissioned the royal gold cup, it appears to have been given by him to his brother Charles, duke of Anjou, between 1389 and 1407. The Holy Thorn set in a gem is surrounded by the Last Judgment scene, in which all the figures, 20, are enameled, many of them being executed wholly in the round. The *Goldene Rössel* of Altötting, a devotional group in the same manner, was made for Charles VI of France, but few other pieces escaped the general destruction of these creations in France.

During the 16th and 17th centuries hat jewels and jeweled pendants were frequently ornamented with enameled figures, often illustrating classical scenes, rendered in high relief or completely in the round.

The application of enamel colours to glass vessels appears in Venice in the 15th century and the enamelers of northern Italy, particularly in Venice, and of Limoges in central France seem to have adapted this technique to their craft by mid-century, creating a new process known as painted enameling. The term is misleading because the result is not achieved by the brush but with a spatula or point. It is significant that the style and the system of production in north Italy remained quite distinct from that in France; in the latter country, production was rapidly commercialized on an industrial scale with a large market at home and abroad, while in Italy it was short-lived, coming to an end in the early 16th century. The Italian enamelers borrowed their subjects from painters of the North Italian schools, while in Limoges for the first 80 years the sources of most of the subjects were Flemish and German prints. The earliest Limoges enameler known is "Monvaerni" (the spelling of the name is doubtful), who, flourishing in the second half of the 15th century, produced a powerful style in thickly applied enamel outlined in black, recalling painting on glass. His best work is in the Louvre. Better known is Nardon (Léonard) Pénicaud (1474–*c.* 1539), whose rich translucent colours are applied to a white ground, freely enhanced by "jeweled" borders (little disks of foil beneath dots of glaze). His style had many imitators. During the period 1530–80, the great enamelers flourished: Pierre Reymond, K.I.P. (Jean Poillevé), the three Jean Pénicauds, Léonard Limousin (*q.v.*) and Couly Nouailher. It was a time when they turned to Italy for their ideas, and the engravings of Marcantonio after Raphael, the Master of the Die, etc., were slavishly copied, though in the hands of Jean Pénicaud III an individual interpretation was created. Some of the latter's designs are even his own creations, and his popular technique of painting in grisaille on a black or dark blue enamel group was rivaled by Léonard Limousin and Pierre Reymond. By the last quarter of the 16th century the quality had degenerated and the enamelers Jean and Suzanne de Court in particular turned from the soft harmonies of the earlier artists to the use of bright colours enhanced by an excess of foil for gaudy rich effects.

The 17th century remained a period of decadence for the Limoges painted enamel craft despite the Nouailher and Laudin families. However, in Paris the Toutin family introduced a style of exquisite miniature enamel painting which continued to have a great vogue under Louis XIV. Enameling of watchcases in the 17th century led to the enameling of snuff boxes, boxes, étuis, etc., in the 18th century. In England, particularly at Battersea (now in London) in 1753–54, the Parisian quality of enameled miniature painting was copied and even rivaled. The technique of transfer painting on enamel and adding colours was practised to perfection at the Battersea works, while the portrait miniatures of Henry Bone (*q.v.*) and William H. Craft carry the tradition into the early 19th century. (H. Tr.)

FAR EAST

China.—Enamels do not appear to have reached China until long after they were to be found throughout Europe. The Chinese terms for the enameled "Fu-lin ware" or "Fo-lang inlaid work" (hence another name, "Fa-lan") suggest an origin in the eastern Roman provinces, those names having been applied to the Byzantine empire by Chinese writers as early as the 7th century. The derivation of the words has been the subject of controversy; but all authorities are agreed as to the western origin of the art, which in all probability was introduced into China by traders or by traveling craftsmen. The art of glassmaking seems to have been imported into China from the west as early as the 3rd century B.C., as indicated by excavated glassware fragments and glass beads. Some excavated bronze wares of this period are decorated with inlaid glass beads or plaques. The Chinese in the 3rd century B.C. even applied an enamellike technique in decorating pottery ware with coloured glass paste. However, the first historical record of glass manufacture in China describes the coming of craftsmen from an Indo-Scythian kingdom on the northwestern border of India to the capital of northern Wei in the reign of T'ai Wu (424–452); these craftsmen succeeded in making excellent opaque glass of various colours from local minerals.

Although the Chinese were thus informed as to the production of an essential material for the making of enamels, and were already highly skilled in the working of bronzes and other metals, there is no evidence that the art of enameling was practised before the T'ang dynasty period (618–906). There is in the Shoso-in ("principal storehouse") at Nara in Japan a silver mirror, the back of which is decorated in cloisonné. This piece was deposited there either in the 8th century or in 950. It is generally agreed that the mirror is of Chinese origin, dating from the T'ang dynasty period, as is certainly the case with many other objects in the collection. At present, this is the only known Chinese enameled ware made before the 14th century. But it can be safely assumed from this piece that the art of cloisonné was developed to a respectable height in the T'ang dynasty China. This supposition can be also supported by the fact that the 8th-century Japanese imperial court, following Chinese example, seems to have maintained enamel artists, discussed below. It is remarkable, however, that there is no written Chinese record of the art until 1387, when the *Ko-ku-yao-lun*, a book on antiquities, was published. It records the production on a large and varied scale of enameled ware, called at the time *Ta-shih-yao* ("Arabian" fired ware), which "resembles the cloisonné work of Fo-lang." It also states that the natives of Yünnan had established factories for this ware in Peking and that the enamels made at Yünnan-fu are "fine, lustrous and beautifully finished." It appears therefore that cloisonné work was well established in China at this period, and that Byzantine work of similar character was also so well known as to invite comparison with the native product. The former may well have served as an example for Chinese craftsmen. M. G. Paleologue points out in *L'Art chinois:* "The workmanship presents occasionally, in fact, striking resemblances with certain enamels of the Byzantine school; the mixture of different enamels inside the wall of the same cell, the employment of gold incrustations in the treatment of the figures and the hands, etc." Active trade and cultural intercourse between the near east and China during the Mongolian dynasty must have been the reason for this revival of enamel work, which then flourished through the Ming and Ch'ing dynasties (1368–1644 and 1644–1912, respectively).

Chinese enamels fall into three categories—cloisonné, champlevé

and painted. In none does the technique vary appreciably from that employed in western countries.

Cloisonné.—The earliest example of cloisonné enamel that can be authentically associated with the far east is the silver mirror in the Shoso-in mentioned above. Its cloisonné back is decorated in a design of a six-petaled blossom in three layers, the tips of the outer rows of petals forming 12 points of the mirror. The piece is regarded as a T'ang dynasty Chinese work. Apart from this the sequence of Chinese enamels with which we are acquainted begins only in the Yüan period, so far as is known at present, and the earliest recorded marks hitherto known belong to the reign of the last emperor of that dynasty (1333–68). The great period of the production is certainly that of the Ming dynasty which followed.

The mark most commonly found within this period is that of the Ching-t'ai reign (1449–57); S. W. Bushell suggests that there must have been a great revival of the art at this time, as even in the late 19th century the term *Ching-t'ai-lan* (Ching-t'ai deep blue ware) was "commonly used in Peking as a general synonym for cloisonné enamels." The Ming enamels, bold in design with fine depth and purity of colour, were never surpassed in later epochs. The two shades of blue, a dark lapis-lazuli tone and a pale sky-blue with a very slight tinge of green, are particularly excellent. The red is of dark coral tint and the yellow full-bodied and pure. Greens derived from copper are sparingly used, and Bushell states that *rouges d'or* (reds made of gold) do not come into the colour scheme at all. The black and white are the least successful, the former shallow and dull, the latter clouded and muddy. At the same time, an imperfection of technique is noted, close examination revealing minute pitting in the enamels, caused by inadequate packing of the material, and some want of polish in the surface. These technical defects, however, do not appreciably detract from the artistic value of the Ming enamels and, indeed, serve as a clue to their identification.

To the patronage of the emperor K'ang-hsi (1661–1722) was due a great revival of art industries. In 1680 he established a whole series of imperial factories for this purpose, of which that devoted to the manufacture of enamels was number six on the official list. Here he had made sets of incense vessels of cloisonné enamel for presentation to the numerous Buddhist temples founded under his auspices in the neighbourhood of Peking as well as other objects for the honorific gifts that were characteristic of his enlightened reign. The enamels of his time are marked by an improvement in technical quality as compared with those of the Ming period, while the finer qualities of the latter are still, to a considerable extent, in evidence. In many cases the forms of ancient bronze vessels were revived for these purposes, with the addition of enrichments in enamel. The style of this reign persisted during that of K'ang-hsi's successor, Yung-cheng (1723–35), while the long period on the throne of Ch'ien-lung (1735–96) was marked, in enamel as in the case of many other industrial arts, by a further perfection of technique but by a loss of much of the vigour of design and breadth of execution that distinguished the products of earlier periods. Modern enamels, chiefly imitations of older work, are more hurriedly made and not so well finished.

Champlevé.—Some of the most ancient enamel examples extant belong to this class, and examples in which both champlevé and cloisonné are employed are not uncommon.

Translucent enamels (opposed to the opaque enamels generally used in China) were occasionally made by Chinese artisans. The cloisonné back of the silver mirror in the Shoso-in is decorated with transparent enamels. Otherwise important pieces are rare, but sometimes of fine quality, silver or gold base being used. This method more often appears in Chinese jewelry.

Painted Enamels.—The painted enamels of China, generally known, from the principal seat of their manufacture, as Canton enamels, are practically identical in technique with the Limoges and other painted enamels of Europe. Specimens of the latter are known to have been taken to China by the missionaries of the late 17th and 18th centuries; they not only exercised direct influence on the Chinese ware but also, in some cases, were copied. Representations of European subjects, copies of engravings and armorial decorations, are also found. Painted enamels are termed by the Chinese *Yang-t'zu* ("foreign porcelain"), the palette of colours used being the same as with enameled porcelain, whose decoration under foreign influence is called *Yang-ts'ai* ("foreign colours"). A ground of opaque enamel, generally white, is laid on the copper, and on this the colours are superimposed and fired. The best period of this art was the 18th century. Although imitations have continued to be made, nothing of real quality in this style was produced after the termination of the reign of Ch'ien-lung in 1796. The method has always been looked upon by the Chinese as alien in taste; in fact a great part of Canton enamels were made for export, not only on European commission but also for clients in India, Persia and other Asian countries.

Japan.—The example of enamel in the Shoso-in at Nara has been referred to, and it may now be repeated that it cannot be attributed to Japanese craftsmanship. However, the art of enameling in Japan may date back to the 7th century. One enameled metal piece, discovered in a tomb near Nara, which probably dates from the 7th century, seems to be of Japanese origin. The Civic Code of Taiho, compiled in the 8th century, provides one official to be in charge of founding metals and "painted glass decoration." In later periods, however, this art seems almost to have died out until the 17th century. In the 15th century, wares decorated with *Shippo-ruri* or *Shippo*, as enamel decoration is called in Japan, are mentioned in several records, but they seem to have been imported from Ming dynasty China. When Donin Hirata I (1591–1646) succeeded in making enameled wares, having learned the technique from Koreans, his art was highly appreciated by Ieyasu Tokugawa, then the shogun of Japan, under whose patronage Hirata worked in Kyoto. There are some sword furnitures and a suit of armour with enameled metal fittings ascribed to him. Enameled metal fittings decorating sliding doors and lintels in the Katsura palace, Kyoto, are also said to have been made by him. His family continued the trade until the late 19th century. They made use, on a small scale, of both the cloisonné and champlevé methods. There was no further development of importance until Tsunekichi Kaji (1803–83) of Nagoya and his pupils established in that city an important and successful manufacture of cloisonné which obtained a considerable vogue, especially among foreigners. On this basis further developments of considerable interest have taken place.

While Kaji used brass *cloisons* and opaque enamel colours only, his successors used silver *cloisons* and succeeded in making both transparent and translucent enamels. They employed these different kinds of enamels cleverly, according to the effects desired. They further modified the cloisonné process with remarkable ingenuity and produced work of great interest. First, they reproduced in minutely detailed cloisonné work realistic pictures of trees, flowers, etc. This effort led them then in the 1880s to produce lineless cloisonné enamels, which have all the beauty and brilliance of cloisonné enamels, with thick layers of enamel colours, but which, showing no trace of *cloisons*, permit a gradation of colours as in the less clear and brilliant painted enamels. This was achieved by taking off *cloisons* before each firing, the process being repeated at least three times. The artists working for the factory of Sosuke Namikawa of Tokyo in the late 19th century were most successful in this technique. Another Namikawa of Kyoto worked in true cloisonné. The factory of Jubei Ando of Nagoya has produced more variations. These developments have carried the art of enamel very far from the old traditions; while the skill and ingenuity of technique they evince may be appreciated, there is a danger of deviating from the artistic beauty proper to this art.

See also references under "Enamel" in the Index volume.

(E. F. S.; C. F. Y.)

BIBLIOGRAPHY.—*Europe:* L. de Laborde, *Notice des émaux . . . du Louvre*, 2 pt. (1852–53); J. Labarte, *Recherches sur la peinture en émail dans l'antiquité et au moyen âge* (1856); E. Molinier, *Dictionnaire des émailleurs . . .* (1885); O. von Falke and H. Frauberger, *Deutsche Schmelzarbeiten des Mittelalters* (1904); J. J. Marquet de Vasselot, *Les Émaux limousins de la fin du XVe siècle . . . Étude sur Nardon*

Pénicaud et ses contemporains, 2 vol. (1921), *Les Crosses limousines du XIIIe siècle* (1941), *Les Gémellions limousins du XIIIe siècle* (1952); M. Chamot, *English Mediaeval Enamels* (1930); W. L. Hildburgh, *Medieval Spanish Enamels . . .* (1936); M-M. Gauthier, *Émaux limousins: champlevés des XIIe, XIIIe, XIVe siècles* (1950); S. Collon-Gevaert, *Histoire des arts du métal en Belgique*, 2 vol. (1951); P. Thoby, *Les Croix limousines de la fin du XIIe siècle au début du XIVe siècle* (1953); H. Maryon, *Metalwork and Enamelling . . .*, 3rd ed. rev. (1954); Floridus Röhrig, *Der Verduner Altar* (1955).

Far East: F. Brinkley, *Japan and China*, vol. vii (1904); S. W. Bushell, *Chinese Art*, vol. ii (1906); R. L. Hobson, "Chinese Cloisonné Enamels," *Burlington Magazine*, xxi:137 and 202, and "Canton Enamels," *Burlington Magazine*, xxii:165.

ENAMELED IRONWARE: *see* PORCELAIN ENAMELING.

ENAMELWARE: *see* PORCELAIN ENAMELING.

ENCARNACIÓN, capital of Itapúa department, Parag., on the Alto Paraná river. Pop. (1960 est.) 18,783. Founded by the Jesuits in 1632, its present importance is based on its function as a transportation centre, as it is the only junction enabling rail communication between Paraguay and the rest of South America. A train ferry across the Paraná, from Posadas, Arg., links the Paraguay Central railway with the Argentine system. Encarnación is also the terminus of the main highway from Asunción and the principal commercial centre for southern Paraguay, exporting hides, maté, timber and cotton, and having a number of food-processing industries, sawmills and tanneries. Partially destroyed by a tornado on Sept. 20, 1926, it has since been rebuilt into a modern city. (G. J. B.)

ENCAUSTIC PAINTING. Encaustic or hot-wax painting is one of man's oldest easel painting methods. Perhaps originating in Egypt, it was practised in the earliest periods of Greek civilization and continued through the various schools of antiquity. Some time prior to the middle ages it became a lost art as other ways of painting developed. The encaustic process is extremely simple in principle, although like any other technique it demands that the artist acquire skill in its manipulation and control of its special equipment. Dry-powder pigments, identical with those used in oil colours (*see* OIL PAINTING, TECHNIQUE OF), are mixed with molten beeswax together with variable amounts of resin; these wax colours are applied to the painting surface from a hot palette which keeps them within the proper temperature range. After the painting is completed, it is subjected to a final burning-in treatment by passing a heat source slowly across its surface. The temperature is controlled so that it welds the paint internally into a solid or single mass without running or fusion of the surface. Encaustic derives its name from this final part of the process.

History.—No Greek easel or mural paintings of the classical period survive, but there is extant a group of funerary portraits on wood panels known as the Faiyum paintings, done in Egypt by Greek artists in the 1st century A.D. They are in fine condition, realistic and brilliant. One of the principal guides to scholars in the study of Greek encaustic painting is the *Natural History* of Pliny the Elder, whose remarks on studio procedure and technical details, however, are brief and clouded by vagueness and ambiguity —perhaps because he lacked familiarity with them or thought them familiar to everyone. In the studies of all ancient technical processes many valuable clues also come from nontechnical writings. An example is a line attributed to the poet Anacreon (*c.* 560 B.C.), who, marveling at a likeness, says: "Quickly O wax, thou wilt speak." These words suggest that the use of pigmented wax was of considerable antiquity, that it was a commonplace artists' paint and that it was sufficiently controllable to allow impressive results. The Greeks called their colours and paintings waxes just as tube colours and canvases are now called oils. Wax colours in cruder form were used on ships. They were also used to a considerable extent on statues, by applying the molten colours and polishing them with cloths; many Greek and Roman sculptures that now exist as clean marble were originally polychromed in naturalistic colours. An interesting addition to the relics of encaustic painting was found in 1847 at St. Médard-des-Près in southern France when a tomb was discovered which contained the remains of a woman painter surrounded by her *cauteria* (the instruments used in encaustic painting). The ancients used a barrel-shaped container of glowing charcoal on which rested a bronze or silver palette which kept the waxes at the proper working temperature. They applied the colours to surfaces with brushes and a bronze spatulate implement called a cestrum. The burning-in or "intusion" was done with a portable charcoal brazier.

The Romans used wax as a polish and preservative coating freely over many surfaces, including frescoes, and because it was found on walls in Pompeii some early investigators believed that those murals were done in encaustic. Encaustic was used by the Romans for solid colour or simple-design decorations on walls, but it is doubtful that it displaced fresco for mural painting to any great extent for work of intricate or pictorial design.

During the 18th century, encaustic painting was studied in several countries in the hope of developing a mural technique that would withstand dampness and other unfavourable conditions. Because Pliny's obscure statements implied the use of a second and possibly a third Greek method in which the fluid colours could be applied cold, many combinations of wax with oils or varnishes, and saponified or emulsified wax were investigated. Differences of opinion continued through the 19th century. Pliny's "Punic wax" is now generally agreed to have been equivalent to modern white refined beeswax. The other painting method is thought to have been unrelated to encaustic and was most likely a simple water medium.

Practically all of the 18th- and 19th-century cold-wax recipes have been discarded, including the prominent Gambier-Perry Spirit fresco, because none of them withstood moisture or moisture seepage from the rear any better than oils. The potential decay elements of oil painting proved to be present in any mixture with oil, and no oil mixture could equal the resistance of wax or wax-resin melts. In addition, the cumbersome charcoal heating arrangements of the 18th and early 19th centuries were little improved over the *cauteria* of 2,000 years ago, which precluded the widespread revival of encaustic. Any mixtures that could have been devised in ancient times would not have been very different from those available in more recent periods. The ancients developed a superb manipulative skill in their creative works, and some of their technical processes called for a great amount of patient, laborious operations, but the keynote of their materials was simplicity. The 18th- and 19th-century wax recipes, with their complicated chemical cookery techniques, were more typical of their own time than of the earlier civilizations from which they were purportedly derived.

Another revival of encaustic painting, of more widespread acceptance than that of the 19th century, arose before mid-20th century. Numerous artists found that its manipulative properties, its thermoplastic flow and its textural effects made encaustic highly suitable to meet the demands of several styles of contemporary painting that were not well-served by any of the other standard, acceptable methods in current use.

Technique.—The artist of modern times has little interest in the exact identification of some liquid binder that may have been used 2,500 years ago on pictures which no longer exist. This is the concern of antiquarians. However, the classic hot-wax method, which is well documented, does interest the artist because its unique properties make it a useful addition to his list of approved painting methods.

Electric heating equipment has now made the method relatively convenient, and artists have displayed much ingenuity in adapting it to hot palettes, using heat elements and light bulbs. Manufactured electric palettes also are available. A lightweight heat reflector that can be clamped to the easel or held in the hand so that the canvas or panel can be kept warm allows the painter to manipulate the colours as freely as oils. It can be switched off or withdrawn at will to make the brush strokes congeal more rapidly. The best working temperature range for average use will be found to be 200–220° F. For the final burning-in, the painting is laid face up on a table; a reflector with an infrared lamp or a heat element is passed over it at a uniform distance.

Encaustic is a very clean technique. It is seldom necessary to use solvents; warming the implements and brushes and wiping them with a cloth will usually clean them at once. The brushes

need not be spotless. When a cleaning solvent is necessary, mineral spirit will usually suffice. The finished picture remains clean with relatively little care and no varnishing is required; the wax surface seems to repel dust and grime rather than attract it as oil and varnish surfaces do.

The molten wax colours can be applied with bristle brushes or with warmed palette knives. Any surface or ground suitable for oil painting is suitable for encaustic. Prepared wax-resin colour sticks are available, although it is a simple matter to make the colours by mixing dry-powder pigments into a molten mixture of white refined beeswax and damar resin.

When they are cold, encaustic paintings can be polished to a pleasing dull sheen by rubbing with absorbent cotton. A wide range of brilliant painting effects is possible: transparent, opaque, thin, thick. Many textures can be obtained, from smooth to very rough and bold, the latter without much actual thickness of paint. Thus the method is attractive to painters who desire these effects but do not want to risk either the doubtful longevity of the synthetic-resin enamels or the possible failure of oil colours and other traditional materials used in ways and in thicknesses for which they were never intended.

Not all 20th-century encaustic paintings designated as such are the classic hot-wax variety. Cold-painting wax mixtures with oils, and water dispersions of waxes, have been extensively developed, not only as reconstructions of the ancient procedures or the 18th-century experiments but also as useful inventions to serve the contemporary painter. Although some of these may have considerable merit, most of them are to be classified as wax-modified oil colours rather than encaustic. All true encaustic contains beeswax as its major ingredient and receives a final burning-in treatment.

BIBLIOGRAPHY.—F. Hamilton Jackson, *Mural Painting* (1904); A. P. Laurie, *Greek and Roman Methods of Painting* (1910), *Materials of the Painter's Craft* (1910); Ralph Mayer, *The Artists' Handbook of Materials and Techniques*, rev. ed. (1957). (RH. M.)

ENCEPHALARTOS, a genus of cycads, low, palmlike plants, native to Africa, a few species of which are grown in conservatories and in houses for their ornamental foliage. The habit of most species is an extremely slow-growing, unbranched stem with a crown of large, pinnate leaves. *E. caffer*, Kaffir bread (*q.v.*) or bread palm, is a source of food for some South African tribes. A few species are tuberous.

See GYMNOSPERMS: *Cycadales.*

ENCEPHALITIS, inflammation of the brain (pl. encephalitides; from Gr. *enkephalos*, "brain"; *itis*, "inflammation"). The brain may be involved either initially or secondarily from inflammation of another organ of the body. The inflammation which affects the brain may affect adjoining structures as well. In encephalomyelitis there is inflammation of the brain and also of the spinal cord; in poliomyelitis the inflammation is in the gray matter of the spinal cord; meningoencephalitis is an inflammation not only of the brain substance but also of the meninges (the membranes covering the brain and spinal cord).

Encephalitis is commonly thought of as "sleeping sickness," but this term is more properly used to designate only one type of encephalitis, encephalitis lethargica, which occurred in epidemics in many parts of Europe and the United States about the time of World War I. Following this disease, many persons had seizures, peculiar gait, inexpressive faces and behaviour disturbances, and a few were mentally deteriorated. Encephalitis lethargica is now a disease of primarily historical interest, since epidemics of this type of encephalitis are thought to have disappeared about 1926. Although the influenza virus was suspected, the causative agent for encephalitis lethargica never was established.

Etiology and Pathogenesis.—Encephalitis may be caused by a large number of living and nonliving agents which enter the body by various routes and cause different kinds of damage to the brain. Many cases occur, however, for which no cause can be established.

The encephalitides caused by living agents outnumber by far those caused by nonliving substances. Many different organisms are capable of producing encephalitis. Organisms are classified into groups according to particular common properties, one of these being size. Viruses are the smallest of known living organisms, and helminths (worms) are among the largest of the organisms that may invade the human body. Listed below are the major encephalitides caused by organisms:

Virus
- Arthropod-borne, including eastern equine, western equine, Venezuelan equine, St. Louis, Japanese B, Russian spring-summer, Murray valley, louping ill and others.
- Coxsackie
- Dengue
- ECHO (enteric cytopathogenic human orphan)
- Hepatitis
- Herpes simplex
- Herpes zoster
- Psittacosis
- Rabies (hydrophobia)
- Yellow fever

Rickettsia
- Q fever
- Rocky mountain spotted fever
- Tsutsugamushi disease (scrub typhus)
- Typhus fever

Bacteria
- Bacillary dysentery
- Bartonellosis (Oroya fever)

Bacteria—*continued*
- Brucellosis (undulant fever)
- Plague
- Cholera
- Melioidosis
- Pertussis (whooping cough)
- Rheumatic fever
- Scarlet fever
- Tularemia
- Typhoid fever

Spirochete
- Leptospirosis
- Rat-bite fever
- Relapsing fever
- Syphilis (*Treponema pallidum*)

Protozoa
- Malaria
- Toxoplasmosis
- Trypanosomiasis, African
- Trypanosomiasis, South American (Chagas' disease)

Helminth
- Ancylostomiasis (hookworm)
- Ascariasis
- Filariasis
- Trichinosis

Of these causative agents, the viruses are the most important because they cause the largest number of cases. The encephalitis-producing viruses may be divided into two groups: (1) those which invade the body and produce no damage until they are carried by the blood stream to the nerve cells of the brain where they lodge and multiply (*e.g.*, the arthropod-borne, rabies and Coxsackie viruses); and (2) those which invade the body and first injure tissue other than nervous tissue and then secondarily and rarely invade the brain (*e.g.*, the viruses causing herpes simplex, herpes zoster, dengue, yellow fever and psittacosis). The rickettsiae, bacteria, spirochetes, protozoa and helminths usually affect only nonnervous tissue, but occasionally they may invade the brain and cause encephalitis.

Another type of encephalitis, known as perivascular demyelinating encephalitis, is related to viral diseases (measles, chickenpox, smallpox, influenza) or to vaccination for viral diseases (smallpox, rabies) but is classified separately because the damage to the brain is markedly different from that which occurs in other types of viral encephalitis. In this type damage is done not to the nerve cell body but to the insulation (myelin) about the nerve fibre (*see* NERVE). For reasons not understood, only the myelin of nerve fibres which lie near blood vessels, particularly veins, is injured. Perivascular demyelinating encephalitis, which includes postexanthematous and postvaccinal, is probably responsible for a comparatively large number of the acute encephalitides in children.

Multiple sclerosis is the best known of another group of encephalitides in which there is injury to myelin. In this type, injury is scattered and widespread and not limited to nerve fibres lying near blood vessels. The cause of multiple sclerosis is not known.

Many chemicals occasionally produce encephalitis. Some of the more familiar are lead (used in paints and insecticides), arsenic (used for rodent and insect poisoning), mercury (used in industry and in medicines), ethyl alcohol (used in alcoholic beverages), carbon tetrachloride (a household cleaner), DDT (insecticide), morphine and barbiturates (medicines).

Certain nutritional deficiency diseases and metabolic disturbances—such as pernicious anemia, beriberi, pellagra, diabetes and liver disease—also may occasionally cause encephalitis. In addition, many cases of encephalitis occur which cannot be ascribed to any known cause. Among 800 cases of acute encephalitis occurring in the summer of 1952 in California, the causative agents of only one-half were positively identified, despite the fact that all exhibited the same clinical findings and epidemiologic conditions, and similar laboratory investigations were pursued. It is probable that many encephalitides of undetermined etiology are

caused by unknown or undetected viruses.

Mode of Transmission.—Methods to prevent the encephalitides of infectious origin are dependent on a knowledge of the source and reservoir of the infectious agent and the mode of transmission to man. Healthy wild birds and domestic fowl are known to be reservoirs for some of the arthropod-borne viruses (*e.g.*, of eastern equine encephalitis, western equine encephalitis and St. Louis encephalitis), these viruses being transmitted to man by the bite of a mosquito, principally the *Culex tarsalis*. Sick birds are the reservoir for the virus of psittacosis, the virus being transmitted from bird to man through close contact. Rabies is found in many biting mammals, including dogs, foxes, coyotes, skunks and certain insectivorous and fruit-eating bats; human beings acquire rabies through the bite of a rabid animal.

The rickettsiae which cause Rocky mountain spotted fever are passed from generation to generation in ticks, which in turn transmit the infection by biting man. Scrub typhus is similarly passed by mites.

Man is the reservoir for some infectious agents, and transmission is accomplished through (1) mosquito vectors as in malaria, dengue and yellow fever; (2) the air, as in pertussis and scarlet fever; (3) food, water or contaminated articles, as in dysentery, brucellosis and trichinosis.

Chemicals which cause encephalitis may enter the body through the skin or by inhalation or ingestion.

Immunity.—Little is known about immunity and susceptibility factors in most types of encephalitis. It is believed that immunity to a specific type of encephalitis may be acquired through an attack, but this will not assure immunity against encephalitis caused by other organisms or agents.

Clinical Illness and Diagnosis.—Symptoms common to most types of encephalitis are fever, headache, drowsiness, lethargy, coma, tremors, stiff neck and back. Convulsions may occur in patients of any age but are particularly common in infants. Since some of the symptoms are similar to those which occur in other illnesses, the physician looks for certain neurologic signs which are found commonly in encephalitis; in-co-ordination of muscle movements, involuntary movements, weakness of the arms, legs or some other portion of the body, or sensitivity of the skin to various types of stimuli. Cerebrospinal fluid may be drawn from the lower spine and examined for microscopic and chemical changes.

The symptoms, signs and cerebrospinal fluid examination usually establish whether or not the patient has encephalitis, but they do not establish its cause. Possible causes are suggested by epidemiologic data; for example, endemic or epidemic conditions, exposure to communicable diseases or to chemicals, recent immunization, ingestion of contaminated food, presence of sick birds, insect bites, etc.

Special laboratory tests designed to determine etiology are useful, but even these may not give conclusive evidence because tests do not exist for all the possible causative agents, nor are testing methods sufficiently refined to give consistently reliable results. Laboratory tests are essentially of two types. In one, the organism is actually seen by microscopic methods or isolated by special procedures. In the other, which is used extensively for the detection of viruses, the organism is assumed to be present when the body produces antibodies to a specific organism. If it can be shown that no antibodies to a specific agent were present in the blood at the onset of an illness, but a measurable number appeared from 10 to 30 days later, there is good evidence that the illness was caused by the specific agent against which the antibodies were produced.

Treatment.—In most cases, treatment of encephalitis consists of relieving the patient's symptoms. Nursing care, mechanical aids, certain medications, maintenance of proper nutrition and cleanliness ease the patient's unpleasant symptoms. Specific treatment of encephalitis is very limited because the causative agent is frequently unknown and because there are few drugs capable of destroying or counteracting the known causative agents. Psittacosis is the only virus-caused encephalitis for which there is specific treatment. Rabies can be prevented by the use of hyperimmune serum and vaccine, but treatment is useless once the symptoms have appeared. Certain chemical and biologic products are effective against some of the larger organisms, such as the protozoa and helminths.

Sequelae.—The sequelae (symptoms which remain after recovery from the acute phase) of encephalitis vary considerably, depending on the type of encephalitis, the extent of the inflammation, the severity of the illness, the age of the patient and other factors. Many patients are weak and debilitated following an attack but recover after a few days to a few months and have no adverse aftereffects. Some encephalitides (*e.g.*, eastern equine encephalitis and western equine encephalitis) cause irreparable damage in about 50% of patients under one year of age. In these cases, the immature brain is damaged to such an extent that it is incapable of developing properly; hence there may be intellectual impairment, delay or failure in behaviour maturation, motor disabilities and recurrent convulsions. The sequelae in adults and older children are generally less severe than those among infants. Such sequelae may consist of paralysis of an arm or leg, uncontrollable tremors, seizures, disturbances in speech or some type of intellectual defect.

Mortality.—The proportion of patients who die as a result of encephalitis may vary from a very small proportion, as in St. Louis encephalitis, to 100% in rabies. There appear to be varying rates, not only in different types of encephalitis but also in different outbreaks of the same type. The death rate is usually higher in those rare instances where encephalitis occurs as a late complication of some ordinarily nonencephalitic disease.

See also separate articles on the various encephalitides.

BIBLIOGRAPHY.—Josephine B. Neal, *Encephalitis* (1942); A. B. Baker (ed.), *Clinical Neurology*, 3 vol. (1955); Francis M. Forster (ed.), *Modern Therapy in Neurology* (1957); William S. Fields and Russell J. Blattner (eds.), *Viral Encephalitis* (1958); Thomas M. Rivers and Frank L. Horsfall, Jr. (eds.), *Viral and Rickettsial Infections of Man*, 3rd ed. (1959). (K. H. F.)

ENCINA, JUAN DEL (real name DE FERMOSELLE) (1468–1529?) poet, metrical theorist, first Spanish secular playwright and foremost Spanish composer of his time. The son of a cobbler, he was born on July 12, 1468, in Salamanca, where he probably studied under Antonio de Nebrija. After seven years in the duke of Alba's service, he emigrated at 30 to the papal court, having already produced almost all his work, most of it published in the first edition of his *Cancionero* (1496; facsimile, 1928). At 50 he took orders and went on a pilgrimage to Jerusalem, returning eventually to become prior of León, where he died at the age of 61. Although he is interesting as a dramatist and showed a great advance in technique from his 1492 nativity plays to the late Italianate dramatic eclogue, *Plácida y Vitoriano*, Encina's great appeal lies primarily in his delightful songs and carols. A modern edition of his complete plays appeared in 1893.

See F. Asenjo Barbieri (ed.), *Cancionero Musical de Palacio*, in which many of Encina's songs are transcribed (1890; new ed. by H. Anglés, 1947). (F. S. R.)

ENCKE, JOHANN FRANZ (1791–1865), German astronomer after whom a comet was named, was born at Hamburg, on Sept. 23, 1791. He was educated at Hamburg and at the University of Göttingen, where he worked under the direction of K. F. Gauss. In 1816 he became assistant at the observatory of Seeberg, Switz., was appointed vice-director there in 1820 and succeeded B. A. von Lindenau as director in 1822. In 1819, he proved that the apparently different comets of 1786, 1795, 1805 and 1818 were in fact one, and deduced for this comet, which now bears his name, the remarkably short period of about 3.3 years (*see* COMET). He received a royal medal of the Royal society in 1828 for this work. In 1825 Encke was appointed professor of astronomy and director of the observatory of the University of Berlin. He supervised the publication of the Berlin *Ephemeris*, into which he introduced many improvements, from 1830 to 1863. He also planned and supervised the construction of a new observatory at Berlin, which was completed in 1835.

The value of the solar parallax, 8″.57, derived by Encke from the results obtained at the transits of Venus in 1761 and 1769, was long accepted as authoritative. He died at Spandau on Aug. 26, 1865. (O. J. E.)

ENCLOSURE (Inclosure), a term describing the process by which common fields, common meadows and common pastures gave way to the modern pattern of small hedged fields and consolidated farms (*see* also Commons). In England the movement began in the 12th century, proceeded rapidly in the periods 1450–1640 and 1750–1860, and was virtually complete by the end of the 19th century. In the rest of Europe enclosure made little progress until the 19th century. Agreements to enclose were not unknown in Germany in the 16th century (for example in the Allgäu), but it was not until the second half of the 18th century that the government began to issue decrees encouraging enclosure. Even then little advance was made in west Germany until after 1850. The same policy of encouragement by decree was followed in France and Denmark from the second half of the 18th century, in Russia after the emancipation of the serfs (1861) and in Czechoslovakia and Poland after World War I. But the difficulties of procuring agreement have rendered the process in Europe slow, and it is still far from complete. Common rights over arable land—the most formidable obstacle to modern farming—have for the most part been extinguished, but land is still cultivated in scattered strips, and common rights continue over large areas of pasture and woodland. English experience, which is the subject of this article, spread over seven centuries, well illustrates the practical difficulties and the social disruption attending such a large-scale rearrangement of land holding and property rights.

Before enclosure, an English farmer in an open-field village possessed a number of strips of arable and meadow, lying dispersed in the open fields, and the right to graze stock on the common pastures. These rights were enjoyed by all tenants in common. Their cattle and sheep were put into one large village herd and flock and grazed all the village lands at permitted seasons: the arable fields after harvest and in the fallow year; meadows after hay harvest until the following spring when they were shut up for hay; common pastures all the year round unless, by common agreement, grazing had been limited to certain times of the year or to certain classes of stock. Under the open-field system, in short, the land was under the control of individual cultivators until the grain and hay were harvested; thereafter, it was at the disposal of the community.

To enclose land was to put a hedge or fence around a portion of the open land, whether in field, meadow or on the common pastures, and so prevent the exercise of common rights over it. In fields and meadows, this move was frequently preceded by some consolidation of strips by exchange or purchase. The advantage to the encloser lay in the fact that he could then farm his "close" as he pleased, since he had complete control of his land throughout the year. The loss to the community lay in the loss of common grazing. In theory, an encloser of land had to compensate his fellows for his enclosure by surrendering some of his common pasture rights; *i.e.*, by reducing the number of animals he grazed with the village flocks and herds. In practice, however, this rule was not strictly observed. The loss of grazing by the community as a result of enclosure caused few disputes when land was plentiful, and much enclosure took place in England in the middle ages without arousing opposition. Many farms, and indeed whole hamlets, came into existence as a result of the enclosure of land from waste, which lay usually on the outer edges of parishes. Manorial lords were pleased to allow the growing population to colonize and improve the land, since it increased their rents. Tenants raised no objection so long as plentiful grazing remained. Thus certain fields, and whole farms, were enclosed from the beginning of their existence, and were never subject to open-field regulations.

Enclosure in the Middle Ages.—The first complaints against enclosure were prompted by alarm at the rapid diminution in the size of the common pastures in the 13th century. This was a period of agricultural prosperity and extensive colonization of land. It resulted in the passage of the Statute of Merton in 1236 and the Statute of Westminster in 1285—the first mild legislation against enclosure. The Statute of Merton permitted manorial lords to enclose and improve their common wastes and pastures so long as they left enough for the tenants. It was frequently cited in later disputes between lords and their tenants, but its vagueness caused endless controversy on what was enough land for the tenants' animals. After the Black Death (*q.v.*), when farm labour was expensive and difficult to obtain, much enclosure was undertaken by manorial lords, and particularly by monastic houses, seeking to rationalize their estates and to manage their farms at less cost. Whole populations of hamlets and small villages were evicted or persuaded to move elsewhere, and village sites were left to fall into ruin. The purpose of many manorial lords in enclosing land in the 15th century seems to have been to turn the land over to pasture, and particularly to sheep farming. Labour costs were thus reduced, while the demand for wool increased steadily, at first for export, and later for supplying the expanding cloth industry in England. Manorial lords sought to gain possession of more and more of their tenants' land, extinguished common rights, turned grain fields into sheep pastures, common waste into sheepwalks and found large-scale sheep farming extremely profitable.

By the second half of the 15th century, voices began to be raised in protest at the social distress caused by these proceedings, and for the next century and a half the attack on enclosure was inseparably associated with the attack on sheep farming. Men almost forgot that enclosure did not necessarily mean enclosure for sheep, and assumed in argument that enclosure led inevitably to unemployment and depopulation. In fact the debate for and against enclosure should have been concerned with two problems only: (1) Was farming more efficiently carried out in open fields or in closes? Was there more to be gained by allowing the enterprising and ambitious to farm in complete freedom than by preserving common fields and common-field regulations and so protecting the land against the bad farmer? (2) Was it possible to devise a method of enclosure which adequately compensated everyone for the loss of common rights?

Enclosure Under the Tudors.—The outcry against enclosure in the 16th century at first focused attention on the activities of the enclosing sheepmasters. Later in the century, the attack was directed at both sheepmasters and cattle graziers. In reality enclosure was not confined to any one class of farmers. It was under way all over England, prompted by a rising population, a rising demand for food and rising agricultural prices. The urge to keep more sheep was strongest in the hills—the northern fells, the Yorkshire and Lincolnshire wolds, the Cotswolds and the Downs—where sheep were already indispensable for maintaining soil fertility. It was also strong on the better-drained claylands of the midlands where sheep were fattened for the London meat market.

But other districts of England practised different systems of husbandry. In some the sheep had no place because the land was too damp to support them. In others, the sheep flock was balanced by an equally important and valued herd of cattle. On the heavy clays of Northamptonshire and Leicestershire, for example, land was enclosed in order to put unsuitable arable land into permanent pasture to produce more beef and mutton; in Hertfordshire, in order to grow more wheat and more fodder for fattening pigs and bullocks. In the fenlands of East Anglia, around the Wash, where a pastoral husbandry was the only practicable system, there was little advantage in enclosing good summer grazing lands, and those who did so usually enclosed pasture in order to grow a little more grain. Some enclosures on the moderately heavy clay lands were prompted by the desire to introduce a system of convertible husbandry on land that was suffering from too much grain-growing.

In short, the outhurst of enclosure in the Tudor period was almost everywhere inspired by a desire to put the land to more productive use. The system of managing the new enclosures differed from region to region according to local husbandry practice.

The Tudors were responsible for much antienclosure legislation. Henry VII passed two acts in 1489 which attempted to deal with the problem and with the associated issue of depopulation. His concern was primarily to strengthen the country, and in particular the Isle of Wight, against possible foreign attack. The measure concerned with the decay of towns and buildings was the first, and not the most effective, of much Tudor legislation on the subject. Wolsey attempted vigorous enforcement of antienclosure legislation; an act of 1515 was more specific in its terms for the reconversion of pasture to arable land. This was followed up

in 1517 by 17 commissions of inquiry covering 35 counties, the results of which showed that the main area of trouble and discontent lay in the midland counties. Protector Somerset's enclosure commission of 1548 achieved little and was swallowed up in the political unheavals of 1549. The attitude of Elizabeth I's government toward enclosure varied; although no ban was put on enclosure for arable, an act of 1563 and a commission of 1565 tried to check enclosure for pasture. This act was repealed in 1593 owing to the "great plenty and cheapness of grain," but the move was premature. A series of harvest failures from 1594 onward caused renewed panic, and parliament resorted to fresh antienclosure measures in 1597. But in the house of commons debates of that year, and again in 1601, when repeal was considered, many arguments were heard in favour of enclosure and "leaving every man free."

The opinion of many members of parliament was gradually shifting toward a more laissez-faire attitude. Probably the tillage laws would have been discarded altogether in the next ten years had the harvests been plentiful. But in 1607, with almost no warning, the midland revolt broke out in Northamptonshire, Leicestershire and Warwickshire. It lasted only a month, but caused enough alarm to prompt the appointment of a new enclosure commission.

Early 17th Century.—By the end of the 16th century, the enclosure problem was somewhat better understood. Tudor legislators were beginning to realize that assorted motives lay behind enclosure and to admit that it did not lead inevitably to unemployment and depopulation. Indeed, in some counties it raised hardly any problems. Most counties in the southwest, west midlands and north had either enclosed all their land that was worth enclosing in the middle ages, or were managing to carry through enclosure by agreement in the 16th century without disturbance. Their success seems to have been connected with the fact that they were regions of pastoral husbandry where common grazing rights were not so highly prized (because of the large amount of land in pasture) as in the regions of mixed husbandry. The government finally recognized this truth by omitting certain counties from the operation of its antienclosure acts. It also omitted certain types of enclosure, admitting that some enclosures and conversions of arable to pasture were beneficial measures, designed to enhance fertility by introducing a system of convertible farming. By 1600 attention was confined to the enclosures that caused the worst social distress; namely, those taking place in certain midland counties where enclosure was usually the prelude to the conversion of arable to pasture, where the population was dense, land in short supply, alternative means of employment for the dispossessed few, and where enclosure invariably caused hardship to the poor through the loss of common rights.

The enclosure inquiry in 1607 was limited to Warwickshire, Leicestershire, Lincolnshire, Northamptonshire, Buckinghamshire, Bedfordshire and Huntingdonshire. Prosecutions followed in 1608. But in the next ten years, the activities of informers who had been mainly responsible for bringing enclosers to book, caused so much public outcry that in 1618 parliament appointed a commission to grant exemptions in reasonable cases. In 1624 most of the tillage statutes were repealed, while those of 1597 died for want of enforcement. The tide of opinion had turned as men saw the practical advantages of enclosure and less of the injurious social consequences. Angry attacks on the movement were still being published as late as 1653, but the evidence contained in them was drawn for the most part from an earlier period. The last antienclosure bill was introduced into the house of commons in 1656 and was rejected.

Some explanation for the softening of opinion toward enclosure must be sought in the changing methods by which it was achieved in the 15th to 17th centuries. Many enclosures of large acreages in the 15th century had been carried through by manorial lords with ruthless disregard for the interests of the tenants. In the course of the Tudor period, however, every landholding class in the community was engaged in enclosure. Since individuals wishing to enclose had to consolidate their strips first, and frequently did so by exchange rather than purchase, agreements gradually came to be recognized as the essential preliminary step. Sometimes they were a matter for two people only. Sometimes groups of tenants with intermixed lands arranged a more complicated exchange.

CROWN COPYRIGHT RESERVED—R.A.F. PHOTOGRAPH

THE PATTERN OF ENCLOSURE BY ACT OF PARLIAMENT (1771) AT KILBY, LEICESTERSHIRE, ENG.

Manorial lords, too, began to act in concert with their freeholders and tenants. A case from Leicestershire in 1597 shows a lord paying gratuities to those of his tenants who were reluctant to submit to enclosure.

By the 17th century, agreements were regarded as the usual method of enclosing a large amount of land, and a recognized procedure was adopted. The lord and his tenants chose outside arbitrators and surveyors to redistribute the land. A collusive suit in the court of chancery resulted in the award of a decree which gave the enclosure legal force.

Later Developments.—There are signs that the enclosure movement slowed down in the second half of the 17th and first half of the 18th century, probably because of general economic stagnation, but after 1750 when agricultural prosperity returned, the enclosure movement gathered momentum again. A new procedure was now adopted. Villages seeking to enclose land, which could procure the consent of people holding three-quarters or four-fifths of the land, obtained an act of parliament sanctioning enclosure. They then appointed enclosure commissioners who scrutinized the claims of all owners of land and common rights, procured surveys of the land, and redistributed it among all the just claimants. The lord received land in lieu of rights of soil, the parson in lieu of glebe and sometimes also in lieu of tithe, owners of land received allotments of land equivalent in acreage to their former holdings, and common-right holders received land in compensation for their surrender of rights. New roads were laid out and thousands of miles of new fencing erected.

Much controversy surrounded the new procedure both at the time and since. The whole process was far more costly than the old. Was it too costly? Did it deal fairly with the owners of common rights? Did it cause the decline of the small owners? Did it reduce or increase employment in agriculture? Contemporaries disagreed violently at the time and historians continue to do so, for no single verdict can be passed on a movement which affected thousands of townships and took a century and a half to complete.

It is generally agreed that the economic consequences of enclosure were beneficial. The efficiency of farming was greatly improved, and new techniques spread more rapidly than before. The social consequences were debatable. Even Arthur Young, a fervent contemporary advocate of enclosure, admitted that the poor were injured, sometimes grossly injured, by enclosure. Even when they were compensated for their loss of common rights, the small piece of land they received was worth far less to them than the right to graze, and to get timber, bracken for bedding and stones on the commons. Most modern writers, however, have tended to modify the harsh verdict passed on parliamentary enclosure by J. L. and Barbara Hammond in *The Village Labourer* (1911). The former consider that the commissioners acted with great fairness, often with indulgence, toward the small owners, that the cost of enclosure on average was not burdensome, that the immediate effect of enclosure was to increase rather than reduce the number of owners. They argue that the decline of the small owner was already well under way before the 18th century, and that his continued decline was caused less by enclosure than by general economic changes.

Enclosure procedure was cheapened after 1830 with the passing of several general enclosure acts. But when the open fields and commons had all but disappeared the public awoke to the loss of their green spaces around large towns. Finally, mainly because of the campaign of the Commons Preservation society, legislation was introduced in 1866, 1876, 1893 and 1899 to preserve common land for public enjoyment. By these measures many of the open spaces, for example, in London have been preserved.

BIBLIOGRAPHY.—M. Beresford, *The Lost Villages of England* (1954), "Habitation versus Improvement: the Debate on Enclosure by Agreement," in *Essays in the Economic and Social History of Tudor and Stuart England*, ed. by F. J. Fisher (1961); W. H. Chaloner, "Bibliography of Recent Work on Enclosure, the Open Fields and Related Topics," *Agricultural History Review*, II (1954); A. H. Johnson, *The Disappearance of the Small Landowner* (1909); *Royal Commission on Common Land, 1955–1958, Report*, H.M.S.O. Cmnd. 462 (1958); G. Slater, *The English Peasantry and the Enclosure of Common Fields* (1907); Joan Thirsk, *Tudor Enclosures*, Historical Association Pamphlet, General Series, no. 41 (1959); Marc Bloch, "La Lutte pour l'individualisme agraire dans la France du XVIII^e siècle," *Annales d'Histoire Économique et Sociale*, II (1930); J. H. Clapham, *The Economic Development of France and Germany, 1815–1914*, 4th ed. (1936); T. F. von der Goltz, *Geschichte der deutschen Landwirtschaft* (1902). (I. J. T.)

ENCYCLICAL, a letter designed for general circulation rather than for one particular recipient (from the Greek *en*, "in" and *kyklos*, "circle"). To this class belong the "general epistles" of St. James, St. Peter and St. John in the New Testament, and also such episcopal letters as Cyprian's *De lapsis*, which make regulations to be observed in controversial matters. In A.D. 476 the emperor Basiliscus composed a letter in which he attacked the doctrinal innovations of the Council of Chalcedon (451). This document, circulated for signature to the bishops of the Eastern Church and applied as a test of orthodoxy to clerics and laymen alike, proclaims itself to be a "sacred encyclical."

The word encyclical in modern times is most commonly used to denote letters, not necessarily infallible, issued by the pope to the various bishops (*Pacem in Terris*, however, issued by Pope John XXIII on April 11, 1963, was addressed further to "all men of good will"). The purpose often is to denounce errors in contemporary thought and practice, well known among such encyclicals being the *Quanta cura* of Pius IX (Dec. 8, 1864), which condemns rationalism, Gallicanism, socialism and those teachings that claim that human progress is best achieved independently of religious sanctions.

Widely different were the chief themes of the encyclical *Mater et magistra* of John XXIII (July 14, 1961), issued in the 70th year of the *Rerum novarum* of Leo XIII (May 15, 1891); both dealt mainly with the search for social justice. Still more different was *Pacem in Terris*, which, with the intention of ensuring world peace, proposed the establishment of a world political community.

The letters sent out by the Anglican bishops at the end of the Lambeth conferences are also called encyclicals.

ENCYCLOPAEDIA. The Greeks seem to have understood by encyclopaedia (*enkyklopaideia* or *enkyklios paideia*) instruction in the complete system or circle (*en kyklōi*) of learning (*paideia*). The expression corresponds to the modern "general education." The earliest information concerning the organization of knowledge and of teaching signified by the term is found in Greek and Latin writers of the 1st and 2nd centuries A.D. Pliny, in the preface to his *Natural History*, says that his book treats all the subjects of the circle of learning of the Greeks, and the expression is used to refer to a general education or to a system of arts and sciences by Quintilian, Plutarch, Lucian and Galen. The word encyclopaedia was probably first used in English by Sir Thomas Elyot in *The Governour* in 1531: "the circle of doctrine, whiche is in one worde of greke Encyclopaedia." In his Latin dictionary, in 1538, Elyot defines "Encyclios et Encyclia, the cykle or course of all doctrines," and "Encyclopaedia, that lernynge whiche comprehendeth all lyberall science and studies." Its first appearance in French is in Rabelais's *Pantagruel* (1533). The term does not seem to have been used as the title of a book in antiquity or the middle ages. It was used in titles in the 16th century, but more frequently for pedagogical or philosophical treatises concerning the liberal arts than for reference books covering all knowledge. "Encyclopaedia" means the liberal arts in such titles as Jacobus Philomusus' *Margarita philosophica encyclopaediam exhibens* (Strasbourg, 1508); and titles such as *Orbis disciplinarum liberalium ex P. Cosentino et M. Aurelio Cassiodoro* (Basel, 1528) are translations of "encyclopaedia." The *Lucubrationes vel potius absolutissima κυκλοπαιδεια* [*kyklopaideia*] of Joachimus Fortius Ringelbergius (Basel, 1541) is the second edition of his collected works on the organization of studies, on grammar, rhetoric, dialectic, mathematics and the other arts. Paulus Scalichius de Lika's *Encyclopaediae seu orbis disciplinarum epistemon* (Basel, 1559) is a more heterogeneous collection of essays.

Johann Heinrich Alsted's *Encyclopaedia* (1630), on the other hand, is a compilation of information organized according to a classification of the arts, and treats of everything that can be learned by man in this life. Alsted had earlier published a philosophic organization of knowledge in his *Encyclopaedia cursus philosophici* (1608) and his *Panacaea philosophica i.e. methodus docendi et discendi universam encyclopaediam* (1610). He later published a *Triumphus Biblicus sive encyclopaedia Biblica*. The term "cyclopaedia" was also used and defended, but writers such as Vossius, in the 17th century, argued that it is without classical authority and is incorrect etymologically.

From the 17th century to the present the word encyclopaedia was gradually restricted to mean a work that treats all the various kinds of knowledge. In the process, encyclopaedias were distinguished from pedagogical treatises on the arts and sciences and from dictionaries that define words rather than explain subjects. Pedagogical manuals were often called encyclopaedias in the 16th century; and encyclopaedias were often called dictionaries and lexicons in the 17th and 18th centuries. The term tended further to be restricted to reference books for information as distinguished from philosophical systems for speculation and technical compendiums for research.

In the past, Bacon's *Novum Organum* has often been called an encyclopaedia, and editors of encyclopaedias have sometimes gone to Bacon, or alternatively to Plato, Speusippus, Democritus, Aristotle or the skeptics, for the organization of knowledge. Carl Christian Schmid's *Allgemeine Encyklopädie und Methodologie der Wissenschaften* (Jena, 1810) is a philological organization of knowledge; Hegel's *Encyklopädie der Philosophischen Wissenschaften im Grundrisse* (Heidelberg, 1817) is a philosophical organization.

In the 17th century the term was applied also to works that used all the sciences in application to a particular subject, as in Fortunius Licetus' treatises on antiquities. It is still attached to works that profess to be comprehensive treatises on particular subjects—such as encyclopaedias of wit, sport, agriculture, geography, technology, arts and manufactures, law and, more recently, social sciences and world literature—or of particular religions or cultures, such as Islamic, Catholic or Jewish encyclopaedias.

Following are the major sections and subdivisions of this article:

I. Early Encyclopaedias
  1. Classical
  2. Early Christian
  3. Byzantine
  4. Muslim
  5. Early Oriental
  6. Late Medieval and Renaissance
II. Modern Encyclopaedias
  A. 17th Century
    1. Moréri
    2. Hofmann
    3. Bayle
    4. Furetière
    5. Chauvin
  B. 18th Century
    1. John Harris
    2. Hübner
    3. Jablonski; Walch
    4. Ephraim Chambers
    5. Zedler
    6. French Encyclopaedists
  C. 19th Century
    1. "Conversations-Lexicon" and the Learned Encyclopaedia
    2. England and the United States
    3. France
  D. 20th Century
    1. England and the United States
    2. Encyclopaedia Annuals
    3. France
    4. Italy
    5. Germany
    6. Austria
    7. U.S.S.R.
    8. Slavic and Baltic Encyclopaedias
    9. Hungary
    10. Finland
    11. The Netherlands
    12. Scandinavia
    13. Switzerland
    14. Spanish- and Portuguese-Language Encyclopaedias
    15. Greece
    16. Turkey
    17. Indonesia
    18. Japan
    19. Canada
    20. Australia
    21. Specialized Encyclopaedias
III. Encyclopædia Britannica
  1. First Edition
  2. Third Edition
  3. Supplements and Other Editions
  4. Ninth Edition
  5. 11th Edition
  6. 14th Edition
  7. Britannica Plan of Continuous Revision
  8. Later History

I. EARLY ENCYCLOPAEDIAS

1. Classical.—The Greek philosophers laid the foundations of the classifications used in later encyclopaedias, and many of them, including Democritus, Plato, the Sophists, Speusippus and Aristotle, have been credited with writing "encyclopaedias." The Romans, however, compiled the first encyclopaedias in the strict sense of compilations of information from all the sciences rather than philosophical systems of ideas or handbooks of techniques. Varro (*q.v.*; 116–27 B.C.), who was called by Quintilian the "most learned of the Romans," employed all four of the systems of organization followed by later encyclopaedists: history, arts, biography and words or subject matter. His *Antiquities of Human and Divine Things* was a historical encyclopaedia. His nine books of *Disciplines* were an encyclopaedic organization of knowledge under the nine arts—grammar, logic, rhetoric, arithmetic, geometry, astronomy, music, medicine and architecture; the first seven, grouped into trivium and quadrivium, were the medieval liberal arts.

Varro's *Portraits* was a collection of 700 biographical sketches; *On Agriculture* and *On the Latin Language* set forth information about things and words of interest to the Romans. Varro's work survives only in fragments.

The most ancient extant encyclopaedia, the *Natural History* of Pliny the Elder (*q.v.*; *c.* A.D. 23–79), is organized according to subject matter. It is divided into 37 books and 2,493 chapters. Book 1 sets forth the organization of knowledge to be followed in the remaining books; book 2 treats cosmography, astronomy and meteorology; books 3–6 geography; book 7 anthropology and physiology; books 8–11 zoology; books 12–19 botany; books 20–32 medical uses of plants and animals; books 33–37 metals, minerals, stones, gems. The invention of the arts is treated under the history of man in book 7 and the fine arts are included in the treatment of minerals and gems.

Roman historians after Polybius had tended to treat Roman history as world history and universal history, and many historians, such as Cornelius Nepos and Suetonius, wrote *Lives of Illustrious Men*, including poets, orators, historians and philosophers as well as politicians and generals. The *Satyricon* or *On the Marriage of Philology and Mercury* of Martianus Capella (fl. 470) treats the seven liberal arts in an allegorical framework and in alternating prose and verse. Pliny and Martianus Capella were influential during the middle ages, and their works were published frequently during the early years of printing.

2. Early Christian.—The writings of St. Augustine and St. Jerome provided the basis for a Christian reorganization of knowledge: Augustine's *City of God* and *On the Trinity* establish historical and theological principles; Jerome's *Chronicon* and *On Illustrious Men* are beginnings of historical and biographical encyclopaedias. *The Institutions of Divine and Secular Letters* of Cassiodorus (*q.v.*; *c.* 490–*c.* 585) is a small encyclopaedia of the arts and disciplines in two books, the first concerned with the Old and New Testaments, the second with the seven liberal arts.

Isidore of Seville (*q.v.*; *c.* 560–636) wrote encyclopaedias, using all four modes of organization of information—history, arts, biography and words or subject matter. His *Chronicon* traces the history of the world from creation to A.D. 616. His *On Illustrious Men* supplements Jerome's biographies. His *On the Nature of Things* covers meteorology and astronomy, and *Differences and Properties of Words* treats, in two books, differences of words and differences of things. His *Etymologies* (also called *Origins*) in 20 books combines an organization based on the arts and on subject matter. The liberal arts are treated, with frequent borrowings from Cassiodorus, in the first three books, medicine in book 4, law and chronology in book 5, ecclesiastical books and offices in book 6. The last 14 books proceed according to subject matter from God, angels and the faithful (book 7) through the church (book 8), languages and society (book 9), man, animals (books 10–12), geography (books 13–14) and the practical arts (books 15–20), ending with food and drink and domestic and rural implements (book 20). Book 10 is a dictionary of Latin words arranged alphabetically. The Venerable Bede (*q.v.*; 672/3–735) used a subject matter organization in *On the Nature of Things* and a historical organization in the *Chronicon*. Rabanus Maurus (*see* RABANUS MAURUS, MAGNENTIUS) begins with God and angels and then rearranges materials derived from books 7–20 of Isidore's *Etymologies* in his *On the Universe* (also called *On Universal Nature*, *On the Nature of Things* or *On the Origin of Things*) in 22 books and 325 chapters.

By the 12th century, the liberal arts had been considerably expanded. The *Heptateuchon* of Thierry of Chartres (d. *c.* 1150) is a comprehensive compilation of extracts and quotations on the seven arts. Hugh of St. Victor's (1096–1141) *Didascalion* treats not only the seven liberal arts but also the seven mechanical arts (weaving, armoury, navigation, agriculture, hunting, medicine and theatre) as well as the methods of scriptural interpretation (historical, allegorical and tropological or moral), the parts of philosophy and the parts of magic.

Honorius of Autun (*c.* 1090) systematized and supplemented previous collections of lives of illustrious men. His *De imagine mundi* was followed by other encyclopaedias based on subject matter arrangements: Lambert of St. Omer's *Liber floridus* (*c.* 1120), William of Conches' (1080–1145) *On the Philosophy of the World*, Alexander Neckam's (1157–1217) *On the Natures of Things*, Thomas of Cantimpré's *On the Nature of Things* (*c.*

1228–44), Bartholomaeus Anglicus' (fl. *c.* 1220–40) *On the Properties of Things.* Thomas of Cantimpré's work is in 19 books and treats in sequence man, animals, plants, waters, stones and metals, astronomy and the elements. Bartholomaeus' encyclopaedia is also in 19 books, but it runs a sequence from God through angels, psychology, physiology, domestic economy, medicine, cosmology, chronology, birds, fishes, geography, minerals, trees and animals to a final book on colours, odours, food, drink, weights and measures. Bartholomaeus addressed his book to plain people—*simplices et rudes*—and it was extremely popular for three centuries. An English translation was completed in 1398 by John Trevisa; it was printed by Wynkyn de Worde around 1495. Universal history, finally, was developed in the early 13th century by Roger of Wendover and Matthew Paris (*qq.v.*).

3. Byzantine.—Eastern Christianity continued the speculative tradition of the Greeks as contrasted with the practical concerns of the Romans. This difference is reflected in the lexicographical and encyclopaedic works that flourished from the 9th to the 11th centuries. Photius (*q.v.*) finished his *Bibliotheca* about 857; it is a series of reviews or summaries of books ranging through many fields. A Greek lexicon was completed either by Photius or under his direction, systematizing earlier lexicons. The *Suda Lexicon* (completed between 950 and 976) has information about subject matters as well as words, particularly in the history and interpretation of literature, but also in scientific fields. Psellus (1018–*c.* 1079) organized his *De Omnifaria Doctrina* ("All-Inclusive Instruction") in a series of 193 questions and answers, according to subject matter, beginning with God and creation, running through natural history and astronomy and ending with various curious and practical questions.

4. Muslim.—Muslim philosophers, who early faced the problem of reconciling Greek philosophy with Islam, tended to be encyclopaedic in scope. The *Catalogue of the Sciences* of al-Farabi (*c.* 870–950) was widely known in Latin translation during the middle ages under the title *On the Sciences* or *On the Origin of the Sciences.* It treated the sciences under eight heads: (1) linguistic; (2) logic (including rhetoric and poetic); (3) mathematics (including optics, astronomy, music, the science of weights, mechanics); (4) physics (including meteorology, botany, zoology and psychology); (5) metaphysics; (6) politics; (7) jurisprudence; (8) the *kalam* (scholastic theology or the science of discourse). Al-Masudi's *Meadows of Gold and Mines of Precious Stones* (*c.* 947) is a historicogeographical encyclopaedia. The Brethren of the Purity also prepared in the 10th century an encyclopaedia in 52 treatises, divided into three main heads: (1) disciplinary sciences (including grammar, arithmetic, divination, magic, arts, trades and commerce); (2) juridicoreligious sciences; and (3) philosophical sciences, which in turn are divided into mathematical, logical, natural and divine sciences.

Al-Khwarismi, the Persian encyclopaedist, in *Keys of the Sciences* (*c.* 976) divided the sciences into the sciences of the religious law and exotic sciences or philosophy. The former include jurisprudence, the *kalam*, grammar, secretarial art, prosody and poetry, history; the exotic sciences are divided into theoretical (physics, mathematics, theology) and practical (ethics, economics, politics). Al-Nadim's *Index of the Sciences* (987–988) is in ten sections which range through linguistic, literary and historical sciences, the *kalam*, jurisprudence, philosophical sciences (including mathematics, medicine and engineering), magic and fables, sects and creeds, and alchemy. Avicenna's (980–1037) *Ash-Shifa* is the best known of his philosophical encyclopaedias (*see* AVICENNA and ARABIC PHILOSOPHY). *The Revival of the Religious Sciences* of al-Ghazzali (*q.v.;* 1058–1111) divides sciences into religious and nonreligious; the *kalam* and philosophy (mathematics, logic, natural theology and natural sciences) fall under neither of these heads. Averroës (*q.v.;* 1126–98) wrote a medical encyclopaedia, *Colliget*, and his commentaries on the works of Aristotle constituted a vast philosophical encyclopaedia. The *Prolegomena* of Ibn Khaldun (*q.v.;* 1332–1406) is an encyclopaedic sociology and philosophy of history, and contains a classification of the sciences and some treatment of their history.

5. Early Oriental.—Chinese encyclopaedias, like some western compilations, consist of extracts of works. One of the earliest, the *Shih lei fu* of Wu Shu (947–1002), treats celestial and terrestrial phenomena, mineralogy, botany and natural history in 30 books and 100 chapters. Wu Shu also collaborated with Li Fang on a larger encyclopaedia, *T'ai-ping yü-lan* (987), containing extracts from 1,690 works arranged in 55 sections and 1,000 books. Ma Tuan-lin compiled *Wên hsien t'ung k'ao* in the 14th century. The *Great Standard*, compiled under the direction of the emperor Yung-lo in the 15th century and designed to include all that had ever been written on the Confucian canon, history, philosophy and the arts and sciences, contained 22,937 books, making it easily the largest encyclopaedia. The encyclopaedia prepared for the emperor K'ang Hsi (d. 1722) is in 5,020 volumes.

Hindu literature contains no encyclopaedias in either the Greco-Christian-Muslim or in the Chinese sense, but it is suggestive to think of the puranas (2nd to 6th century A.D.) as encyclopaedic compilations adapted to Hindu culture. They are neither organizations of information from the arts and sciences nor collections of passages from the classics, but didactic epics expounding ancient legendary lore—cosmogony, mythical descriptions of the earth, the cosmic ages, exploits of gods and heroes, genealogies, ancient law and ritual.

6. Late Medieval and Renaissance.—Latin encyclopaedias were profoundly influenced by translations of Aristotle during the second half of the 12th and the first half of the 13th century. Arabic and Hebrew philosophic and encyclopaedic works were translated at the same time. Gundissalinus' *On the Division of Philosophy*, translated sometime after 1140, shows the influence of al-Farabi's classification. It has chapters on natural science, mathematics, divine science, grammar, poetic, logic, medicine, arithmetic, music, geometry, optics or perspective, astrology, astronomy, the science of weights, natural capacities or engines and practical philosophy. The 13th-century philosophers—Albertus Magnus, Thomas Aquinas, Bonaventura and Roger Bacon—were encyclopaedic in scope. The *Opus majus* of Roger Bacon, thus, covers all knowledge in seven parts: (1) the causes of error and the obstacles to the discovery and propagation of truth; (2) the dependence of all philosophy and science on revealed truth; (3) languages; (4) mathematics; (5) perspective or optics; (6) experimental science; (7) moral philosophy.

Vincent de Beauvais (*q.v.; c.* 1190–*c.* 1264) proposed in his *Speculum majus* to reflect "all things of all times." It was divided into four parts—a natural, a doctrinal, a moral and a historical mirror. Three were compiled by Vincent; the *Moral Mirror* was completed between 1310 and 1325. Like the Chinese encyclopaedias it is largely a collection of extracts, some of them very long. About 450 authors are represented, Greek, Arabic and Hebrew as well as Latin. The *Natural Mirror* is organized according to subject matter in 32 books, and the *Doctrinal Mirror* is organized according to the sciences in 17 books. The *Historical Mirror* is organized on a chronological basis and traces the history of the world from creation to 1254, while the *Moral Mirror* is organized according to problems and distinctions in three books.

Brunetto Latini, noted for his association with the young Dante, wrote a French encyclopaedia, *Li Livres dou Trésor* (the "Treasure Books"), in the 1260s.

Encyclopaedic developments from the 14th to the end of the 16th century took two directions: (1) combination of the different modes of organizing knowledge; and (2) separation of encyclopaedic from philosophical, pedagogical, philological, lexicographical, biographical and historical compilations. The works that are usually listed as encyclopaedic follow the first tendency. Pierre Bersuire (Berchorius; d. 1362) wrote an encyclopaedia in three parts, using three bases of classification for moral ends: the Bible in his *Moral Reduction of the Whole Bible* in 34 books; things in his *Moral Reduction of the Properties of Things* in 14 books; and words in his *Dictionary* of 3,514 terms, arranged alphabetically and expounded morally. Gregory Reisch (d. 1525) used a classification according to the arts and another according to subject matter in his *Margarita philosophica* (first published in 1496, followed by eight editions to 1535). The *Commentarii urbani* of Raphael Maffei (or Volaterranus) (first edition 1506,

followed by eight editions to 1603) is in three tomes: *Geography, Anthropology* and *Philology*. The third part is a survey of knowledge according to subject matter, including the "cyclic" sciences.

The second direction led to the distinction of encyclopaedias in the modern sense from other universal organizations of knowledge. Most philosophers of the period were "encyclopaedic," but four influenced encyclopaedias directly. Raymond Lully (*see* LLULL, RAMON; *c.* 1235–1315) developed an "art" which had three parts: a universal characteristic or language, a combinatory art or grammar and an "encyclopaedia" of the knowledge acquired and the meanings of the words. Petrus Ramus (1515–72) reformulated and reorganized the liberal arts and philosophy. Francis Bacon (1561–1626) devoted his philosophical ingenuity to the organization and advancement of learning. He developed an elaborate classification of learning in three parts, history, poetry and knowledge, corresponding to the faculties of memory, imagination and reason. Comenius (Jan Amos Komensky) (*q.v.;* 1592–1670), who was acquainted with the work of Bacon, proposed a reform of education that would spread "universal light" by means of universal books, universal schools, a universal college and a universal language. He planned three kinds of universal books: a *Pansophia* containing the marrow of truth, a *Panhistoria* treating particular things and a *Pandogmatica* reviewing theories about things. J. T. Jablonski, a grandson of Comenius, published an *Allgemeines Lexicon der Künste und Wissenschaften* in 1721.

Early works of erudition were called polymathy, polyhistory, philology, literary history and sometimes encyclopaedia, but for the most part scholars criticized "encyclopaedias" as superficial and secondary. Pedagogical treatises were concerned with the relations of the arts and sciences. Such works sometimes used the word encyclopaedia in their titles, and the larger treatises on the arts are sometimes classed as encyclopaedias. Giorgio Valla (*c.* 1430–99) in *On Things to be Sought and Things to be Avoided,* two folio volumes, divided "things" into three kinds—mental, bodily and external—and then treated them in 49 books. The numerous Renaissance collections of biographies laid the foundations both of modern bibliographies and modern encyclopaedias. The beginnings of modern lexicography were involved in controversies concerning the separation of dictionaries of words from dictionaries of things. Early historical and geographical dictionaries developed with little change of basic plan into encyclopaedias.

II. MODERN ENCYCLOPAEDIAS

A. 17TH CENTURY

Antonio Zara in his *Anatomia ingeniorum et scientiarum* (1614) based his organization of knowledge on an analogy similar to Bacon's. The work is divided into four sections: (1) on the dignity and excellence of man in all his bodily and mental aspects; (2) on 16 sciences of the imagination; (3) on 8 sciences of the intellect; and (4) on 12 sciences of memory. Johann Heinrich Alsted (1588–1638) was influenced by both Lully and Ramus. His *Encyclopaedia septem tomis distincta* (1630, two folio volumes) is organized according to the arts or disciplines in seven parts: the presuppositions of the disciplines, philology, theoretic philosophy, practical philosophy, the three superior faculties (theology, jurisprudence, medicine), mechanical arts and the remaining disciplines under the title *Medleys of Disciplines* (*Farragines disciplinarum*), including mnemonics, history, chronology, architecture, magic and many other curious "arts."

Jean de Magnon planned a historical encyclopaedia in French heroic verse, in ten volumes, which was to render libraries a useless ornament. He completed only one volume of *La Science universelle* (1663), beginning with the nature of God and ending with the history of the fall of man. Charles Estienne's *Dictionarium historicum ac poëticum* (1553) was often reprinted before its final edition, augmented and perfected, by Nicholas Lloyd in Oxford, 1670. Juigné-Broissinière's *Dictionnaire théologique, géographique, poétique, cosmographique, et chronologique* (1644), a translation and inaccurate expansion of Estienne's work, went through 12 editions in less than 30 years in spite of its defects.

1. Moréri.—The Juigné-Broissinière *Dictionnaire* was superseded by Louis Moréri's (1643–80) *Le Grand Dictionnaire historique, ou le mélange curieux de l'histoire sacrée et profane* (one folio volume, 1674). Moréri prepared a second edition which was completed after his death (two volumes, 1684). The sixth edition was edited by the scholar and philosopher Jean le Clerc (four volumes, 1691). The 20th and last edition (ten volumes, 1759) appeared after the publication of the seventh volume of the French *Encyclopédie*. Moréri's *Dictionnaire* was translated twice into English, a translation of the sixth edition in 1694 and a translation and augmentation of the eighth edition by Jeremy Collier under the title *The Great Historical, Geographical, Genealogical, and Poetical Dictionary* in four volumes, 1701–21, and with additions in 1727. It was translated with additions peculiar to each country into German three times (1709, 1722 and 1726), Dutch (1725) and Spanish (1773). Peter the Great is reported to have ordered a Slavonic translation, and an Italian translation was begun.

Moréri is arranged alphabetically. It consists of proper names of persons and places; general historical and geographic terms are also explained, and there are articles on books, such as the Koran and Genesis, and on arts. The article "Architect" gives the extract from Vitruvius on the encyclopaedic knowledge required by the architect. The articles on writers, such as Aristotle, contain sketches of their doctrines and influences.

2. Hofmann.—Johann Jacob Hofmann (1635–1706) published his *Lexicon universale historico-geographico-chronologico-poëtico-philologicum* in 1677 in two folio volumes. It is a dictionary of proper names of persons and places. In 1683 he published a continuation in two folio volumes, with additional proper names and also the history of animals, plants, stones, metals, elements, stars, man and his affairs, honours, laws, magic, music, rites and other subjects. The second edition of the *Lexicon* (four folio volumes, 1698) incorporates these additions in their alphabetical places. Whereas Moréri's vernacular *Dictionnaire* grew in successive editions in the 18th century, Hofmann's Latin *Lexicon* was not reprinted after the 17th century.

3. Bayle.—Pierre Bayle (*q.v.;* 1647–1706) published "Project for . . . a Critical Dictionary" (1692), which was to contain correction of errors and omissions of Moréri and other dictionaries. This project was abandoned because later editions of Moréri, particularly Le Clerc's edition, removed most of the errors. Bayle published instead his *Dictionnaire historique et critique* (two folio volumes, Rotterdam, 1697). The editors of the eighth edition of Moréri (1698) acknowledge their use of some of Bayle's corrections. The original project was to include "real" as well as "personal" articles, but the *Dictionnaire* was limited to persons, and definitions of things (two had appeared in the work of 1692) were dropped. The arrangement is alphabetical and each article is divided into two parts: a factual account and "remarks"—critical, controversial and philosophic. Bayle was a skeptic in philosophy and was attacked not only by French Catholic theologians but also by Le Clerc, who followed Locke in philosophy. Bayle published a second edition in three folio volumes in 1702, and the fourth edition (four folio volumes, 1720) was enlarged from his manuscripts. Two English translations were made. Des Maizeaux translated the second edition (four folio volumes, 1709; second edition, five volumes, 1734–38).

An adaptation of the fifth edition by J. P. Bernard, T. Birch and J. Lockman appeared in ten volumes in 1734–41. Birch, who was author of *The History of the Royal Society of London,* wrote most of the English lives included; they were so highly esteemed that many of them (*e.g.,* W. Chillingworth, Mrs. Cockburn, R. Cudworth, John Greaves, Milton, Raleigh) were prefixed to editions of works of those authors published between 1742 and 1753. These articles and other additions were published in J. G. Chaufepié's *Nouveau Dictionnaire historique* as a supplement to Bayle (four folio volumes, 1750–56). Prosper Marchand, editor of the fourth edition, left materials, published posthumously, for a supplementary *Dictionnaire historique* (two folio volumes, 1758). The 11th and last edition (16 octavo volumes, 1820–24) includes the supplementary and additional materials. Bayle's dictionary was also translated into German (four folio volumes, 1741–44) with a preface by J. C. Gottsched.

4. Furetière.—Dictionaries of arts and sciences were distinguished from dictionaries of words by their concern with subject matter, much as encyclopaedias were distinguished from philosophical and pedagogical treatises on the arts by their emphasis on information rather than on erudition or teaching; historical dictionaries were distinguished from bibliographies, biographies, histories and gazetteers by their scope. The Italian dictionary of the Accademia della Crusca was published in 1612. The French Academy started work, in 1639, on a French dictionary that was to exclude words belonging specifically to the arts and sciences. Abbé Antoine Furetière (*q.v.*), one of the editorial committee, was expelled from the academy for planning another dictionary giving "all the terms of the Sciences and the Arts." César Pierre Richelet's dictionary, published in Geneva in 1680, had covered history, grammar, criticism, jurisprudence and other arts, and went through many editions. Furetière's *Dictionnaire universel* appeared in three folio volumes in 1690. The French Academy, in response, commissioned Thomas Corneille (*q.v.*), brother of the dramatist, to prepare *The Dictionary of Arts and Sciences*, reversing the order of Furetière's "Sciences and Arts" (two folio volumes, 1694). It was reprinted in 1696, 1720 and with revisions, 1732. Corneille acknowledged indebtedness to existing dictionaries in the particular arts and sciences. Corneille also wrote a *Dictionnaire universel géographique et historique* (1708). Furetière's dictionary formed the basis of the *Universal French and Latin Dictionary of Trévoux*, edited by the Jesuits. The editors of the *Dictionary of Trévoux* distinguished between dictionaries that gave only the significance of words and those that described things indicated by words, between "vocabularies" and "dictionaries of notions." The French *Encyclopédie* was directed in part against the *Dictionary of Trévoux*, and the *Encyclopédie* project was called ridiculous and impossible in the fifth edition of the *Dictionary* (1752), after two volumes of the *Encyclopédie* had appeared.

5. Chauvin.—Descartes's philosophy provided the basis for a kind of encyclopaedia different from those based on Lully, Ramus, Bacon or the skeptics. He emphasized method rather than doctrine and, following his lead, Étienne Chauvin (1640–1725) wrote a *Lexicon rationale sive thesaurus philosophicus* (one folio volume, Rotterdam, 1692; second revised edition under title *Lexicon philosophicum*, 1713) to explain the various senses in which words have been used by different philosophers. The terms are drawn from philosophy, including logic and rhetoric, and from the natural sciences. The *Lexicon* contains no proper names of places and almost no names of persons. Many of the terms are from physics, astronomy and physiology; there are articles on instruments such as barometer, microscope, telescope, pendulum; and 30 pages of plates of astronomical figures and "machines." The longer articles (on subjects such as light, magnetic bodies, local motion and moral virtues) differentiate the positions of different philosophers, usually the Peripatetic, the Cartesian and the atomic, although other positions are frequently treated.

B. 18th Century

Vincenzo Maria Coronelli, a Franciscan friar, worked on a general alphabetical encyclopaedia, written in Italian, *Biblioteca universale sacroprofana*. It was to have treated more than 300,000 words, to have covered other subjects as well as history and biography and to have extended to 45 folio volumes. However, only seven volumes were published, "A-Caque" (Venice, 1701–06).

1. John Harris.—The first alphabetical encyclopaedia in English, John Harris' (*q.v.*; *c.* 1666–1719) *Lexicon technicum, or an Universal English Dictionary of Arts and Sciences* (one folio volume, 1704), professes to explain not only the terms of art but the arts themselves. Harris finds Chauvin too full of school terms and defective in the modern improvements in mathematics and physics, and he considers the dictionaries of Corneille and Furetière to be of language rather than of arts and sciences. Harris treats no proper names; his treatments of terms of logic, metaphysics, ethics, grammar and rhetoric are few and brief (there is thus no article on "syllogism," which is treated at length in Chauvin); his interest centres in mathematics, physics, anatomy, law, heraldry, navigation, ships and similar subjects. The first volume was republished in 1708, and a second volume, consisting largely of mathematical and astronomical tables, appeared in 1710. It also contains Sir Isaac Newton's treatise on acids. A fifth edition appeared in 1736 in two volumes, and a supplement in 1744.

2. Hübner.—The "conversation-lexicon," a simplified handy reference book of short entries, was to have profound effect on the later development of encyclopaedias. Johann Hübner (1668–1731), in his preface to *Reale Staats-Zeitungs und Conversations-Lexicon* (one volume octavo, 1704), draws up a list of 100 words (such as exchequer, Whigs, Tories, tariff, divan, pragmatic sanction, lord-mayor, Caballa) to show the utility of the work as an aid in reading newspapers. The contents resemble Moréri's, much abbreviated and with biography omitted: religion, orders, states, rivers, towns, castles, mountains, genealogy, war, ships and in general the political condition of the world.

Hübner is not the author of the *Lexicon*. It went through five editions by 1712; the 31st edition, in four volumes, was published in 1824 and a revised edition appeared in 1874. After the appearance of the fifth edition Hübner was asked to write the preface to a second part: *Curieuses und Reales Natur-Kunst-Berg-Gewerb- und Handlungs-Lexicon* (one volume octavo, 1712, frequently reprinted to 1792). It treats of nature, science, art and commerce and was one of a series of *Reallexicons* in one volume, each on a different subject. Another, J. Mencke's *Compendiose Gelehrten-Lexicon*, or biographical dictionary of authors, published first in 1715 and revised by C. G. Jöcher in 1726 and 1733, was the basis of Jöcher's *Allgemeines Gelehrtenlexicon* (1750–51; frequently reprinted and revised in 11 volumes, 1897).

3. Jablonski; Walch.—Johann Theodor Jablonski (1654–1731) returned from proper names to arts and sciences in his *Allgemeines Lexicon der Künste und Wissenschaften* (one volume quarto, 1721); it does not treat history, geography, biography or genealogy and it omits theology from the sciences. It was republished, enlarged, in 1748 and again, in two volumes, in 1852.

Johann Georg Walch's *Philosophisches Lexicon* (one volume octavo, 1726) undertakes to treat all the parts of philosophy—logic, metaphysics, pneumatics, ethics, natural theology, jurisprudence and politics. Walch enumerates a dozen earlier philosophical lexicons (including Chauvin and Jablonski, but not Harris). He finds Chauvin deficient in two respects: emphasis on the moderns to the neglect of the ancients and use of verbal definition. Walch undertakes to consider things as well as words and to explain the meanings of the ancients as well as the moderns.

4. Ephraim Chambers.—In 1728 Ephraim Chambers (*q.v.*) published his *Cyclopaedia; or an Universal Dictionary of Arts and Sciences, Containing an Explication of the Terms, and an Account of the Things Signified Thereby, in the Several Arts, both Liberal and Mechanical, and the Several Sciences, Human and Divine . . . the Whole Intended as a Course of Ancient and Modern Learning* in two folio volumes. In his preface he acknowledges his indebtedness to previous dictionaries—among others, those of the French and Italian academies, Furetière, Trévoux, Chauvin and Harris. He carefully defines each of the terms of his title—art, science, term of art, definition and finally dictionary. Sciences are discovered by use of sense and reasoning; arts are portions of sciences considered relative to circumstances. A schematism of knowledge is constructed according to three faculties—senses (either alone, as in natural history, or with art, as in anatomy and chemistry), imagination (grammar, rhetoric, poetry) and reason (physics, metaphysics, logic, mathematics and the arts subordinate to these, as agriculture, painting, trade, manufacture, policy, law). This schematism is elaborated in a classification of knowledge into natural and scientific (divided into sensible and rational) and artificial and technical (divided into internal and external) and eventually into 47 sciences and arts. This system and the cross references provide interconnections among the articles which are arranged alphabetically.

Chambers differentiates three kinds of dictionaries: grammatical (which define words by words), philosophical (which give the general force of words) and technical (which give the particular senses attached to them in one or more arts). Unlike the historical or biographical dictionaries, Chambers' *Cyclopaedia* treats no

names of persons or places. There are articles on Aristotelian and Aristotelians, but not on Aristotle; on Cartesian and Cartesians, but not Descartes; etc. There are articles on each of the arts and sciences and on the important terms of art. The differences of the schools on important terms are expounded; thus under "descent" in mechanics, two opposite hypotheses of the cause of descent are given—as an internal principle (the Peripatetics, Epicureans and Newtonians) and as an external principle (the Cartesians and Gassendists); and in the article on gravity the opinions of 12 writers, from Aristotle to Newton, are set forth concerning the cause of gravity. Seven editions had been published in London by 1751–52, as well as one in Dublin in 1742. Chambers declined an invitation in 1739 to publish a French edition, but an Italian translation was prepared (nine volumes quarto, Venice, 1748–49) and was the first complete Italian encyclopaedia.

The materials for seven additional volumes, which Chambers had collected, were reworked after his death in 1740, first by John Lewis Scott, then by John Hill as a *Supplement* (two folio volumes, 1753). Abraham Rees published a revised and enlarged edition in two folio volumes (1779–88), which was reprinted in London (five volumes, 1786–88 and 1789–91) and Dublin (four volumes, 1787), and formed the basis for Rees's *Cyclopaedia* in 45 volumes.

5. Zedler.—Johann Heinrich Zedler, a Leipzig bookseller (1706–63), published *Grosses Vollständiges Universal Lexicon*—"The Great Complete Universal Lexicon of all Sciences and Arts Which Have Yet Been Discovered and Improved by Human Understanding and Ingenuity" (to translate only the first part of his title)—in 64 volumes (1732–50). The *Supplement* came to a halt, however, at almost the same point as Coronelli had (four volumes, 1751–54, "A-Caq"). The materials were prepared by nine editors, but although they are compared to the nine muses in the preface by J. P. von Ludewig, their names are not mentioned and the work is known by the name of the publisher. Two other editors can be identified: Longolius and Frankenstein.

Zedler says, in his preface to vol. xviii, that his *Universal Lexicon* is a work such as no time and no nation could show, and that both in its plan and execution it is more comprehensive and complete than any previous encyclopaedia. It treats not only the liberal arts and theoretic sciences, but also the practical arts and trades, not only history, geography and biography but also genealogy, topography and, beginning with vol. xviii, the lives of illustrious living persons. It has been valued as a reference work, particularly on topography, genealogy and biography.

Dennis de Coetlogon, a Frenchman naturalized in England, professed to follow a new plan of organization in *An Universal History of Arts and Sciences* (two folio volumes, 1745). To avoid repetitions the work was arranged in separate treatises for each of the sciences. Theology was divided into several treatises; philosophy into ethics, logic and metaphysics; physics into anatomy, botany, geography, geometry, etc.; military art into army, fortification, gunnery. The treatises were printed in alphabetical order. The style is diffuse, and the work was not republished.

Gianfrancesco Pivati (1689–1764), secretary of the Academy of Sciences at Venice, who had published a *Dizionario universale* (quarto, 1744), wrote a *Nuovo dizionario scientifico e curioso sacro-profano* (ten folio volumes, 597 plates, Venice, 1746–51). It is a general encyclopaedia, including geography but not history or biography. The preliminary discourse of 80 pages contains a history of the several sciences from mathematics to geography. The plates are remarkable for their number and quality.

6. French Encyclopaedists.—The French *Encyclopédie* was a literary and philosophic enterprise that had profound political, social and intellectual effects. A French translation of Ephraim Chambers' *Cyclopaedia* was begun in 1743 and finished in 1745 by John Mills, an English bank clerk, assisted by G. Sellius, a German professional translator. The translation was offered to the publisher André Le Breton, who obtained the *privilège royal* required for publication in Feb. 1745. The publication of a five-volume (four of text, one of plates) *Encyclopédie, ou Dictionnaire universel des arts et des sciences . . . traduit de l'Anglois d' Ephraim Chambers* was announced in a printed prospectus, dated 1745. After a quarrel with Mills, Le Breton had his contract with Sellius and Mills annulled, enlarged his plans to a ten-volume work (eight of text, two of plates) and sought the collaboration of three other publishers. Mills returned to England, while Sellius continued as translator of articles not only from Chambers but also from Harris' *Lexicon*. A new privilege was secured in 1746 and another in 1748 to cover changes and additions. In 1745 D'Alembert, the mathematician, and the abbé Jean Paul de Gua de Malves, the mathematical economist, were associated with the project. Early in 1746 Denis Diderot, the philosopher who was then known primarily as a translator, was added in an advisory capacity. He was working at the time with two other scholars on the translation of Robert James's *Medicinal Dictionary*, which appeared under the title *Dictionnaire universel de médecine* in 1746–48. On Oct. 16, 1747, Diderot undertook the general direction of the *Encyclopédie*, except for the mathematical parts, which were edited by D'Alembert.

Contributors were supplied with translations from Chambers, but for the most part the translations were so bad and so full of errors and omissions that they were not used. Diderot was imprisoned in Vincennes on July 29, 1749, for his *Letter on the Blind*. His prospectus, in eight folio pages, including a tabular scheme of the organization of knowledge later reproduced in the *Encyclopédie*, appeared in Nov. 1750. The first volume was published in July 1751, the second in Jan. 1752. The prospectus and the two volumes were attacked for plagiarism and irreligion by the Jesuits in their *Journal de Trévoux* and in the new edition of the *Dictionnaire de Trévoux*. After the appearance of the first volume, three Jesuit censors were appointed to examine all articles. The Jansenists joined in the attack on Abbé de Prades's article "Certitude," which had passed the censors in the second volume, but which based certitude in the analysis of knowledge after the fashion of Locke rather than in revelation. The *conseil d'état* of the king suppressed the first two volumes and ordered the manuscripts to be seized in 1752, but Chrétien de Malesherbes, director of publications of Louis XV, warned Diderot in time to save the manuscripts from the police and hid them in his own house. The editors were requested by the government to resume work since the enterprise had assumed national proportions and was considered to contribute to the honour of France. Vol. iii appeared in Oct. 1753, with an introduction by D'Alembert explaining the delay, and publication continued at the rate of a volume a year until vol. vii, completing the letter G, appeared in Nov. 1757. It contained D'Alembert's article "Geneva" praising the Protestant clergy for Socinian virtues which the Catholics did not possess and suggesting the establishment of a theatre to civilize the Genevans. The Catholic clergy resented the criticism; the Protestant clergy denied they were Socinians; Rousseau, who had contributed the articles on music to the *Encyclopédie*, argued in his *Lettre à D'Alembert sur les Spectacles* (1758) that the theatre is invariably a corrupting force and denounced Diderot. D'Alembert resigned as editor in Jan. 1758, and in Feb. 1759 the *Encyclopédie* was included in the general condemnation of dangerous books provoked by the publication of Helvétius' *De l'esprit* in 1758. His friends urged Diderot to abandon the project and even to flee. Instead he persuaded the publishers to secure permission to publish the harmless volumes of plates, while the remaining ten volumes of text were edited and printed. In 1764 Diderot had occasion, by chance, to consult his article "Saracens or Arabs" and discovered that Le Breton and his compositor had been censoring articles, after Diderot's final corrections, to remove whatever might, in their opinion, be objectionable. Since the manuscripts had been destroyed, Diderot had no way of repairing the damage or even of knowing its extent. Diderot's daughter reports in her biography of her father that Diderot had required Le Breton to print a copy with the deleted material restored. No such copy was discovered, but 318 pages of proof with Le Breton's corrections were discovered in 1933. The last 10 volumes of text were ready in 1765 and distributed in 1766; the 11 volumes of plates, of which the first was published in 1762, were completed in 1772, making a total of 28 volumes.

D'Alembert sets forth in his "Preliminary Discourse" the double

purpose that is indicated by the two parts of the title, *Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers:* it was to be an "encyclopaedia," that is, expound "the order and concatenation of the parts of human knowledge," and it was to be an "analytical dictionary of the sciences, arts and trades," that is, set forth "for each science and for each art, liberal and mechanical, the general principles which are its basis and the most essential details which are its body and substance." The "encyclopaedic tree" divides the branches of knowledge, following Bacon, according to the three faculties, memory, reason and imagination; the literary world is similarly divided into three "republics": *érudits, philosophes* and *beaux esprits* whose respective talents are memory, sagacity and pleasure (*agrément*). The encyclopaedic order or the "chain" by which it is possible to descend from the first principles of any art or science to its remote consequences or ascend from consequences to principles, is supplemented by the historical order of the acquisition of knowledge. Since the Renaissance, influence has centred successively in scholars, artists and, finally, philosophers.

D'Alembert acknowledges indebtedness to Bacon for the encyclopaedic system, but Diderot points out, doubtless to answer allegations of plagiarism, in his *Observations on the Division of the Sciences of Chancellor Bacon,* the respects in which the *Encyclopédie* departed from Bacon's organization.

D'Alembert states the purpose of the editors to provide a dictionary that could be consulted on all matters of the arts and the sciences and that would serve both to guide those who have the courage to instruct others and also to assist those who instruct only themselves. Diderot, after giving the etymology of the word in the article "Encyclopédie," goes on to say: "In fact, the purpose of an encyclopaedia is to assemble the knowledge scattered over the surface of the earth; to explain its general system to the men with whom we live, and to transmit it to the men who will come after us; in order that the labours of centuries past may not be useless for the centuries to come; that our descendants, by becoming better instructed, may as a consequence be more virtuous and happier, and that we may not die without having deserved well of the human race."

In two respects the conception and execution of the *Encyclopédie* go beyond previous encyclopaedias. One is signalized by the addition of "trades" to "arts and sciences" in the title. D'Alembert in the "Preliminary Discourse" and Diderot in the article "Art" criticize the distinction between liberal and mechanical arts, and the consequent neglect of the latter. One of their chief criticisms of Chambers was that he had read books but had not visited workshops; they do not refer to Zedler, who also has some treatment of the useful arts. Diderot himself undertook responsibility for the trades as well as philosophy, and his articles are based on observation and experience. The second advance, also indicated in the title which attributes authorship to "a society of men of letters," marks a stage of evolution. Early encyclopaedias were the work of individual authors. Later, Hübner wrote a preface to a work by other hands; Zedler depended on nine anonymous editors; the successive editors of Moréri made it eventually a co-operative product; Chambers acknowledges that his borrowings associated others with him in his work; the 1744 *Supplement* to Harris was prepared by "a society of gentlemen." The group of writers that Diderot and D'Alembert assembled about them at first were relatively unknown apart from Rousseau and Baron d'Holbach, but as the fame of the *Encyclopédie* and the attacks on it grew, distinguished and expert contributors were attracted: Turgot, Duclos, Bordeu, Boulanger in vol. iv; Voltaire, Marmontel, Forbonnais and Deleyre in vol. v; De Brosses, Saint-Lambert, Morellet, Necker, Quesnay in vol. vi. The "society of men of letters" became more than a phrase.

Three devices were used, according to D'Alembert, to relate the encyclopaedic order of arts and sciences to the alphabetical order in which the articles were printed: a table of the system of knowledge placed at the beginning of the work; an indication usually placed after each word of the art or science to which it referred; and relations established between the articles. Proper names of places (geography) and even of persons (history and mythology) are included, but not biographies of writers. There are therefore articles on Abraham (sacred history), Elizabeth (history of England) and Eleutho (mythology), but none on the ecclesiastical writers that occupied a large place in biographical dictionaries. There are no articles on the philosophers, but there are articles on the philosophic schools—"Aristotelianism" and "Peripatetic Philosophy," "Academics," "Skeptics," "Epicurianism," "Pyrrhonian Philosophy," "Machiavellianism," "Socratic Philosophy." The greater part of the articles, however, are on the arts and sciences and on important terms or distinctions in them. The contributors to the *Encyclopédie* did not expound a single philosophy; the Baconian scheme of the sciences was adjusted to Locke's philosophy and to Newton's science, and the Lockean analysis of human understanding was adjusted to the Epicurean atomism of Pierre Gassendi and to Bayle's skepticism. The articles on philosophy and religion, particularly after Le Breton's deletions, seem neither original nor shocking. The philosophy of the *Encyclopédie* was an attitude of tolerance and liberalism and of concern with the practical; it had widespread and revolutionary effects as an expression of attitude of mind and of perspective.

Charles Joseph Panckoucke (1736–98) had proposed to issue, in collaboration with other publishers, a new and revised edition about 1761 and had bought the plates of the first seven volumes. When Diderot refused to edit the new edition, a reissue of the original volumes was begun. Three volumes were published and were seized by the government in 1770. When the last of the original publishers died in 1775, Panckoucke published his *Supplement* to the *Encyclopédie* in four volumes of text and one of plates (Paris, 1776–77); he later added two volumes of index, prepared by Pierre Mouchon (Amsterdam, 1780). These 7 volumes plus the 28 prepared under the supervision of Diderot constitute the first edition of the *Encyclopédie* in 35 folio volumes, 1751–80.

Several editions of the *Encyclopédie,* some expurgated and some enlarged with supplements and additions, appeared in Italy and Switzerland before 1782. In 1780 Panckoucke secured the *privilège* to issue a new and enlarged edition of the *Encyclopédie* arranged as a system of separate dictionaries under the title *Encyclopédie méthodique ou par ordre de matières.* The articles of the *Encyclopédie,* rearranged according to the sciences and arts to which they were assigned in the original schematism, were the basis of this longer project. There were to be 26 parts treated in distinct dictionaries entrusted to different editors, and completed by a *vocabulaire universel* with references to all the places in which each word occurred. Publication started in 1782 and the work was to be finished in 1787. In 1788 the plans were revised, including a new division of knowledge into 44 parts and 51 dictionaries to be published in about 124 volumes. The work continued through the Revolution and was completed in 1832, with the appearance of vol. 166, 50 years after the appearance of the first volume. The largest dictionaries are medicine, 13 volumes, 10,330 pages; zoology, 7 dictionaries, 13,645 pages, 1,206 plates; botany, 12,002 pages, 1,000 plates: geography, 3 dictionaries and 2 atlases, 9,090 pages, 193 maps and plates.

(For *Encyclopædia Britannica,* first published in Edinburgh in three volumes, 1768–71, *see* below.)

C. 19th Century

During the 19th century the development of encyclopaedias was marked by three tendencies: (1) combination of articles on arts and sciences with articles on biography, geography and history (although separate gazetteers and dictionaries of biography and history continued to appear); (2) development of simple short-entry encyclopaedias, of systematic long-entry encyclopaedias and of encyclopaedias that combined long articles about subjects with short articles about terms; and (3) extension of the alphabetical arrangement of articles balanced by experiments in systematic or topical arrangements.

1. "Conversations-Lexicon" and the Learned Encyclopaedia.—In 1796 Gotthelf Renatus Löbel (1767–99) began publication of an encyclopaedia for women under the title *Frauenzimmer-Lexicon.* Designed to take the place of Hübner's, it included history and biography as well as natural history, art, science and

the political conditions of the world. It was continued after Löbel's death: vol. 1 through 4 ("A–R") appeared from 1796 to 1800 and vol. 5 in 1806. In 1808 the publisher Friedrich Arnold Brockhaus (*q.v.*; 1772–1823) bought the copyright and issued the sixth and final volume. Edition succeeded edition, and the number of volumes increased. The fourth edition returned to Hübner's combination of *real* and *conversation* in the title: *Allgemeine Deutsche Real Encyclopädie für die gebildeten Stände: Conversations-Lexikon.* The title was changed to *Brockhaus' Konversations-Lexikon* in the 13th edition (see *20th Century* below).

Brockhaus gives information on all kinds of subjects in simple popular form without extended details. The articles are extremely brief, but they provide references; biographies of living men are included and articles on German subjects are numerous. The influence of *Brockhaus* proved extensive. It provided, in translation or adaptation, a core for encyclopaedias in many languages about which articles of particular national interest were assembled. Many national encyclopaedias influenced by *Brockhaus* to a greater or less extent appeared in the 19th century: Danish (*Dansk Conversations-Lexikon* of H. A. Kofod in 28 volumes, 1816–28; *Dansk Konversations-Lexikon* of P. Larsen in 8 volumes, 1849–60, second edition, 6 volumes, 1870–73); Swedish (*Conversations-Lexikon,* 4 volumes, 1821–26, 4 supplementary volumes, 1831–39; *Svenskt Konversations-Lexikon* of P. G. Berg, 4 volumes, 1845–52; Bonnier's *Konversations och Universal-Lexikon,* 4 volumes, 1856–66); Norwegian and Danish (*Nordisk Conversations-Lexicon,* 5 volumes, 1858–63); Dutch (*Algemeene Nederlandsche Encyclopedie,* 5 volumes, 1865–68); Hungarian (*Ismerettar,* 10 volumes, 1859–64); Russian (*Entsiklopedecheski Lexikon* begun by Gresch in 1835, abandoned in the 17th volume and the fourth letter; later *Entsiklopedecheski Slovar,* containing articles by the best Russian men of science and letters, in 41 volumes and 2 supplements published from 1890, and an interrupted revised edition, vol. 1–29, "A–Otto," 1912–17?); French (*Dictionnaire de la conversation et de la lecture,* 68 volumes, 1833–51, second edition, 16 volumes, 1853–58, 5 supplementary volumes, 1865–82; *Encyclopédie des gens du monde,* 22 volumes, 1833–45; and *Encyclopédie moderne,* 26 volumes, 1823–32); at a further remove, Italian (*Enciclopedia italiana,* which formed the basis of *Nuova Enciclopedia italiana* published in 14 volumes in 1841–51, revised in 25 volumes in 1875–88; and *Enciclopedia moderna,* 13 volumes and 2 volumes of plates, 1838–53); and Spanish (*Enciclopedia moderna,* 34 volumes and 3 volumes of plates, 1851–53).

The Encyclopedia Americana (13 volumes, 1829–33) was edited by Francis Lieber and based on, and in part translated from, the seventh edition of *Brockhaus.* It went through a number of editions down to 1858 and added, in the last three editions, a supplementary 14th volume. A new work under the same title but continuing some of the old material was published in 1903–04 in 16 volumes. The translation used by Lieber was reprinted at Glasgow in seven volumes in 1841 and revised in 1862, with some additional introductory matter, under the title *The Popular Encyclopaedia; or, "Conversations Lexicon."*

Chambers's Encyclopaedia (ten volumes, 1859–68; named after Robert and William Chambers, the publishers, and not related to Ephraim Chambers' *Cyclopaedia*) was based on a translation of the tenth edition of *Brockhaus.* After several reissues and revisions, a new edition, in ten volumes, was published in 1888–92, followed by other later editions. *The International Cyclopaedia,* based partly on *Chambers's Encyclopaedia* of 1878–80, was first published in 1884 and revised in 1891, 1894 and 1898. It was superseded in 1902 by *The New International Encyclopaedia* in 17 volumes.

Meyers Grosses Konversations-Lexikon was first published in 46 volumes in 1840–52. The fourth edition in 17 volumes appeared in 1885–90, the fifth in 18 volumes in 1894–98. The articles are brief; like *Brockhaus, Meyers* is profusely illustrated with maps and plates. *Meyers* was supplemented by an annual. *Herders Konversations-Lexikon,* a Catholic encyclopaedia, was first published in five volumes in 1853–57.

At the other extreme from the short unsigned articles of the *Konversations-Lexikon* are the detailed signed articles of Ersch and Gruber's *Allegmeine Enzyklopädie der Wissenschaften und Künste* (167 volumes, 1818–79). The work was designed and begun in 1813 by Johann Samuel Ersch, and the successive editors of the various sections were men of learning, including G. Hufeland, J. G. Gruber, M. H. E. Meier, Hermann Brockhaus, W. Müller and A. G. Hoffmann. The work was divided into three sections: (1) "A–G," 99 volumes, 1818–79; (2) "H–N," 43 volumes, 1827–55; (3) "O–Z," of which 25 volumes, "O-Phy," appeared. The last and some of the intermediate parts of the alphabet are lacking. The articles are signed and the contributors are among the most learned German scholars of the period. The longest article is "Griechenland," vol. 80 through 87, 3,668 pages. "Gross Britannien" is treated in 700 pages, "Indien" in 356. Many of the articles are based on original research and rank among the best authorities on their respective subjects.

2. England and the United States.—In the development of English encyclopaedias the tradition of brief articles based on *Brockhaus* was balanced by the tradition of *Encyclopædia Britannica* (*see* below), which combined brief with long systematic articles. Several encyclopaedias appeared early in the 19th century before the Conversations-Lexicons: *English Encyclopaedia* (10 volumes, 1802); Brewster's *Edinburgh Encyclopaedia* (1810–30); *Encyclopaedia Perthensis* (second edition, 24 volumes, 1816); *Encyclopaedia Edinensis* (6 volumes, 1827); J. Wilkes' *Encyclopaedia Londinensis* (24 volumes, 1810–29); *London Encyclopaedia* (22 volumes, 1829); and Rees's *Cyclopaedia,* based on Ephraim Chambers' (45 volumes, 1819–20).

The *Encyclopaedia Metropolitana* (25 volumes, 1817–45, 3 of plates and an index volume), on the other hand, represented a determined attempt to return to the systematic organization of encyclopaedias, expounding the sciences and systematic arts "entire and in their natural sequence." In the "General Introduction or Preliminary Treatise on Method," S. T. Coleridge expresses the conviction that because of disregard of method no encyclopaedia has ever been methodically arranged. Method, as conceived by Coleridge, is concerned with relations discerned by the mind among things—relations of law in the pure and mixed sciences, of theory in the scientific arts (such as medicine, chemistry and physiology), while in the fine arts, method lies between law and theory. Coleridge combined Plato and Francis Bacon in his organization of the sciences, and argued that the method of Bacon is identical with Plato's dialectical method. The only difference between them is that "the one more especially cultivated Natural Philosophy, the other Metaphysics. Plato treated principally of Truth, as manifested in the world of Intellect; Bacon, of the same truth as manifested in the world of sense." The organization is in four parts, the last only being alphabetical: (1) pure sciences in 2 volumes and 28 treatises; (2) mixed and applied sciences in 6 volumes and 42 treatises; (3) history and biography in 5 volumes; and (4) miscellaneous, including geography, a dictionary of English (the first form of Richardson's) and descriptive natural history in 12 volumes. The authors of the treatises included Archbishop Richard Whately, Sir John Herschel, Peter Barlow, George Peacock and Augustus de Morgan. A second edition was published in 40 volumes in 1848–58.

The *British Cyclopaedia* (ten volumes, 1835–38) of C. F. Partington is likewise distributed into four divisions: (1) arts and sciences; (2) natural history; (3) biography; and (4) literature, history and geography. The arrangement within each division is alphabetical. A third "divisional" encyclopaedia, the excellent and useful *English Cyclopaedia* (23 volumes, 1854–62, and 4 volumes of supplements, 1869–73) of Charles Knight is based on the *Penny Cyclopaedia* (29 volumes, 1833–46). It is in four parts, each alphabetically ordered: (1) geography, (2) natural history, (3) biography (with 703 lives of living persons) and (4) arts and sciences. The synoptic index brings the four parts together alphabetically in four columns, one for each division.

The *New American Cyclopaedia* (16 volumes, 1858–63) was a popular dictionary of general knowledge, edited by George Ripley and compiled by 364 contributors, chiefly from the United States. The *American Annual Encyclopaedia,* a volume of about 800 pages, was issued yearly from 1861 to 1874 and was continued as *Appleton's Annual Cyclopaedia* until 1902. A new edition of the *Ameri-*

can Cyclopaedia (16 volumes, 1873–76) was prepared by the same editors. A. J. Johnson's *New Universal Cyclopaedia* first appeared in four volumes (1875–77); it was expanded in an excellently planned new edition in eight volumes (1893–95), under the title *Johnson's Universal Cyclopaedia*. It was reissued under the editorship of Charles K. Adams as *Appleton's Universal Cyclopaedia* (12 volumes, 1902).

3. France.—The *Encyclopédie méthodique*, completed in 1832, was an extensive application of the systematic organization. In the 20th century, French encyclopaedias were to return to experiments in systematic organization, and the Conversations-Lexicons of short articles had some vogue at the beginning of the 19th century; but French encyclopaedias took three characteristic lines of development during the 19th century.

Diderot's interest in the details of the useful arts was continued in the *Encyclopédie moderne, Dictionnaire abrégé des lettres, des arts, de l'industrie, de l'agriculture et du commerce*, first published in 26 volumes, 1823–32. A new edition was prepared by L. Renier in 27 volumes and 3 volumes of plates (1846–51); supplements were edited by N. des Vergers and L. Renier in 12 volumes, 1856–62.

The *Grand Dictionnaire universel du XIX^e siècle* (15 volumes, 1866–76, and 2 supplements, 1878 and 1888) of Pierre Larousse is a French encyclopaedia of "vulgarization." Combining dictionary with encyclopaedia, it has many entries on small subjects, including articles on individual works of literature, entered under their titles, and many minor biographies. It is particularly useful for European literature, biography and history. The *Nouveau Larousse illustré* (seven volumes and supplement, 1896–1904) was a new work, more popular in character and profusely illustrated.

The *Grande Encyclopédie, inventaire raisonné des sciences, des lettres et des arts* (31 volumes, 1886–1902), like Diderot's *Encyclopédie*, was the work of "a society of scholars and men of letters." Alphabetical in organization and strongly historical in orientation, it is one of the best encyclopaedias in any language, and has authoritative signed articles, excellent bibliographies and many entries on small subjects. (RD. MCK.)

D. 20TH CENTURY

The appearance of the entire 11th edition of *Encyclopædia Britannica* within six months (1910–11) ended one era of encyclopaedia publishing. A few years later World War I hurled the 20th century into a quickened pace, which by mid-century had outdistanced the learning of all previous time and continued to race toward an incalculable accumulation of knowledge. Following World War II, improved communication and advanced techniques of book design, production and distribution combined with an appetite for information and an urgent world-wide need for education to create a revolution in encyclopaedia production.

Along with an increasing number of scholarly compendiums of knowledge, an apparently insatiable buying public consumed innumerable generalized encyclopaedias available in supermarkets or in kiosks, short encyclopaedias of one to five volumes, specialized encyclopaedias or encyclopaedic series dealing with one discipline or one nation and encyclopaedic dictionaries. New concepts of education and of the needs of children evoked encyclopaedias for the young in most major languages. Co-editing by leading publishers in a number of countries led to the production of the same encyclopaedia in as many as ten native languages.

1. England and the United States.—Among the ideals evolved for encyclopaedias—which include being comprehensive, authoritative, accurate, clear in organization and provided with adequate bibliographies—the requirement of being up to date first reflected the change. Several devices have been used. Two loose-leaf sets in English, *Nelson's Perpetual Loose-Leaf Encyclopaedia* (12 volumes, 1909) and *Winston's Cumulative Loose-Leaf Encyclopedia* (10 volumes, 1912–42), unable to supply adequate revisions, were discontinued. The *Catholic Encyclopedia* (15 volumes and index, 1907–14) used loose-leaf format for supplements to bring some articles up to date. Because of the availability of moderately priced private binding, some European publishers at mid-century revived the use of loose-leaf format and soft-bound or fascicule issues, but in English-speaking countries, frequent editions with periodical supplements or annual printings partially revised, augmented by yearbooks, have proven more satisfactory.

The *New International Encyclopædia* (second edition, 23 volumes, reissued in 1922, supplements 1924, 1930) included good bibliographies and many U.S. and Latin-American biographies still useful for that period. *Chambers's Encyclopædia* was issued in new editions of 15 volumes (1955, 1959). The *Grolier Encyclopedia*, revised and issued annually (10 volumes after 1944, 20 volumes in 10 after 1957), was discontinued after 1963 and replaced by *Grolier's Universal Encyclopedia* (10 volumes, 5,000,000 words) in 1965. *The Encyclopedia Americana* (28,500,000 words) developed from a 20-volume set (1911, 2 supplements, 1912) to a 30-volume edition (1918–20), issued annually after 1936, the year in which *Encyclopædia Britannica* (36,500,000 words) began continuous revision (*see* below). *The American Peoples Encyclopedia* (8,500,000 words) was published and issued annually in 20 volumes from 1948. *Collier's Encyclopedia* appeared in 20 volumes from 1950–61 and in 24 volumes (22,000,000 words) thereafter. *Encyclopedia International* was produced annually (20 volumes, 9,500,000 words) from 1963. Among current short encyclopaedias, one of the oldest is *Pears' Cyclopaedia* (1 volume). The 72nd edition appeared in England in 1964. Nonalphabetical, *The Lincoln Library of Essential Information* has an alphabetical index. The 28th edition (1 or 2 volumes) appeared in 1964. *The Columbia Encyclopedia* (1 volume, first edition 1935 with periodical supplements; third edition, 7,500,000 words, 1963) contains about 70,000 brief entries, useful for "spot" information.

The Children's Encyclopaedia, issued in England in 1908, was published in the United States as *The Book of Knowledge* (24 volumes, 1910; 20 volumes, 4,000,000 words) from 1912 and in France as *Encyclopédie de la jeunesse*. *Compton's Pictured Encyclopedia*, issued in 1922 (15 volumes, 7,700,000 words), now annually, is a school reference set that also provides information for parents on education and family life. Each volume contains a "Fact-Index" that adds short articles on subjects not treated in the alphabetically arranged text. A Spanish edition, periodically revised, was first published in 1957 as *Enciclopedia Cultural*. The *World Book Encyclopedia* (1917–18) was revised annually from 1925, with 20 volumes (8,500,000 words) after 1960. Written for high-school students, it also is produced in Braille and in a large-type edition. *Britannica Junior*, a 1934 reissue of *Weedon's Modern Encyclopedia*, was completely revised in 1947 for elementary-school use in 15 volumes (5,000,000 words) including a "Ready Reference Index" volume with short entries and definitions. It was retitled *Britannica Junior Encyclopedia* with the annual revision of 1963. The *Children's Britannica* (12 volumes) appeared in England in 1960. The *Oxford Junior Encyclopaedia* (12 volumes, 1948–54; index, 1956) is arranged topically, but is alphabetical within each volume. An American edition was produced in 1964. *The American Educator Encyclopedia* was enlarged to 14 volumes (5,000,000 words) in 1964. Two regional encyclopaedias in 3 volumes appeared in 1961: a new edition of the *Australian Junior Encyclopaedia* and the *New Zealand Junior Encyclopaedia*.

2. Encyclopaedia Annuals.—*The New International Year Book* was published with one interruption from 1899, for many years as a supplement to the encyclopaedia, continuing after it was defunct. Yearbooks have proliferated in all languages since World War II, for adults and children, as adjuncts to encyclopaedias, as independent publications and in specialized areas, until they have become too numerous to list.

The general characteristics are feature articles on pressing issues, a calendar of major events, a necrology, and updating articles on geographical areas and major fields of current interest.

3. France.—*Larousse du XX^e siècle*, edited by Paul Augé, appeared in 6 volumes (1928–33; revised, 1948–50; supplement, 1954). Among other Larousse publications are *Grand memento encyclopédique* (2 volumes, 1936–37, revised and retitled *Encyclopédie Larousse méthodique*, 1955) in systematic organization; *Nouveau Larousse universel* (2 volumes, 1948–49); and *Petit Larousse illustré* (in numerous editions [new edition, as *Petit Larousse, 1959*, issued 1960 and annually] and in Spanish translá-

tion), remarkably complete in brief biographies and place names: *Grand Larousse* (10 volumes, 1958–63) in alphabetical arrangement; and *Encyclopédie Larousse pour la jeunesse* (1958–62), a topical work in 5 volumes for secondary-school children. The *Dictionnaire encyclopédique Quillet* (6 volumes, 1934; new edition, 1962) includes some systematic expositions.

The *Encyclopédie française*, organized according to interrelated problems, appeared in loose-leaf (1935–64) in 20 volumes, of which some departed from the original framework, and others were rewritten, delaying publication of a much needed index. Supplementary pamphlets have brought some material up to date. *Clartés, encyclopédie pratique*, also loose-leaf, was designed for secondary-school use. Sixteen of a projected 18 volumes appeared from 1949, with frequently issued updating inserts and a quarterly journal on current nonencyclopaedic subjects. The *Encyclopédie de la Pléiade*, edited by R. Queneau and planned for 40 volumes in several series, appeared slowly from 1955, and was not quite half completed in a decade. Of little value as a quick reference work, it presents profound analyses of the disciplines so far covered.

Encyclopédie internationale Focus, a 5-volume alphabetical work, began to appear in 1963. Adapted from the Swedish *Focus uppslagsbok* (1959), it is one of many encyclopaedias being published in several languages. The Finnish version, *Fokus*, appeared 1963–65.

4. Italy.—The *Enciclopedia italiana di scienze, lettere ed arti* (36 volumes, 1929–39, 5 appendices covering the years 1937–60). One of the great encyclopaedias, it includes excellent long articles, profusely illustrated both in the text and with sepia and coloured plates. Major articles are signed. A Fascist bias is prominent in the political articles, but does not intrude in most other articles. A 12-volume *Dizionario enciclopedico italiano* also published by the Istituto della Enciclopedia italiana (1955–61) offers many concise encyclopaedic articles and very fine illustrations. The third edition of the *Enciclopedia Motta* (12 volumes, alphabetical) was completed in 1963. The *Nuovissima enciclopedia pratica Bompiani* (5 volumes in topical arrangement) appeared in its 20th edition in 1962.

Children's encyclopaedias, probably more numerous in Italy than in any other country, include: the *Enciclopedia dei ragazzi* (33rd edition, 10 volumes, 1962), *Il tesoro del ragazzo italiano* (third edition, 1960–), *A-Z Panorama* (9 volumes, topical, 1957–63) with many original applications of graphic representation and organization of material, as well as translations of American sets of minor value. The *Enciclopedia cattolica* was published by the Vatican in 12 volumes, 1948–54.

5. Germany.—The present title of *Brockhaus* appeared with the 15th edition (20 volumes and supplement), *Der Grosse Brockhaus* (1928–35). *Der Kleine Brockhaus* was revised as *Der Neue Brockhaus* (4 volumes and atlas, 1936–38; completely revised third edition, 5 volumes and atlas, 1958–60; supplement, 1964). *Meyers Grosses Konversations-Lexikon* appeared in a sixth edition (24 volumes, 1902–13) and a seventh edition (17 volumes, 1924–35). This edition is still found in many libraries, valued for its concise articles and for information difficult to find elsewhere. An eighth edition, highly coloured by Nazi influence, was abandoned at the ninth volume in 1942.

New editions of German encyclopaedias, which not only brought them up to date but also removed traces of Nazi orientation, appeared shortly after World War II. *Der Grosse Brockhaus* (12 volumes of a 16th edition, 1953–55, an atlas, 1960, and supplements in 1958 and 1964) retained its position as a reliable reference work. *Der Grosse Herder* (10 volumes, 1952–56) was followed by a revised fifth edition in 12 volumes, 1956–62. The Meyer firm was liquidated in 1945. The old name was revived for *Meyers Neues Lexikon*, the first German Socialist dictionary (8 volumes, 1961–64). The great Zedler encyclopaedia (*see* above) was reprinted, unchanged, in 1959.

6. Austria.—The second edition of the *Neues Welt-Lexikon* appeared in two volumes, 1948–49.

7. U.S.S.R.—The first edition of the "Great Soviet Encyclopaedia," *Bol'shaia sovetskaia entsiklopediia*, was completed in 65 volumes over a period of 20 years. The first volume appeared in 1926 and succeeding volumes irregularly with vol. 52 and 55, published in 1947, completing the set. A historical supplement on the U.S.S.R. appeared in 1947. The articles are uneven in quality and there were many changes of editors. The officially approved manner of interpreting history was changed abruptly in 1932. The basic articles are, on the whole, satisfactory; bibliographies are fairly full; maps and illustrations are fairly abundant and usually good. Articles on non-Communist countries, personalities and movements, however, are distorted. By the time it was completed, it no longer had official approval.

A second edition was begun in 1949 under the editorship of S. I. Vavilov and (after vol. 8) completed by B. A. Vvedensky. Vol. 46, 47 and 48 appeared in 1957, completing the alphabet except for vol. 40, which was issued finally in 1958. It contained a five-page article and a full-page photograph of Stalin. The article reflected Khrushchev's criticism of Stalin's errors but undertook to give him credit where it was due. Vol. 50 is a special volume on the U.S.S.R., translated into English in the United States in 1962. Vol. 51 is a general supplement to the second edition, containing a number of political biographies, tables of weights and measures, currencies, foreign words and phrases, and an author and general index. A two-volume author-and-title index was added. An annual supplement was published from 1957.

The purpose of the encyclopaedia is set forth in the preface to the first volume in the words of the 1949 decree of the council of ministers of the Soviet Union:

> The second edition of the Great Soviet Encyclopaedia should widely elucidate the world-historical victories of socialism in our country, the achievement of the U.S.S.R. in the fields of economics, science, culture and art. The superiority of socialist culture over the culture of the capitalist world must be shown with exhaustive completeness. Operating on the Marxist-Leninist theory, the encyclopaedia should give a party criticism of contemporary bourgeois tendencies in the various provinces of science and techniques.

There was more "slanting" in the second edition than in the first; the bibliographies paid scant attention to foreign achievements unless they were favourable to the U.S.S.R. The printing, paper and binding of the second edition were better than those of the first; the illustrations and maps were abundant and frequently good. The third edition of a short version, the *Malaia sovetskaia entsiklopediia* (10 volumes and a supplement) appeared in 1958–60. A 10-volume topical encyclopaedia for children, *Detskaya entsiklopedia*, was published in 1961–62.

8. Slavic and Baltic Encyclopaedias.—A Yugoslav encyclopaedia, the *Enciklopedija Jugoslavije*, was begun in 1955, planned for eight volumes. A Serbo-Croatian encyclopaedia, the *Enciklopedija Leksikografskog zavoda*, was begun in the same year, to be completed in six volumes. The *Lietuvių Enciklopedija*, begun in Lithuania before World War II, was continued in the United States, with 30 volumes (A-Tu) appearing between 1953 and 1964. Vol. 15, a special volume on Lithuania, had not yet been completed.

9. Hungary.—The *Révai Nagy Lexikona: az Ismeretek Enciklopédiája*, 20 volumes, was published 1911–27, with a supplement in 1935, and in a short 2-volume version, 1947–48. A 6-volume encyclopaedic dictionary, *Új magyar lexikon*, designed to give a Marxist-Leninist selection and interpretation of facts, was published in 1959–62.

10. Finland.—*Iso tietosanakirja* was published in 15 volumes, 1931–39, and the second edition in 10 volumes, 1959–65.

11. The Netherlands.—*Winkler Prins' Geïllustreerde encyclopaedie*, an excellent short-article encyclopaedia originally based on *Brockhaus*, was completely revised (sixth edition, 18 volumes 1947–54) with a supplementary and index volume in 1955, followed by another supplement in 1960 and a seventh edition, titled *Algemene Winkler Prins encyclopaedie*, in 1956–60. The *Eerste nederlandse systematisch ingerichte encyclopaedie* (ENSIE) appeared in 10 volumes, 1946–52, and a supplement, 1959–60. The last volume of this systematic work contains a dictionary and index with additional material not found in the text.

12. Scandinavia.—The *Svensk Uppslagsbok*, beautifully printed and bound, with excellent long articles, appeared in a second edition in 32 volumes, 1947–55. It has good maps, city plans

and portraits and biographical material not easily accessible elsewhere. *Nortstedts Uppslagsbok*, a 1-volume encyclopaedia, was issued in a fifth edition in 1952. An earlier edition was used as the basis for the Danish and Norwegian *Gyldendals* 1-volume encyclopaedias, which were in turn frequently revised and issued in new editions. *Bonniers folklexikon* (5 volumes, 1952–55) followed *Bonniers Konversations lexikon* (second edition, 14 volumes and supplement, 1937–50), the leading Swedish encyclopaedia prior to *Svensk Uppslagsbok*. The Norwegian *Aschehougs Konversations-Leksikon* (first published as the *Illustreret Norsk Konversations-Leksikon* in 6 volumes, 1907–13) was reedited in a fourth edition of 18 volumes, 1954–61. The *Norsk Allkunnebok*, with short, signed articles, the first encyclopaedia written entirely in the *nynorsk* (*landsmål*) language, was published in 1948–61, with 10 volumes and an atlas.

In Danish, *Salmonsens Konversations Leksikon* (19 volumes, 1891–1911) was published in a second edition in 26 volumes, 1915–30, and a third edition was begun in 1949. Two shorter versions followed, one in two volumes and *Den lille Salmonsen* (12 volumes, 1937–40) which has been supplemented since 1941 by a well-indexed monthly. *Raunkjaers Konversations Leksikon* was published in 12 volumes and supplement, 1948–57.

13. Switzerland.—The *Schweizer Lexikon* (7 volumes, 1945–48) on good paper and nicely bound, with clear, short articles, was one of the first new encyclopaedias to appear after World War II. It has been revised, in a 2-volume edition (1949–50) and in other editions, variously titled, in 1952, 1954 and 1958.

14. Spanish- and Portuguese-Language Encyclopaedias.—The *Enciclopedia universal ilustrada europeo-americana* (70 volumes and a 10-volume supplement, 1905–33, with a *Suplemento anual* published at irregular intervals from 1934, usually about four years behind the date of publication) is an outstanding reference work. Its articles, bibliographies, maps and illustrations are good; it is particularly useful for Spanish and Spanish-American biography. The fifth edition of a Spanish version of *Herder* appeared in 1963. Many of the Larousse encyclopaedic dictionaries were adapted for Hispano-American readers during the 1950s and frequently revised. *L'enciclopedia methodica Larousse* was issued in 3 volumes in 1964. The *Diccionario Enciclopédico Salvat* (12 volumes, 1955; 8 volumes of an 11th edition in 1964) is useful for Latin-American material. Salvat's *Universitas* (seventh edition, 20 volumes and index, 1964) is a topical encyclopaedia in attractive format, designed for secondary school reference. The 8-volume German topical *Bok-förlaget Fakta A.B.* was issued as *Facta* in 1964. *Enciclopedia Barsa* was published in Spanish in 15 volumes by *Encyclopædia Britannica* in 1957 and continually revised thereafter, with an added volume in 1965. It is a general encyclopaedia emphasizing interests of South American readers. *Enciclopédia Barsa* in Portuguese appeared in 1964 in 16 volumes to fill a similar need. Yearbooks were published for each of the *Barsas*, from 1960 and 1965 respectively.

The *Diccionario Enciclopédico U.T.E.H.A.* in 10 volumes (1950–52, 2-volume supplement, 1965) stresses Mexico and Latin America. The *Grande Enciclopédia Portuguesa e Brasileira* began to appear in 1935 and was completed in the 37th volume in 1958, including a general index that ran through the 40th volume.

15. Greece.—The "Great Greek Encyclopaedia," *Megale hellenike enkyklopaideia* (24 volumes, 1926–34) was published in a second edition in 1958 with a 6-volume supplement. The *Eleutheroudake Enkyklopaidikon leksikon*, 1927–31, constituted 12 volumes. The modern Greek *Neōteron enkyklopaidikon lexikon* (18 volumes) began to appear in a new expanded edition in 1957.

16. Turkey.—The *Turk Ansiklopedisi* started publication in 1946 under the title *Inonu Ansiklopedisi*, producing 10 volumes in 15 years. It promises to be a very large work; the first 8 volumes completed the letter B. General in scope, it places emphasis on Turkish and Islamic subjects.

17. Indonesia.—The *Ensiklopedia Indonesia* appeared in 1954–57 in 3 volumes. It is a general encyclopaedia but emphasizes Asian affairs. It contains a 15-page time chart showing parallel developments of civilization from prehistoric times to 1955 in Indonesia, Asia Minor, China, Japan, Africa, Europe and America.

18. Japan.—The "Great Encyclopaedia," *Dai-hykka Jiten*, was published in 28 volumes, 1931–35, and supplement, 1939. The *Sekai dai hyakkajiten* (32 volumes, 1955–59), with supplementary volumes planned, is especially important for Oriental subjects. *Gendai kyoyo koza* (7 volumes, 1956–57) is strong in coverage of history and biography.

19. Canada.—*The Encyclopedia of Canada* (6 volumes, 1935–36, reprinted in 1948, with a Newfoundland supplement in 1949) covers Canadian history, biography and geography. It was succeeded by the Grolier society's *Encyclopedia Canadiana* (10 volumes, 1957–58, new edition, 1960).

French-Canadian encyclopaedias include the *Encyclopédie Grolier* (10 volumes), the *Larousse canadien-complet* (2 volumes) and the *Encyclopédie du Canada français*—(6 volumes), all of which appeared in the 1950s.

20. Australia.—*The Australian Encyclopaedia* (2 volumes, 1952–57) was expanded and issued in a new 10-volume edition in 1958. Essentially a national work, it stresses distinctive features of the country, particularly natural history and geography, the aborigines and other matters of special interest in Australia.

21. Specialized Encyclopaedias.—*Jewish:* The *Encyclopaedia Hebraica* (*Entziklopedya Ivrit*) was published in Hebrew in 18 volumes, 1948–64. Earlier general Jewish encyclopaedias are *The Jewish Encyclopaedia: a Descriptive Record of the History, Religion, Literature and Customs of the Jewish People* (12 volumes, 1901–06); *The Universal Jewish Encyclopedia* (10 volumes, 1939–43); and the *Encyclopaedia Judaica, das Judentum in Geschichte und Gegenwart* (10 volumes, 1928–34, in German and Hebrew editions and a revised edition begun in 1963). *The Standard Jewish Encyclopedia* (1959; revised, 1962) is a useful 1-volume source.

Muslim.—The *Enzyklopädie des Islam* appeared in 4 volumes, 1908–34; an English edition, *Encyclopaedia of Islam: a Dictionary of the Geography, Ethnography and Biography of the Muhammadan Peoples* (full title), appeared in 4 volumes and a supplement, 1911–38. A second English and French edition reached the fifth volume (1954–64) and a Turkish edition under the title *Islam ansiklopedisi* appeared in 6 volumes (1940–57). A *Shorter Encyclopaedia of Islam* was published in 1 volume in 1953. An Islamic encyclopaedia in Urdu, *Urdu Dairā'ira Ma'arif-i-Islamaiy*, planned in 20 volumes, began publication in 1954, and was particularly useful for geographical and biographical information.

Other Specialized Encyclopaedias.—Such encyclopaedias as those named above are to be distinguished from those that are truly specialized in the sense that they deal with subjects of particular interest but present the "whole circle of knowledge" within their stated confines. These include works devoted to a major geographical area or to a single nation, or exhaustive studies of a single discipline or subdiscipline; for example, an *Encyclopaedia Africana* being planned in Ghana in English and for translation later into French; *Tutt'Italia*, appearing in weekly magazine format, to be privately bound or purchased in bound form when completed; and the *McGraw-Hill Encyclopedia of Science and Technology* (15 volumes, 1960); and *The Encyclopedia of Mental Health* (6 volumes, 1963).

Mention may also be made of the large-scale national biographies, of which the *Dictionary of National Biography*, a British work begun in the latter part of the 19th century, is the prototype. (*See* also BIOGRAPHY; *The 19th Century.*)

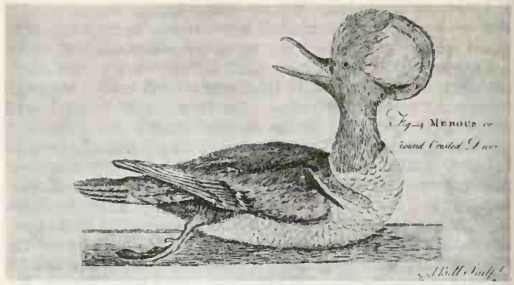
Well after the mid-century mark, the impact of two world wars was reflected in such ways as the publication in western countries of national encyclopaedias like *Ukraine*, a revised condensation of the *Entsyklopediia ukrainoznavstva*, of which vol. 1 was issued in Canada in 1963; and by new works attempting to orient readers in a new age, such as *Epoche, Atom und Automation; Entzyklopädie des technischen Jahrhunderts* (an encyclopaedia of the atomic age), a 10-volume work on recent advances in technology and their social implications, published in quick succession in French, German and Italian, from 1959. In the 1960s, the need for knowledge seemed to be divided between an appetite for facts and a need to synthesize basic concepts, and encyclopaedists around the world attempted to fill both needs. (X.)

III. ENCYCLOPÆDIA BRITANNICA

1. First Edition.—The *Encyclopædia Britannica or Dictionary of Arts and Sciences*, "by a society of gentlemen in Scotland, printed in Edinburgh for A. Bell and C. Macfarquhar, and sold by Colin Macfarquhar at his printing-office in Nicolson-street," was completed in 1771 in 3 volumes containing 2,459 pages (actually numbered as 2,659) and 160 copperplates engraved by Andrew Bell. The first two numbers were issued in Dec. 1768, "price 6*d*. each, or 8*d*. on a finer paper," and was to be completed in 100 weekly numbers. It was compiled, as the title page says, on a new plan. The different sciences and arts were "digested into distinct treatises or systems," of which there are 45 with crossheadings (*i.e.*, titles printed across the page) and about 30 other articles more than three pages long. The longest are "Anatomy," 165 pages, and "Surgery," 238 pages. "The various technical terms, etc., are explained as they occur in the order of the alphabet." "Instead of dismembering the sciences, by attempting to treat them intelligibly under a multitude of technical terms, they have digested the principles of every science in the form of systems or distinct treatises, and explained the terms as they occur in the order of the alphabet, with references to the sciences to which they belong." This plan, as the compilers say, differs from that of all the previous dictionaries of arts and sciences. Its merit and novelty consist in the combination of De Coetlogon's plan with that in common use—on the one hand keeping important subjects together, and on the other facilitating reference by numerous separate articles. It is doubtful to whom the credit of this plan is attributable. The editor, William Smellie, a printer (1740–95), afterward secretary and superintendent of natural history to the Society of Scottish Antiquaries, is said by his biographer to have devised the plan and written or compiled all the chief articles. Archibald Constable, who was interested in the work from 1788 and was afterward intimately acquainted with Bell, says Macfarquhar was the actual projector of the encyclopaedia and the editor of the first two editions, while Smellie was merely "a contributor for hire." George Gleig, in his preface to the third edition, says, "The idea had been conceived by him [Colin Macfarquhar] and his friend, Mr. Andrew Bell, engraver." Macfarquhar, according to Constable, was a person of excellent taste and very general knowledge, though at starting he had little or no capital and was obliged to associate Bell, then the principal engraver in Edinburgh, as a partner in his undertaking.

The second edition was begun in 1776 and was published in numbers, of which the first was issued on June 21, 1777, and the last, no. 181, on Sept. 18, 1784, forming ten volumes dated 1778 to 1783 and containing 8,595 pages and 340 plates. The pagination is continuous, ending with page 9,200, but 295 pages are inserted in various places, and page 7,099 is followed by 8,000. The number and length of the articles were much increased, 72 have crossheadings and more than 150 others may be classed as long articles. At the end is an appendix ("Abatement" to "Wood") of 200 pages, containing, under the heading "Botanical Table," a list of the 931 genera included in the 58 natural orders of Linnaeus, and followed by a list of 526 books, said to have been the principal authorities used. All the maps are placed together under the article "Geography." Most of the long articles have numbered marginal titles; "Scotland," 82 pages, has 837, "Medicine," 305 pages, and "Pharmacy" have each an index. The plan of the work was enlarged by the addition of history and biography, which encyclopaedias in general had long omitted. Smellie was applied to by Bell to edit the second edition and to take a share of one-third in the work; but he refused on the ground that the introduction of a system of general biography was inconsistent with the character of a dictionary of arts and sciences. James Tytler seems to have been selected as the next most eligible compiler. Tytler (outlawed by the high court of justiciary in 1793, buried at Salem, Mass., in 1805, aged 58) "wrote," says Watt, "many of the scientific treatises and histories, and almost all the minor articles."

SINGLE MICROSCOPE WITH REFLECTION: FROM A PLATE IN THE SECOND EDITION OF ENCYCLOPÆDIA BRITANNICA, 1777–84. ENGRAVING BY ANDREW BELL



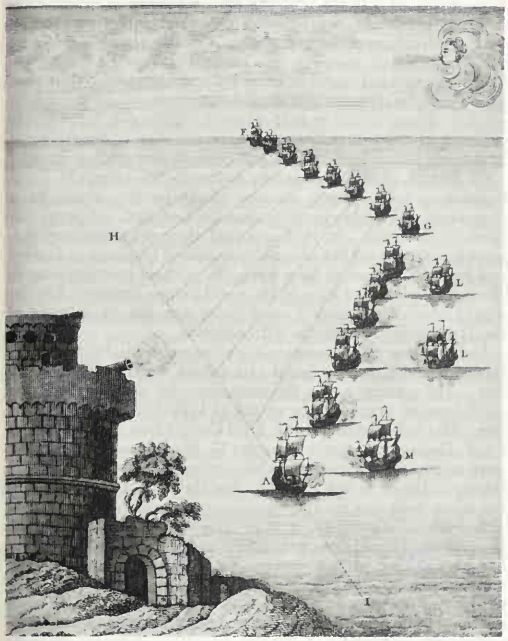
ROUND-CRESTED DIVER (MERGUS): FROM A PLATE IN THE SECOND EDITION OF ENCYCLOPÆDIA BRITANNICA, 1777–84

2. Third Edition.—After about a year's preparation, the third edition was announced in 1787; the first number was published early in 1788 and the first volume in Oct. 1788. There were to be 300 weekly numbers, price 1*s*. each, forming 30 parts at 10*s*. 6*d*. each, and 15 volumes, with 360 plates. It was completed in 1797 in 18 volumes containing 14,579 pages and 542 plates. Among the multifarious articles represented in the frontispiece, which was required by the traditional fashion of the period, is a balloon. The maps are, as in subsequent editions, distributed among the articles on the respective countries. It was edited by Macfarquhar as far as the article "Mysteries." In 1793 Macfarquhar died at the age of 48, "worn out," says Constable, "by fatigue and anxiety of mind." His children's trustees and Bell requested George Gleig of Stirling (consecrated on Oct. 30, 1808, assistant and successor to the bishop of Brechin), who had written about 12 articles, to edit the rest of the work. According to Robert Kerr (*Memoirs of . . . William Smellie*, vol. i, pp. 364–365), about 10,000 copies were printed, and the profit to the proprietors was £42,000, besides the payments for their respective work as tradesmen in the conduct of the publication—Bell as engraver of all the plates and Macfarquhar as sole printer. According to Constable, the impression was begun at 5,000 copies and concluded with a sale of 13,000. James Hunter, "an active bookseller of no character," who had a shop in Middle Row, Holborn, sold the book to the trade, and on his failure Thomson Bonar, a wine merchant, who had married Bell's daughter, became the seller of the book. He quarreled with his father-in-law, who would not see him during the ten years before his death. When the edition was completed, the copyright and remaining books were sold in order to wind up the concern, and "the whole was purchased by Bell, who gave £13 a copy, sold all the complete copies to the trade, printed up the odd volumes, and thus kept the work in the market for several years."

3. Supplements and Other Editions.—The supplement of the third edition, printed for Bonar and edited by Gleig, was published in 1801 in two volumes, containing 1,624 pages and 50 copperplates engraved by D. Lizars. In the dedication to the king, dated Stirling, Dec. 10, 1800, Gleig says: "The French *Encyclopédie* has been accused, and justly accused, of having disseminated far and wide the seeds of anarchy and atheism. If the *Encyclopædia Britannica* shall in any degree counteract the tendency of that pestiferous work, even these two volumes will not be wholly unworthy of your Majesty's attention." Thomas Thomson wrote "Chemistry," "Mineralogy" and other articles, in which the use of symbols was for the first time introduced into chemistry;

and these articles formed the first outline of his *System of Chemistry*.

The fourth edition, printed for Bell, was begun in 1800 or 1801 and finished in 1810 in 20 volumes, containing 16,033 pages, with 581 plates engraved by Bell. No articles were reprinted from the supplement, as Bell had not the copyright. William Wallace's articles on mathematics were much valued, and raised the scientific character of the work. Thomson declined the editorship, and recommended James Millar, afterward editor of the *Encyclopaedia Edinensis*. He was fond of natural history and a good chemist, but, according to Constable, slow and dilatory and not well qualified. The edition began with 1,250 copies and concluded at 4,000, of which two-thirds passed through the hands of Constable's firm. Early in 1804 Bell had offered Constable and his partner Hunter the copyright of the work, printing materials, etc., and all that was then printed of the fourth edition, for £20,000. This offer was in agitation in March 1804, when the two partners were in London. On May 5, 1804, after Lord Jeffrey's



MANEUVER IN NAVAL TACTICS, FROM A PLATE IN THE SECOND EDITION OF ENCYCLOPÆDIA BRITANNICA, 1777–84. TO CHANGE FROM A RETREATING FORM TO A LINE OF BATTLE (SHIPS LLM REPRESENTING THE ENEMY): AS SHIP A TURNS TOWARD POINT I, SHIPS IN LINE AG FOLLOW IN ITS WAKE WHILE REMAINING SHIPS MAINTAIN FORMATION UNTIL REACHING LINE IH, WHEREUPON THEY TURN IN CONCERT TOWARD POINT I

arrival in Edinburgh, as he relates to Francis Horner, they entrusted him with a design, on which he found that most of his friends had embarked with great eagerness, "for publishing an entire new encyclopaedia upon an improved plan. . . . W. Scott has embraced it with great affection. . . . The authors are to be paid at least as well as reviewers, and are to retain the copyright of their articles for separate publication if they think proper." It was then, perhaps, that Constable gave £100 to Bonar for the copyright of the supplement.

Bell died in 1809, and a fifth edition was begun immediately after the fourth as a mere reprint. Bell's trustees mismanaged the new edition so badly that, after the issue of five volumes, both the stock and the copyright were sold to Constable, who paid between £13,000 and £14,000. Bonar, who lived next door to the printing office, thought he could conduct the book, and had resolved on the purchase. Having a good deal of money, he seemed to Constable a formidable rival, whose alliance was to be secured. After "sundry interviews" it was agreed that Constable should buy the copyright in his own name, and that Bonar should have one-third, and also one-third of the copyright of the supplement, for which he gave £200. Millar corrected and revised the last 15 volumes. The edition dated 1817 was published in 20 volumes, 16,017 pages, 582 plates, price £36.

Soon after the purchase of the copyright, Constable began to prepare for the publication of a supplement, to be of four or, at the very utmost, five volumes. Dugald Stewart, in a letter to Constable, Nov. 15, 1812, though he declines to engage to execute any of his own suggestions, recommends that four discourses should "stand in front," forming "a general map of the various departments of human knowledge," similar to "the excellent discourse prefixed by D'Alembert to the French *Encyclopédie*," together with historical sketches of the progress since Bacon's time of modern discoveries in metaphysical, moral and political philosophy, in mathematics and physics, in chemistry and in zoology, botany and mineralogy. He would promise only to undertake the general map and the first historical sketch, if his health and other engagements permitted. For the second he recommended John Playfair, for chemistry Sir Humphry Davy. Stewart received £1,000 for the first part of his dissertation (166 pages) and £700 for the second (257 pages), the right of publication being limited to the supplement and encyclopaedia. Constable next contracted with Playfair for a dissertation "to be equal in length or not to Stewart's." He at first intended to have two editors, "one for the strictly literary and the other for the scientific department." He applied to Thomas Brown, who "preferred writing trash of poetry to useful and lucrative employment." At last he fixed on Macvey Napier (1776–1847), whom he had known from 1798 and who "had been a hard student, and at college laid a good foundation for his future career, though more perhaps in general information than in what would be, strictly speaking, called scholarship." Napier went to London and obtained the co-operation of many literary men. The supplement was published in half-volume parts from Dec. 1816 to April 1824. It formed six volumes containing 4,933 pages, 125 plates, 9 maps, 3 dissertations and 669 articles, of which a list is given at the end. In his dissertation on the "progress of metaphysical, ethical and political philosophy," Stewart criticizes D'Alembert's account of the genealogy of knowledge and his encyclopaedic arrangement of the objects of human knowledge. Stewart emphasizes the contribution of the gifted individual rather than the history of the human species, and, although he thinks Bacon's classification of the sciences faulty, he acknowledges that no philosopher since Bacon's time has improved on it. The practical influence of Bacon and Locke, with which the dissertation closes, derives from their realization that the improvement of man depends less on cultivating reasoning powers than on controlling impressions and associations, formed early and riveted by habit, which may hold reason in bondage. Among the distinguished contributors were James Mill, David Ricardo, T. R. Malthus, D. F. J. Arago, J. B. Biot, William Hazlitt and Sir Walter Scott, who, to gratify his generous friend Constable, laid aside *Waverley*, which he was completing for publication, and in April and May 1814 wrote "Chivalry." There were about 160 biographies, chiefly of persons who had died within the preceding 30 years. Signatures, on the plan of the *Encyclopédie*, were annexed to each article, the list forming a triple alphabet, A to XXX, with the full names of the 72 contributors arranged apparently in the order of their first occurrence. At the end of vol. vi are addenda and corrigenda.

The sixth edition, "revised, corrected and improved," appeared in half-volume parts, price 16*s.* in boards, vol. xx, part ii completing the work in May 1823. Constable, thinking it not wise to reprint so large a book year after year without correction, in 1820 selected Charles Maclaren (1782–1866) as editor. "His attention was chiefly directed to the historical and geographical articles. He was to keep the press going, and have the whole completed in three years." A new edition in 25 volumes was contemplated, not to be announced until a certain time after the supplement was finished; but Constable's house stopped payment on Jan. 19, 1826, and his

copyrights were sold by auction. Those of the encyclopaedia were bought by contract, on July 16, 1828, for £6,150, by Thomas Allan, proprietor of the *Caledonian Mercury*, Adam Black, Abram Thomson, bookbinder, and Alexander Wight, banker, who, with the trustee of Constable's estate, had previously begun the seventh edition. Not many years later Black purchased all the shares and became sole proprietor.

The seventh edition, 21 volumes (with an index of 187 pages), containing 17,101 pages and 506 plates, edited by Napier, assisted by James Browne, was begun in 1827 and published from March 1830 to Jan. 1842. It was reset throughout and stereotyped. Mathematical diagrams were printed in the text from woodcuts. The dissertations of Stewart and others and their index of 30 pages filled vol. i. As they did not include Greek philosophy, "Aristotle," "Plato" and "Socrates" were supplied by R. D. Hampden, afterward bishop of Hereford. Among the numerous contributors of eminence, mention may be made of Sir David Brewster, Thomas De Quincey, Sir Anthony Panizzi and Robert Stephenson. Zoology was divided into 11 chief articles, "Mammalia," "Ornithology," "Reptilia," "Ichthyology," "Mollusca," "Crustacea," "Arachnides," "Entomology," "Helminthology," "Zoophytes" and "Animalcule," all by James Wilson.

The eighth edition, 1853–60, 21 volumes (and index of 239 pages, 1861), containing 17,957 pages and 402 plates, with many woodcuts, was edited by Thomas Stewart Traill, professor of medical jurisprudence in Edinburgh university. The dissertations were reprinted, with one on the "Rise, Progress, and Corruptions of Christianity" by Archbishop Richard Whately. Lord Macaulay, Charles Kingsley, Robert Chambers, the Rev. Charles Merivale, David Forbes, Austen Henry Layard, Baron Christian von Bunsen, Sir John Herschel, Edward Everett, William Thomson and John Stuart Blackie were some of the many eminent new contributors found among the 344 authors, of whom an alphabetical list is given, with a key to the signatures. This edition was not wholly reset like the seventh, and many long articles were retained almost or entirely intact.

4. Ninth Edition.—The publication of the ninth edition (A. and C. Black) was commenced in Jan. 1875 under the editorship of Thomas Spencer Baynes until 1880 and subsequently of W. Robertson Smith, and completed in 1889, 25 volumes including an index. This great edition retained a certain amount of the valuable material in the eighth but was substantially a new work; and it was universally acknowledged to stand in the forefront of the scholarship of its time. Its contributors included the most distinguished men of letters and of science.

The sale of the ninth edition by its publishers from 1875, when the first volume appeared, until 1898, a period of 23 years, was about 10,000 sets. The authorized American distributors, Charles Scribner's Sons, sold in the United States 45,000 sets of the genuine ninth edition. At that time, however, so far as the United States was concerned, there was no international copyright. Any enterprising American was at liberty legally to appropriate a work published abroad and reissue it in the United States as his own property. The ninth edition presented a tempting morsel to a number of publishers and consequently several hundred thousand sets of mutilated and fraudulent ninth editions were sold in the United States before the courts intervened and the bogus plates were finally melted down and the fraud ended.

In 1897 four Americans united to form a company to push the sale of *Encyclopædia Britannica* throughout the world. One of the four, Horace E. Hooper, representing the company, concluded a contract with A. & C. Black, Ltd., for the printing of 5,000 sets and a contract with the *Times* of London for the sale through and in the name of that influential paper. The sale began in March 1898 and was successful. The first sale was almost immediately followed by a second, and a third, and a fourth, and for the first time in its history *Encyclopædia Britannica* was distributed in considerable quantities not only in Great Britain and the United States but to the ends of the earth.

Of the company of four, Hooper and Walter M. Jackson were in London. The other two partners remained in the United States and in 1900 sold their half-ownership to Hooper and Jackson. On May 9, 1901, Hooper and Jackson purchased from A. & C. Black, Ltd., all rights, copyrights, plates, manuscripts, etc., of *Encyclopædia Britannica*.

At that time work had been in progress on the supplementary volumes which appeared in 1902. Following the old Scots custom, these 11 new volumes, added to those of the ninth edition, comprised the tenth edition. The editors of that edition were Sir Donald Mackenzie Wallace, Hugh Chisholm, Arthur T. Hadley and Franklin H. Hooper, the brother of Horace Hooper.

5. 11th Edition.—In 1903 work was begun on the 11th edition, with Chisholm editor in London and Franklin Hooper American editor. This edition bore the imprimatur of Cambridge university through an arrangement devised by Horace Hooper. The university received royalties on sales and exercised a certain amount of editorial supervision. The edition was issued in 1910–11 in 29 volumes, one being the index. During the period of editorial composition Horace Hooper purchased the Jackson interests in the property and by 1915 was sole owner, barring a few qualifying shares, of both the U.S. and British companies.

World conditions and financial difficulties increased the distribution problems of the 11th edition. Finally arrangements were made with Sears, Roebuck and Co. of Chicago to handle the sale of a reprinted, small-volume, although complete, edition of the 11th, to be sold at a considerably lower price than the regular edition. The first sales in 1915–16 were encouraging, but with the entrance of the United States into World War I, *Encyclopædia Britannica* found itself in a serious position financially.

On Feb. 24, 1920, *Encyclopædia Britannica* was bought by Sears, Roebuck and Co., largely because of the interest which Julius Rosenwald, the philanthropist, then head of that company, had taken in the work. In 1922 appeared three supplementary volumes prepared under the editorial direction of Chisholm in London and Franklin Hooper in New York city. These volumes were an attempt to record the state of the world before, during and immediately after World War I, and along with the 29 volumes of the 11th edition became the 12th edition.

In 1926 three new supplementary volumes, displacing the earlier three and added to the original 11th edition, became the 13th edition. The editorship was the same, and the substitute volumes were designed to present the events of 1910–26 in better perspective. Marshal Ferdinand Foch, Lord Cecil, Albert Einstein, Marie Curie, Gustav Stresemann, Thomas G. Masaryk and Elihu Root were among the contributors whose names gave international authority to the three new volumes. But the bulk of the material in the whole work was growing old.

6. 14th Edition.—In spite of the supplementary volumes, the 12th and 13th editions were actually the old 11th edition, which had been printed without revision for almost a quarter of a century. Moreover, work on the 11th edition having been begun as early as 1903, many of the contributions were as much as seven years old before that edition appeared in 1910. A brand new 14th edition was indicated.

The sum of approximately $2,500,000 was invested in the assembling and promotion of the 14th edition. It was edited by J. L. Garvin in London and Franklin Hooper in New York, and it took approximately three years to complete this work. More than 3,500 distinguished men and women of all nationalities contributed articles. The set of 24 volumes, one of which contained an index and a complete atlas, was published in Sept. 1929, on the eve of the great depression. Although sales in the beginning were good, the next four or five years proved to be trying ones in *Encyclopædia Britannica* history.

7. Britannica Plan of Continuous Revision.—The long history of *Encyclopædia Britannica* had been a series of fat and lean periods. Once a new edition of the books had been completed, the editorial staff was disbanded and a distributing organization created. Sales, for a time, were generally good. This was the fat interval. For 10 to 20 years the same numbered edition was printed without any revision of its contents. Soon the age of the set was noted, and fewer and fewer sales were made. In order to prepare a completely new edition, or even supplementary volumes, it was necessary then to organize a new editorial staff.

Once the new edition was in process the fact became publicly known and the sales of the then current edition ceased and the selling department was disbanded. Vast sums of money had to be expended to cover the editorial costs. This was the lean interval, and, because it required from 3 to 14 years to make a new edition, the lean interval was a long, hard one. The traditional method of publishing—which tolerated only a sales organization or an editorial staff but never the two together—was not only inefficient and uneconomical but often financially and editorially disastrous. In 1933 E. H. Powell, at that time secretary and treasurer of Sears, Roebuck and Co., was appointed president of Encyclopædia Britannica, Inc. He began the formulation of a plan which would permit the maintenance, year in and year out, of the sales and operating division of *Encyclopædia Britannica* by creating continuously the financial reserve necessary to keep the volumes always timely and always salable.

The editorial problem was solved first by dividing all the thousands of articles in *Encyclopædia Britannica* into a number of broad classifications. These classifications were scheduled for complete revision at varying intervals so that each article would be reviewed and revised, if necessary, on a regular schedule. It meant that *Encyclopædia Britannica* would no longer be issued in numbered editions; that new printings or editions (identified by year of copyright) would be made each year and only enough sets printed to meet the year's demand; that *Encyclopædia Britannica* would be revised, in part, annually under a systematic classification schedule; that, consequently, the material in each annual issue of the encyclopaedia would be more current than the material had ever been in any of the earlier numbered editions, and that, therefore, the books would always be in demand. Editorial offices were maintained in both Chicago and London.

8. Later History.—Several *Britannica* publications were begun in the 1933–43 decade. One was the 1,000,000-word volume called *Britannica Book of the Year*, which covers the significant developments of the year preceding publication. The first of the annuals in this series was issued in 1938, edited by Franklin H. Hooper, who had been made editor in chief of *Encyclopædia Britannica* in 1933. Important also was a 12-volume encyclopaedia for children of elementary-school age called *Britannica Junior Encyclopædia*. Revised completely in 1947 to appear in 15 volumes, it was revised continuously thereafter.

Another was *The March of Man*, a chronological record of people and events from prehistoric times to 1934 in the form of a historical atlas and comparative time charts. A fourth was the *Encyclopædia Britannica World Atlas*, first published in 1942 and distinguished particularly for the extensive geographical summaries prepared under the direction of G. Donald Hudson.

Late in 1941 William Benton (*q.v.*), then a vice-president of The University of Chicago, obtained from Gen. Robert E. Wood, chairman of the board of directors of Sears, Roebuck and Co., the offer of all rights to *Encyclopædia Britannica* as a gift to the university. When the trustees of the university decided that it should not undertake the financial risk of operating the enterprise, Benton supplied the working capital. Under an agreement reached in 1943, Benton became chairman of the board of directors of Encyclopædia Britannica, Inc., and majority stockholder. Robert M. Hutchins, then president of the university, was named chairman of the Board of Editors.

Walter Yust (*q.v.;* 1894–1960) became editor in chief of the *Encyclopædia Britannica* and affiliated publications in 1938 and continued in that post until 1960, when he retired. Harry S. Ashmore was appointed editor in chief in Oct. 1960 and in April 1963 became director of research and development for special editorial projects. John V. Dodge, managing editor of *Encyclopædia Britannica* from 1950, was executive editor of all *Britannica* publications from 1960 until June 1964, when he was named senior vice-president—editorial, with responsibility for international publishing projects. Warren E. Preece was named editor of *Encyclopædia Britannica* in June 1964. Maurice B. Mitchell, president of the company since 1962, became editorial director also in 1963.

The University of Chicago assumed no management authority or responsibilities for *Britannica* enterprises, although its interest remained active. As in the case of Cambridge university and the 11th edition, The University of Chicago receives a royalty on sales. The encyclopaedia is published with the editorial advice of The University of Chicago faculties; a committee of members of the faculties of Oxford, Cambridge and London universities; and a committee at the University of Toronto.

In the years after the transfer other projects were undertaken by the firm. In 1943 *Britannica* acquired Erpi Films and renamed it Encyclopædia Britannica Films Inc. Within a few months The University of Chicago was given the educational film library of the Eastman Kodak company. Encyclopædia Britannica Films Inc. became one of the largest makers and distributors of educational films in the United States. In 1947 *Britannica* published *Ten Eventful Years*, a four-volume compendium covering the events of the decade 1937–46. In 1952 *Britannica* published the 54-volume *Great Books of the Western World*, edited by Hutchins and Mortimer J. Adler. Its original feature is the two-volume *Syntopicon* (literally, "a collection of topics"), a guide to "syntopical reading" in the 443 works in the set and an index based on 102 great ideas with 3,000 subtopics and about 163,000 references to material in the books themselves. In 1963 appeared *Gateway to the Great Books*, a ten-volume set of more than 200 shorter works by 135 authors, with a *Syntopical Guide*, bibliographies and graded reading plans. The editors were Hutchins and Adler, who also edited *The Great Ideas Today*. First published in 1961, it is an annual that seeks to illuminate contemporary events in the light of the great ideas.

In 1957 the Spanish-language *Enciclopedia Barsa* was published in 15 volumes, and *Libro del Año Barsa*, an annual, followed in 1960. A 16-volume Portuguese-language *Enciclopedia Barsa* was produced in Brazil in 1964. Meanwhile, in 1960 the *Children's Britannica* had been published in London in 12 volumes under the direction of the London editor of *Encyclopædia Britannica*, John Armitage. In 1961 the company purchased *Compton's Pictured Encyclopedia*. Encyclopædia Britannica Press, publisher of trade and educational books and materials, was established in 1962. The Merriam-Webster company was purchased in 1964.

BIBLIOGRAPHY.—L. N. Malclès, *Les sources du travail bibliographique*, vol. 1 (1950); A. Walford, *Guide to Reference Material* (1959; supplement, 1963); G. Zischka, *Index lexicorum: Bibliographie der lexikalischen Nachschlagwerke* (1959); R. Collison, *Encyclopaedias: Their History Throughout the Ages* (1964); H. Kogan, *The Great EB* (1958).

(W. Yu.; Hn. K.; X.)

ENDECOTT (ENDICOTT), **JOHN** (*c.* 1588–1665), colonial governor of Massachusetts under whose leadership the new colony made rapid progress, was probably born in Devonshire, Eng. The only record in existence touching on Endecott before 1628 suggests that he may have been a friend of Sir Edward Coke. The six grantees making up The New England Company for a Plantation in Massachusetts selected him as manager and governor of their plantation lying between lines drawn 3 mi. N. of the Merrimack river and 3 mi. S. of the Charles and extending from the Atlantic to the Pacific. In 1628 Endecott with about 60 fellow settlers came to Naumkeag, a location already occupied by a settlement of seceders from Plymouth led by Roger Conant. The initial difficulties between the two groups of settlers were short-lived and tradition has it that the establishment of good relations was the occasion for the change of the name of the settlement to Salem, the Hebrew word for peace. Endecott's differences of opinion with Thomas Morton (*q.v.*) and his settlers at Merry Mount (Quincy) were not so amicably resolved. Endecott did not approve of the conduct of Morton's colony and used rather highhanded methods, including the chopping down of the controversial maypole, to correct the situation. When the jurisdiction of the New England company was supplanted by that of the Massachusetts Bay company in 1629, Endecott was for a time (April 30, 1629, to June 12, 1630) the local governor. Although he was, as a contemporary described him, "a fit instrument to begin this Wildernesse-worke," the company appointed John Winthrop (*q.v.*) governor of the colony at the time the charter was transferred to Massachusetts.

There is no evidence that Endecott resented his successor. From Winthrop's arrival in 1630 until his death in 1649, the two men formed with Thomas Dudley (*q.v.*) a kind of ruling triumvirate

for the Bay colony. Endecott, the soldier, was blunt, tactless and hot-tempered. Winthrop, the country squire, was even-tempered and diplomatic. Endecott was a thoroughgoing Separatist and encouraged and sympathized with Roger Williams. Winthrop leaned toward a reformed Church of England. Despite these differences, which might have made them natural enemies, the two had considerable mutual respect which developed into a warm and enduring friendship. Both strong men, they complemented each other and served the young colony well.

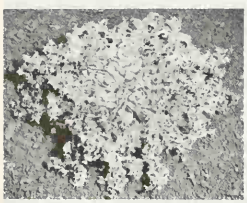
Endecott occupied, almost without a break, one or more prominent official positions in the colony. He served as a member of the board of assistants; member of the standing council of the colony; commander in chief of the militia, leading an expedition against the Pequots in 1636; commissioner of the United Colonies of New England; deputy governor in 1641–44, 1650–51 and 1654–55; and governor of the Colony of Massachusetts Bay in 1644–45, 1649–50, 1651–54 and 1655 until his death on March 15, 1665.

See L. S. May, *John Endicott: a Biography* (1936). (RA. MU.)

ENDEMIC, a biological term meaning restricted to a given area, *i.e.*, native or indigenous. In medicine endemic refers to any infectious disease that is constantly present in a population in the sense that a few cases may always be detected on careful search. An endemic disease is likely to assume epidemic proportions under conditions that are not fully understood. No doubt in endemic diseases, as in the occurrence of epidemics, the existence of chronic carriers of the disease (*see* CARRIER) is of fundamental importance, but the factors of susceptibility and immunity among the healthy (*see* IMMUNITY) are equally important and far more difficult to determine. *See* EPIDEMIOLOGY.

ENDERS, JOHN FRANKLIN (1897–), U.S. microbiologist, received the Nobel prize in physiology and medicine in 1954 for cultivation of the poliomyelitis viruses in tissue culture. Born in West Hartford, Conn., on Feb. 10, 1897, Enders received his early education at public school, St. Paul's school and Yale university (B.A., 1920). He served as a navy pilot during World War I. He received an M.A. degree in English literature from Harvard university in 1922 and a Ph.D. in bacteriology and immunology in 1930. His earlier research contributed new and basic knowledge to problems of tuberculosis, pneumococcal infections and resistance to bacterial diseases. He then began the study of viral diseases, including herpes or "cold sores," measles, influenza and mumps. The search for methods of growing these viruses in test tubes and flasks resulted in the use of various mammalian cells in tissue culture. In 1946 Enders established a laboratory at the Children's Medical center, Boston, where he, F. C. Robbins and T. H. Weller, with whom he shared the Nobel prize, were successful in cultivating the poliomyelitis viruses. This work made possible the development of poliomyelitis vaccine, new diagnostic methods and the isolation of many new viruses. (J. H. DE.)

ENDIVE, *Cichorium endivia,* an edible annual leafy plant of the family Compositae, commonly reputed to have been introduced into Europe from the East Indies, but, according to some authorities, probably indigenous to Egypt. It has been cultivated in England since the 16th century and is grown in America. There are many varieties, forming two groups, the much-curled or narrow-leaved (var. *crispa*), and the Batavian or broad-leaved (var. *latifolia*), slightly curled. The former are those most used for salads, the latter for culinary purposes. The plant requires a light, rich and well-drained soil, unshaded.



JOHN H. GERARD

ENDIVE (CICHORIUM ENDIVIA)

In central and southern U.S., sowing should be done in June or later lest the plants shoot to seed instead of forming desirable rosettes of edible leaves. When sown late in the season the plant behaves as a biennial. Transplantation, for early crops, is inadvisable. Bleaching of endive, to prevent development of the natural bitter taste of the leaves, is begun about three months after sowing, and is best effected either by tying the outer leaves around the inner or by covering the plants in some way to exclude light. Bleaching may be completed in about ten days in summer, but later in the season it takes three or four weeks. In England sowing for the main crop should begin about the second or third week in June; but for young plants it may be as early as the latter half of April and for winter crops up to the middle of August.

For French endive, or witloof chicory, *see* CHICORY.

(V. R. B.)

ENDLICHER, STEPHAN LADISLAUS (1804–1849), Austrian botanist known for his scheme of plant classification as put forward in his best-known work, *Genera Plantarum* . . . (1836–41), was born June 24, 1804, in Presburg, Hung., later Bratislava, Czech. At the universities of Budapest and Vienna (Ph.D., 1823), his studies included theology (minor vows, 1826), but he gave up plans for the priesthood to concentrate on natural history, philology and finally, medicine (M.D., Vienna, 1840). In 1827 he entered the Imperial library, Vienna, where he catalogued the manuscript section. In 1836, appointed curator of botany at the Museum of Natural History (also in Vienna), he gave the museum his large herbarium of over 30,000 numbers, purchased with his own funds. While reorganizing the collections of the herbarium, Endlicher wrote his *Genera Plantarum secundum Ordines Naturales disposita.* In 1840 he was appointed professor of botany at the University of Vienna and director of the botanical garden, which he completely reorganized. He published many important literary and historical works, including *Monumenta Arpadiana,* a Chinese grammar (1844–45) and an atlas of China (1843). He founded the first journal of natural history in Austria (the annals of the Museum of Natural History) and was co-founder with Carl Philipp von Martius (Munich) of the monumental *Flora Brasiliensis.* Endlicher underwrote the cost of publication of most of these works. He died in Vienna, March 28, 1849. (T. K. J.; X.)

ENDOCRINE GLANDS. In the animal body, groups of specialized cells that produce a specific type of secretion are called glands. When the secretion is conveyed from the cells directly into the blood stream, the gland is known as a ductless or endocrine gland. These tissues elaborate one or more highly specific chemical substances called hormones that act in minute amounts in a catalytic manner on other tissues, organs or physiological processes. Surgical removal of an endocrine gland produces a typical deficiency state in the animal, which is restored to normal by the administration of the appropriate hormone(s).

In the mammal the presently known endocrines are: the pituitary (posterior and anterior lobes), thyroid, parathyroids, pancreatic islets, adrenal cortex, adrenal medulla, ovary and testis. It has not been established whether the thymus and pineal glands secrete any hormones. In the invertebrate phyla there have been found neuro-endocrine structures, secreting hormones that regulate growth and metamorphosis. Certain tissues that structurally do not look like special glands probably secrete hormonal materials. Thus some cell groups in the intestinal tract exert hormonal influences on secretion and motility of the alimentary tract. In the central nervous system there are groups of cells that seem to serve in double capacity—as neurons and as producers of hormones. *See* ENDOCRINOLOGY; HORMONES, VERTEBRATE and articles on the individual glands. (R. LE.)

ENDOCRINOLOGY. Endocrinology is the science concerned with the normal and pathological physiology of the glands of internal secretion, the endocrine glands. The pituitary, thyroid, pancreas, adrenals, parathyroids, ovaries and testes are the glands with proved internal secretory functions. These glands synthesize hormones and secrete them into the blood stream, in contrast with the glands of external secretion, such as the salivary glands, whose products are conveyed via ducts to a surface and there serve their purpose. The hormones or chemical messengers are highly specific substances which, although their concentration in blood is always minute, may have profound effects upon tissues some distance from their source. Hormonal action may be highly selective, affecting the activity of just one gland or tissue, or it may be more or less generalized in its influence.

The hormones are regulators of cellular metabolism. In a sense

they may be regarded as catalytic agents. Thus the biological effects of internal secretions are brought about by virtue of their influence on the rates of critical intracellular biochemical reactions; *i.e.*, they are either inhibitors or accelerators. The cellular activities affected by hormones do not appear to be absolutely dependent upon their presence. There is no evidence that hormones are actually utilized as a constituent of the cellular process which they affect or that they serve as a source of energy. However, hormones are metabolized and inactivated by tissues, and they are also excreted in the urine, so that a relatively continuous secretion is required in order to maintain an effective concentration in the circulating blood.

The physiological effects of hormones may be grouped in the following categories: (1) the maintenance of homeostasis; *i.e.*, the preservation of a stable functional status for the organism as a whole; (2) the control of body growth; and (3) the development of secondary sex characteristics. The prime functions of each of the hormones would enable them to be listed beneath one or another of these three headings. In almost every case, however, there are actions of another sort, perhaps subordinate but nonetheless important. For example, the male hormone not only stimulates the development of masculine traits but also stimulates general growth.

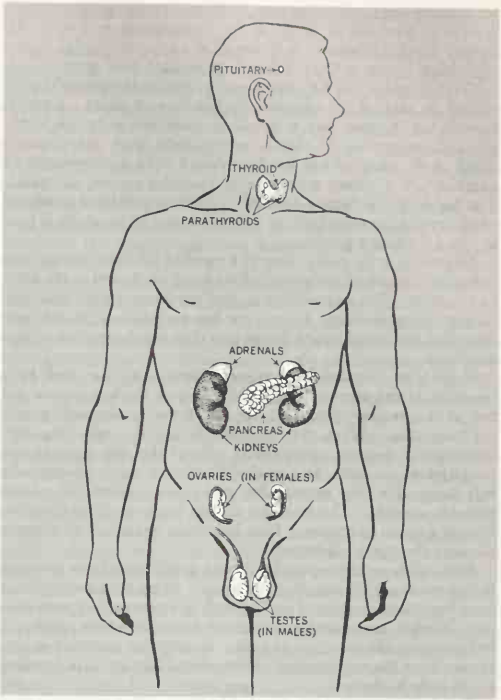
It is obvious that the quantity of these potent hormonal substances presented to tissues must be carefully regulated. The circulating concentration of any hormone is dependent upon the secretory activity of the gland producing it and the rate at which it is eliminated by inactivation and excretion. The situation is complicated by the fact that the requirement for the hormone or the sensitivity of the tissues to the secretion may vary under different conditions. Basically a proper balance is usually maintained by a system somewhat reminiscent of the feedback mechanism familiar to engineers. The blood concentration of the hormone or a product whose concentration is affected by it exerts a controlling influence on the rate of secretion. Thus an elevation in the hormone concentration or of the key product may inhibit the secretory process, and a lowered concentration may activate the gland. Neurohumours, hormonal substances of nervous origin, and nervous impulses may also play a role in the regulation of glandular activity.

The elegant precision of the regulatory mechanisms is amazing when the complexities of biological systems are taken into consideration. Disturbances of function are relatively rare, but when they occur the clinical disorders are often spectacular. Hormonal excesses are caused by functioning tumours or hyperplastic glands that have escaped from control. Deficiency states usually result from destructive tumours, infections or vascular lesions or from atrophy of the gland from unknown causes. Much of the modern knowledge in the field resulted from the careful study of endocrine disorders in experimental animals and in man.

In the study of endocrinology, anatomists, biochemists, physiologists, pathologists, pharmacologists and clinicians have cooperated and collaborated to a remarkable degree in advancing the science, as is exemplified in the approximately equally proportioned membership of physicians and preclinical scientists in the Endocrine society in the United States. It is noteworthy that not infrequently clinical endocrinological observations have anticipated laboratory developments.

In the discussion that follows an attempt will be made to survey what is known of the physiology of each of the endocrine glands and to relate the physiology to clinical endocrinology.

Pituitary Gland.—The pituitary gland or hypophysis is a small body (weight 0.5 g. in man) located beneath the central portion of the brain, to which it is attached by a thin stalk. The gland may be divided anatomically into an anterior portion, which secretes at least six active principles, and the posterior lobe or pars nervosa, which secretes two hormones. All the pituitary hormones are proteins or protein fragments (polypeptides). The gland is comprised of at least three types of cells differentiable by their reaction to stains. They have been termed eosinophiles, basophiles and chromophobes. It has been possible to deduce with some accuracy the type of cell that secretes each of the active principles.



APPROXIMATE LOCATIONS OF THE ENDOCRINE GLANDS IN THE HUMAN BODY (KIDNEYS, FALLOPIAN TUBES AND SPERM DUCTS ARE SHOWN ONLY TO INDICATE POSITION OF SOME OF THE GLANDS)

Growth hormone (*somatotrophin*), which is secreted by the eosinophilic cells, is a growth-promoting principle affecting practically all tissues of the body. When it is secreted in deficient quantities growth is retarded. Excessive amounts result in accelerated and at times abnormal growth. A diabetes-producing effect has been observed in the presence of large amounts of growth hormone.

Highly purified growth hormone has been obtained from the pituitary glands of man and a number of species of animals. The active portion of the hormone is a polypeptide that may or may not be attached to a protein molecule when secreted. With these products an enhancement of growth has been produced in experimental animals and man. There are, however, chemical differences in the growth hormone obtained from several species, and these differences are reflected in a degree of species specificity of the products. Growth hormone obtained from sheep or cattle is ineffective in man, but that from human and other primate sources has induced growth reactions in man.

Eosinophilic tumours of the anterior pituitary gland result in gigantism when they occur before puberty and acromegaly (excessive growth of face, hands, feet and soft tissue) when they develop afterward. Dwarfism occurs when hypopituitarism develops before full growth is attained. There is thus every indication that the growth hormone serves a vital function in the control of somatic growth; other possible roles have not been so clearly defined.

Thyrotrophin (*thyroid-stimulating hormone*) is secreted by basophile cells. It not only supports a functionally competent gland but also, as its name implies, enhances the rate of secretion of the thyroid hormone. It may also possess some influence aside from its effect on the thyroid. An exophthalmos-producing effect (exophthalmos means pathological prominence of the eyes) can be secured by the administration of purified thyrotrophin. An exophthalmos-producing factor has been separated from thyroid-

stimulating hormone. However, it is not certain whether this is a separate hormone or a part of the thyrotrophin molecule. Pituitary gland dysfunction is involved in a type of clinical hyperthyroidism known as Grave's disease or exophthalmic goitre.

Corticotrophin (*ACTH*) has a relationship to the adrenal cortex similar to that of thyrotrophin to the thyroid gland. Corticotrophin too is apparently secreted by basophilic cells. A highly purified preparation of this hormone made from the pituitary gland of the sheep or hog has been found to be a polypeptide (a combination of amino acids smaller than the protein molecule). The hormone maintains a normally functioning adrenal gland and when present in excess leads to a hyperadrenal state, which in man is called Cushing's syndrome.

Adipokinin is the name given to a purified pituitary extract that enhances the rate of movement of fat from fat depots in the body to the liver. The relationship of this substance to the other pituitary hormones is not known, nor has a clinical syndrome been described that could be attributed to a disordered secretion of this substance.

Melanocyte-stimulating hormone (*intermedin*) may well be a secretion of the narrow intermediate zone of the hypophysis and not of the anterior pituitary. It has chemical similarities with corticotrophin. When the hormone is injected into frogs the melanophore pigment contained in cells of the skin expands and thus darkens the skin. In man the injection of purified intermedin will darken the skin by inducing an increased deposition of the pigment melanin. The secretion of this hormone is inhibited by adrenal cortical hormones. Thus in adrenal deficiency (Addison's disease) the skin is darkened.

Prolactin is an active principle found in relatively large amounts in the pituitary glands of both sexes. After mammary glands have been fully developed and primed by estrogen and progesterone (ovarian hormones) and perhaps other secretions, prolactin will promote the secretion of milk. In rats this secretion is also required for the preservation of the corpus luteum. Its function in the male is obscure.

Two *gonadotrophins* are secreted by the anterior pituitary; they are the follicle-stimulating hormone (FSH) and luteinizing, or interstitial cell-stimulating, hormone (LH). Although separate functions have been defined for the two secretions, their effects on the gonads are often synergistic; *i.e.*, the presence of both hormones is required to complete a particular function. They are probably secreted by basophile cells.

In the female, follicle-stimulating hormone causes the Graafian follicle containing the maturing egg to grow to its full size and at the same time stimulates the secretion of estrogens by the ovaries. The rising estrogen concentration eventually suppresses FSH secretion and apparently encourages the secretion of pituitary luteinizing hormone. This secretion causes the ripened follicle to rupture and extrude the egg and then facilitates the conversion of the ruptured follicle into the corpus luteum, which under the continuing influence of luteinizing hormone (and perhaps prolactin) secretes progesterone. If pregnancy does not occur the corpus luteum atrophies, probably due in part to a suppression of LH secretion. The decline in progesterone secretion leads to menstruation in primates. In males the follicle-stimulating hormone, luteinizing or interstitial cell-stimulating hormone (ICSH) and male hormone or androgen combine to bring about the maturation of spermatogenic tissue and to support normal sperm production by the seminiferous tubules. ICSH controls the development and the activity of the interstitial cells of Leydig, which secrete androgen.

The destruction of most of the anterior pituitary leads to a clinical condition termed panhypopituitarism. The gonads atrophy; evidence of hypothyroidism and adrenal deficiency becomes apparent. Cachexia (general ill-health) is a rare accompaniment. Fatigue and inability to withstand stress situations are common.

The *posterior pituitary* (*pars nervosa*) is an extension of the hypothalamus, that portion of the brain to which the pituitary stalk is attached. Nerves pass from nuclei in the hypothalamus to secretory cells, pituicytes, and apparently exert a controlling influence on their activity. The secretions of the posterior pituitary possess at least three primary functions: (1) pressor or vasoconstricting and blood pressure elevating action, (2) antidiuresis and (3) promotion of uterine contractions. These properties of the posterior pituitary power can be resolved into two hormones, vasopressin, which possesses the vasoconstricting and antidiuretic effects, and oxytocin. The exact chemical structure of these polypeptides has been determined.

The place of the vasoconstricting element in cardiovascular physiology is not known, but it is probable that oxytocin exerts an influence in the process of parturition. The antidiuretic principle is an important factor in the control of the hydration of the body. Its rate of secretion is regulated by osmoreceptors that attempt to maintain a constant osmotic pressure in circulating plasma and extracellular fluids. When plasma solutes (predominantly sodium chloride) become more concentrated, vasopressin secretion is increased and urine output is diminished, thus conserving water.

Diabetes insipidus is the disease resulting from the destruction of the posterior pituitary gland. Huge volumes of dilute urine are excreted, and serious dehydration develops if the patient is unable to keep up with the diuresis by consuming sufficient water. No sugar is excreted in the urine. The syndrome is observed only when the posterior lobe is destroyed and some anterior lobe function is preserved.

Hypothalamus. The hypothalamus is a small compartment of the brain whose principal function is to control a number of vital processes such as sleep, body temperature and appetite. The hypothalamus does not appear to be directly responsive to the conscious mind. Nerves pass from the hypothalamus to the pars nervosa and have a regulatory influence on posterior lobe activity. Some hypothalamic cells possess glandular characteristics, and there is evidence that anterior lobe function may be regulated by hypothalamic neurohumours passing through the blood stream to the pituitary body. Some measure of this sort of hypothalamic control appears to bear on the secretion of gonadotrophins, thyrotrophin and corticotrophin.

Thyroid Gland.—The thyroid gland is located in front and on either side of the trachea just beneath the larynx. The gland synthesizes the only iodine-containing hormones in the body, the amino acid thyroxin and several close chemical relatives. Thyroxin is stored in the gland in a chemical bond with protein (thyroglobulin). The hormonally active amino acids are secreted without the attached globulin but are loosely bound to plasma proteins while circulating.

The thyroid hormone affects the metabolism of practically all the tissues of the body. The principal function of the hormone is to regulate the rate of oxygen consumption; *i.e.*, the metabolic rate of tissues. An excess of thyroid produces hypermetabolism; a deficiency produces a low metabolic rate. The hormone is also required for the normal growth and development of the brain and musculoskeletal system. In its absence the gonads and sometimes the adrenal cortex do not function effectively. As noted above, the normal thyroid gland is maintained by the presence of an optimum amount of circulating pituitary thyrotrophin, and the rate of secretion of thyroid hormone is governed by the quantity of circulating thyrotrophin.

Iodine is also required in order to maintain a normally functioning thyroid gland. If the iodine content of the diet and drinking water are deficient or a drug prevents the gland from utilizing the circulating iodine, then synthesis of the hormone is impeded, the thyroid enlarges and a goitre develops. Some drugs and other chemicals interfere with the biosynthesis of the hormone and thus also result in goitre formation. Some of these are used in the clinical management of hyperthyroidism.

Hyperthyroidism (thyroid excess) is characterized by an elevated basal metabolic rate. The characteristic symptoms are nervousness, weight loss despite an increased appetite, a rapidly pounding heart and heat intolerance. In Grave's disease, in which there is an association of exophthalmos and a diffuse goitre, pituitary function is also disturbed. When nodular goitres become hyperactive the primary defect is thought to lie in the thyroid gland itself.

Hypothyroidism (thyroid deficiency) leads to a slowing of

both physical and mental processes. The skin becomes dry and puffy, hair is lost and the patient is sensitive to cold. In the adult the condition is known as myxedema. When it occurs in intrauterine life or early infancy, cretinism with retarded growth and sexual development is the result. Hypothyroidism is very satisfactorily treated orally with dried thyroid extract.

Parathyroid Glands.—The parathyroid glands are usually four in number and are located behind or beside the thyroid. There may, however, be from two to six or more glands, some of which may be within the chest in the region of the thymus gland. They are concerned with the regulation of calcium and phosphorous metabolism. No pituitary hormone governs the activity of the parathyroid glands. The parathyroid secretion, which is a protein, enhances the rate of calcium dissolution from bone and simultaneously increases the excretion of urinary phosphorus and ultimately urinary calcium. There is also evidence that connective tissues, in addition to bone, are also under the influence of the parathyroids. Normally functioning parathyroid glands maintain a remarkably stable concentration of serum calcium. An increase in serum calcium concentration suppresses the parathyroids, and a lowered level results in an increased secretion of parathyroid hormone.

Hyperparathyroidism, which may be due to functioning tumours or hyperplastic glands, produces a high serum calcium and a lowered serum phosphorus concentration. Fibrocystic bone disease (Recklinghausen's disease) with pathological fractures and distortion of bone is sometimes seen. If the patient's dietary calcium is high, however, the bone disease is minimized and multiple renal calcifications develop because of the large amount of calcium in the urine.

In hypoparathyroidism the serum calcium concentration is lowered, the serum phosphorous elevated. The patient has tetany, a state of muscular hyperirritability with spasms, and occasionally generalized convulsions. Cataracts are a frequent complication. Treatment with very large quantities of vitamin D or dihydrotachysterol is uniformly successful. Parathyroid extract is useful for short periods only.

Pancreas.—The pancreas is located in the abdomen within a bend of the duodenum, the first portion of the small intestine. The pancreas secretes the hormone insulin and digestive ferments that enter the duodenum. Insulin (*q.v.*), a protein, is synthesized by the beta cells of the islets of Langerhans, small conglomerates of cells located between the acini (smallest lobules) of the exocrine (outwardly secreting) portion of the gland. Although a number of the other hormones have some effect upon carbohydrate metabolism, it must be conceded that all are secondary in importance to insulin in this regard.

Insulin is the prime hormone concerned with the regulation of the blood glucose (sugar) level. Its effect is to lower the glucose concentration. An elevated blood glucose level evokes an increased secretion of insulin, and the pancreas responds to a lowered blood sugar by secreting less insulin. When the glucose concentration is lowered by insulin, the glucose is not converted to some other compound within the blood but the rate of its removal and utilization by tissues is enhanced. Some of the glucose is consumed for energy, and much of it is stored as glycogen (a starch-like compound composed of many glucose molecules) in muscles and in the liver. Insulin is absorbed and firmly held by cell membranes, and there are indications that it functions in part at least by facilitating the passage of glucose from extracellular fluid into the cells.

Insulin deficiency leads to diabetes mellitus, a disease characterized by hyperglycemia (elevated glucose in the blood) and spillage of sugar in the urine. The loss of large amounts of glucose is followed by weight loss and ultimately inanition since body tissue is utilized for energy purposes. The excretion of large amounts of sugar cannot be accomplished without a large urine volume, so that thirst is also a common symptom. Treatment with insulin, which must be administered by subcutaneous injection, is satisfactory; in certain cases, oral hypoglycemics control the disease. A moderate restriction in dietary carbohydrate is usually imposed as well. Unfortunately even treated diabetics are sometimes prone to premature development of degenerative diseases such as arteriosclerosis and cataracts.

The administration of excessive amounts of insulin produces hypoglycemia with attendant symptoms of weakness, sweating, confusion, double vision and sometimes loss of consciousness and generalized convulsions. Insulin-secreting tumours of the islets of Langerhans are associated with a similar clinical picture. A hyperglycemic factor, sometimes called glucagon, has also been isolated from pancreatic extracts. Whether it is actually a hormone or not is unknown.

Adrenal Glands.—The adrenals are rather small glands located just above the upper pole of each kidney and are in reality two glands in one, the medulla or inner portion and the cortex or outer portion.

The *adrenal medulla* is composed of chromaffin cells similar to those found scattered near the ganglia of the sympathetic nervous system. The medulla receives nerves from the sympathetic nervous system and secretes its hormone in response to nervous stimulation. Epinephrine (adrenaline) and closely related norepinephrine (noradrenaline) are secreted simultaneously. The proportions of each secreted vary widely from species to species (70%–90% epinephrine and 10%–30% norepinephrine in man). These secretions comprise an important element of the sympathetic nervous system's regulation of vascular tone and cardiac action. The effects of the two compounds differ somewhat. Considered together, however, as they are secreted, epinephrine and norepinephrine constrict peripheral blood vessels, increase the cardiac output, elevate the blood pressure and increase the concentration of blood glucose. In response to severe injuries, hemorrhages and other shocks to the body, the secretions of the medulla are increased; thus it plays an important part in supporting the organism in times of stress. The secretion of the hormones of the medulla may also be increased as a result of emotional stimulation. Epinephrine is useful in the treatment of allergic states, especially bronchial asthma, but its beneficial effects are of short duration. *See* also ADRENALINE AND NORADRENALINE.

Pheochromocytomas, hormone-secreting tumours of the adrenal medulla, may cause nervousness, hypertension and diabetes. The metabolic rate is always elevated during the active phases of this condition. In some instances the hormones are liberated only periodically by the tumour, giving rise to paroxysms of nervousness and palpitation with associated hypertension and hyperglycemia.

The *adrenal cortex* secretes several steroid hormones whose combined effects are to preserve homeostasis in several metabolic areas. The principal secretions are probably hydrocortisone, corticosterone and aldosterone. These secretions (corticoids) play a role in regulating rates of protein tissue breakdown, utilization of carbohydrate by peripheral tissues, mobilization of fat for energy purposes, maintenance of vascular tone and renal excretion of sodium, chloride, potassium and water. Hydrocortisone is the most effective hormone with respect to the influence on carbohydrate, protein and fat metabolism. Aldosterone is particularly potent in conserving salt. Maintenance of the gland requires the presence of pituitary corticotrophin, and the rate of secretion of hydrocortisone and corticosterone is governed by the pituitary through the secretion of ACTH. Aldosterone secretion is more or less independent of pituitary action. The adrenal cortex-pituitary axis is activated by a wide variety of bodily injuries. Apparently at time of stress the body requires more adrenal hormones. The adrenal-deficient organism is unable to withstand even moderately severe stress.

Chronic hypersecretion of corticoids as a result of adrenal stimulation by basophilic pituitary adenomas or by functioning adrenal tumours produces Cushing's syndrome in man. This is characterized by hypertension, diabetes mellitus, obesity confined to the trunk, hirsutism and gonadal hypofunction. Adrenal cortical dysfunction (tumours or hyperplastic glands) may also cause virilization due to the oversecretion of a male hormone, and more rarely feminization of men has been observed. Whether these syndromes result from the hypersecretion of a normal adrenal hormone or from the abnormal product of a disorganized and per-

verse gland is uncertain. Chronic adrenal cortical deficiency is termed Addison's disease. In this disorder the skin and mucous membranes acquire tan or black pigmentation, the blood pressure is poorly maintained, salt and water are lost in excessive amounts, and the end result when the condition is untreated is complete collapse and shock. Treatment with hydrocortisone or cortisone and desoxycorticosterone (a salt-retaining steroid) is very successful.

Developments after 1948 made it apparent that a number of serious allied disorders often grouped as connective tissue diseases may be favourably influenced by the administration of excesses of corticoid hormones or by stimulation of the adrenal glands with corticotrophin. Among diseases in which dramatic improvement was secured by such treatment are rheumatoid arthritis and rheumatic fever. Excessive amounts of adrenal hormones have also been found to block antibody-antigen reactions within the body and thus spectacularly overcome certain allergic states. The mechanism of these effects of adrenal secretions is not known. It must be emphasized that hypercorticism is induced in this treatment, and that Cushing's syndrome will accordingly complicate such therapeutic efforts whenever they are protracted.

Testes.—The male gonads are responsible for the production of the male gamete, sperm, and the synthesis and secretion of the male hormone androgen. In mammals the testicles are located in a sac, the scrotum, suspended from the groin. The temperature of the testicles is thus usually several degrees cooler than that of the rest of the body. Body temperature is too warm for spermatogenic tissue.

The sperm develop and mature within the coiled, microscopic spermatogenic tubules comprising the main bulk of the testis. In the absence of gonadotrophic secretions the infantile rudimentary spermatogenic tubules fail to develop and produce sperm. Follicle-stimulating hormone, interstitial cell-stimulating hormone and androgen are required, each in optimum concentration, for normal spermatogenesis. An excess as well as a deficiency may be detrimental to the process.

The male hormone testosterone is a steroid secreted by the interstitial cells of Leydig. These cells are located between the spermatogenic tubules. In addition to aiding in the support of spermatogenesis, testosterone induces the growth and development of male secondary sex characters. These include the male accessory organs, the penis, prostate gland and seminal vesicles, and in man the beard, body hair and the potential for baldness. Androgen is also a potent general anabolic agent that stimulates the growth of muscle, skeleton and viscera and thus accounts in part for the greater musculature and stature of men as compared with women. The secretion of androgen is under the control of pituitary interstitial cell-stimulating hormone.

Androgen deficiency may be due to a pituitary defect or to a primary interstitial cell failure. If the deficiency is present from puberty, secondary sex characteristics fail to develop; if it occurs after maturity has been attained, a partial regression is the rule. The absence of androgen during adolescence leads to a delayed closure of the epiphyses of long bones and consequently to disproportionately long arms and legs. The administration of testosterone can satisfactorily substitute for normally produced androgen.

Ovaries.—The ovaries are located within the pelvis on either side of the uterus. The female gamete, the ovum, matures in and is extruded by the ovary. It secretes two female hormones, estrogen and progesterone.

Several estrogenic substances have been isolated from biological material. In women the closely related steroids estradiol, estrone and estriol are excreted in urine. Estradiol is the most potent and is probably the estrogen that is secreted; the others are metabolic derivatives. Estrogen is the female counterpart of androgen, except that for certain of its effects the synergistic action of progesterone is required. The female accessory organs, the uterus, Fallopian tubes, vagina, external genitalia and breasts, grow and mature under estrogenic influence. Estradiol is secreted by the Graafian follicles and interstitial cells under the control of the follicle-stimulating hormone.

The Graafian follicle ruptures as a result of luteinizing hormone action and extrudes the egg. The ruptured follicle is then converted to the corpus luteum, which secretes progesterone. Progesterone with estrogen induces the growth of the alveolar or secretory tissue of the breasts and the increased proliferation of the endometrium, the internal uterine lining, in preparation for implantation of the fertilized ovum. After conception progesterone preserves the proper tone and contractibility of the uterine muscle, the myometrium, so that pregnancy is maintained. Progesterone also plays an important role in the control of salt and water balance. It inhibits the salt-retaining action of the adrenal hormone aldosterone. If no pregnancy develops, the corpus luteum fails, progesterone secretion stops and the thickened endometrium sloughs. This is menstruation.

Prepubertal ovarian failure, which may well be primarily a pituitary or an ovarian defect, results in an absence of female sex characteristics and the persistence of infantile accessory sex tissues. Primary ovarian failure at age 35–55 is physiological. This is menopause. In addition to the cessation of menses, the menopausal woman may experience hot flushes and nervous tension. The uterus and vagina slowly atrophy because of the absence of estrogen. Treatment with any of a variety of synthetic or natural estrogenic compounds is generally effective in controlling the symptoms. Progesterone replacement is not necessary except in some prepubertally hypogonadal women.

Placenta.—The placenta secretes chorionic gonadotrophin, which is predominantly luteinizing and prevents the atrophy of the corpus luteum during early pregnancy. Later it secretes large amounts of estrogen and progesterone, so that after the fourth or fifth month the ovaries of a woman may be removed without interrupting the pregnancy. The growth of the uterus during pregnancy is largely estrogen-induced.

See also ENDOCRINE GLANDS and specific articles on the individual glands and their diseases. *See* HORMONES, VERTEBRATE; STEROIDS.

For endocrine effect on the sympathetic nervous system, *see* NERVOUS SYSTEM: *Autonomic Nervous System. See* also REPRODUCTION: *Physiology of Reproduction: Hormonal Control; see* also references under "Endocrinology" in the Index.

See R. H. Williams (ed.), *Textbook of Endocrinology,* 2nd ed. (1955); G. Pincus and K. V. Thimann (eds.), *The Hormones,* 3 vol. (1948, 1950, 1955). (R. L. L.)

ENDODONTICS is the division of dentistry that concerns itself with the prevention, diagnosis and treatment of diseases of the dental pulp. Pierre Fauchard described early endodontic procedures in the 18th century. Factors influencing later developments in endodontics were use of the rubber dam to secure isolation of the tooth under treatment (1864); introduction of gutta-percha as a permanent root canal filling material (*c.* 1870); discovery of the association of microorganisms with disease of the pulp (1880); utilization of X-rays in diagnosis and guidance of treatment (1896); discovery of a satisfactory local anesthetic, procaine (1906); and the concept of oral sepsis (W. D. Miller, 1889; W. Hunter, 1910; F. Billings, 1914).

The primary function of the dental pulp (commonly known as the nerve) is the formation of dentine. Dentine constitutes the main substance of the tooth, which grows around the pulp so that in the fully formed tooth the pulp occupies a central cavity. When the formation of dentine ceases, the primary function of the pulp is completed. A second function is related to the pulp's extreme sensitivity to heat, cold, touch and some chemical substances. Contact with any one of these will produce pain, which warns of a threat to the tooth.

The pulp may become inflamed or die from injury inflicted by dental decay, a blow to the tooth, deep fillings or the dental drill in preparing a cavity for filling. Death of the pulp usually results in root-end abscess; infection may pass by the blood stream to other parts of the body and produce inflammation, secondary abscess or other disease. Endodontic treatment, if initiated immediately after injury to or death of the pulp, can prevent formation of the dental abscess, or if the abscess has developed can usually eliminate infection without extraction. Treatment may

prevent recurrent infection at the root end. This permits maintenance of the tooth in position and function in the mouth.

The dentist gains access to the pulp through an opening in the crown of the tooth into the pulp canal. He removes the pulp or its remnants and enlarges the canal with the most delicate surgical instruments. Medicated irrigations aid in removing debris and in cleansing the canal. A medicament, which is replaced at intervals, is sealed in the canal to destroy microorganisms and thus promote healing. At times, treatment may require that a window be opened through the bone over the root end to remove masses of diseased tissue (cysts) that would delay or prevent healing.

After sterilization the root canal is filled permanently with solid cones, usually of pure silver or gutta-percha, that have a length and taper corresponding to those of the canal. The cones are then cemented into place with a sealer containing zinc oxide and eugenol. This operation obliterates the space formerly occupied by the pulp and seals the canal. *See* also TEETH. (R. G. K.)

ENDOEUS, an early Greek sculptor who worked at Athens in the middle of the 6th century B.C. He is said to have made an image of Athena dedicated by Callias, the contemporary of Pisistratus, at Athens about 564 B.C. An inscription bearing Endoeus' name has been found at Athens, written in Ionian dialect. The tradition which made him a pupil of Daedalus is apparently misleading, since Daedalus had no connection with Ionic art.

ENDOGAMY, the law or custom requiring a person to marry within his own social group, as opposed to exogamy (*q.v.*), which requires marriage outside the group. Endogamy refers to marriage only, not to extramarital relations. The most famous example is among the Hindus, who do not permit men and women of different castes to marry (*see* CASTE, INDIAN). Among primitive peoples it is relatively rare and occurs where there is stratification by rank or occupation, such as in Polynesia, among the Masai of east Africa or among the Flathead Indians of North America. The disapproval of marriages between persons of different classes may assume the form of a definite regulation, and this is true also of the feeling against a woman's marrying outside her own tribe or local group, but such unions can generally be condoned and diplomatic marriages between different tribes do take place. Exogamy and endogamy are by no means mutually exclusive. In any community there may be one type of social group (*e.g.*, the clan) which is exogamous and another type (*e.g.*, the occupational class) which is endogamous. *See* also MARRIAGE, PRIMITIVE. (C. H. W.; X.)

ENDOR, an ancient town of Palestine, 4 mi. S. of Mt. Tabor and 6 mi. S.E. of Nazareth, in early history the home of those members of the tribe of Manasseh living in the territory of Issachar (Josh. xvii, 11). It was the scene of the rout of Jabin and Sisera (Ps. lxxxiii, 9–10) and the abode of the witch (medium) whom Saul consulted on the eve of the battle of Bilboa (I Sam. xxviii).

An Israeli settlement, Ein Dor, is located at the site.

(E. Ro.; X.)

ENDOSCOPY is a special medical examination of the interior of the body, usually through a natural body opening, by the insertion of a lighted optical shaft or open tube. It is customarily performed with local anesthesia. Instruments used are the esophagoscope, a rigid pipe for examination of the esophagus; the bronchoscope, a similar pipe for examination of the bronchial tuhes; and the gastroscope, a flexible lighted shaft for examination of the stomach. These are passed through the mouth into the respective organs. The examinations are usually performed in a hospital but the patient ordinarily is dismissed immediately after. The open-tube proctosigmoidoscope is passed through the anal orifice without anesthesia for examination of the rectum and lower colon and is a common office procedure. The cystoscope, a lighted rod, is passed through the urethra with local anesthesia for examination of the bladder.

Three endoscopic procedures require incisions for the introduction of the lighted shaft where natural openings do not exist. The thoracoscope permits examination (for adhesions in patients with tuberculosis) of the chest cavity and surface of the lungs through a small incision between the ribs. The peritoneoscope allows examination of the abdominal cavity and lower surface of the liver and gall bladder by a small incision in the abdominal wall. The culdoscope permits examination of the female pelvic organs through a small vaginal incision. (A. P. K.)

ENDOTHELIUM, strictly speaking, is the layer of simple flattened (squamous) cells lining the inner surface of blood vessels, lymphatic vessels and the heart. The term is occasionally applied also to the lining of all closed cavities of the body, such as the pleural, pericardial or abdominal cavities and the anterior chamber of the eye. Endothelium is similar in histological appearance to squamous epithelium, but the former is derived embryologically from the mesoderm, whereas the latter is either of ectodermal or endodermal origin.

See also COELOM AND SEROUS MEMBRANES; EPITHELIUM; HISTOLOGY.

ENDOWMENT INSURANCE: *see* LIFE INSURANCE.

ENDYMION, in Greek mythology, son of Aëthlius and king of Elis. He was loved by Selene, goddess of the moon, by whom he had 50 daughters. Otherwise, he was a beautiful youth, a shepherd or hunter whom Selene visited every night while he lay asleep in a cave on Mt. Latmus in Caria. Zeus offered him anything he might desire, and he chose an everlasting sleep in which he might remain youthful forever. According to others, Endymion's eternal sleep was a punishment inflicted by Zeus upon him because he ventured to fall in love with Hera. The usual form of the legend, however, represents Endymion as having been put to sleep by Selene herself in order that she might enjoy his society undisturbed.

ENERGY, in physical science, a term which may be defined as accumulated mechanical work, which, however, may be only partially available for use (from the Gr. *energeia: en,* "in"; *ergon,* "work"). A bent spring possesses energy, for it is capable of doing work in returning to its natural form; a charge of gunpowder possesses energy, for it is capable of doing work in exploding; an electric current possesses energy, for it is capable of doing work in turning a motor; light possesses energy, as is shown in its ability to produce working currents in a photo cell. The motions of bodies, or of the ultimate parts of bodies, also involve energy, for stopping them would be a source of work.

Measurement.—All kinds of energy are ultimately measured in terms of work. If we raise 1 lb. of matter through 1 ft. we do a certain amount of work against the earth's attraction; if we raise 2 lb. through the same height we do twice this amount of work, and so on. Also, the work done in raising 1 lb. through 2 ft. will be double that done in raising it 1 ft. Thus we recognize that the work done varies conjointly as the resistance overcome and the distance through which it is overcome.

Now, we may select any definite quantity of work we please as our unit, as, for example, the work done in lifting 1 lb. 1 ft. high from the sea level in the latitude of London, which is the unit of work generally adopted by British engineers, and is called the foot-pound. A more appropriate unit for scientific purposes is one which depends only on the fundamental units of length, mass and time, and is hence called an absolute unit. Such a unit is independent of gravity or of any other quantity which varies with the locality. Taking the centimetre, gram and second as our fundamental units of length, mass and time, the most convenient unit of force is that which, acting on a gram for a second, produces in it a velocity of a centimetre per second; this is called a dyne. The unit of work is that which is required to overcome a resistance of a dyne over a centimetre, and is called an erg. In the latitude of Paris the dyne is equal to the weight of about $\frac{1}{981}$ of a gram, and the erg is the amount of work required to raise $\frac{1}{981}$ of a gram vertically through one centimetre.

Energy is commonly defined as the capacity for doing work. Since, however, we can not always bring about the change the term capacity is somewhat misleading; it is better to define energy as that which diminishes when work is done by an amount equal to the work so done. The unit of energy should therefore be the same as that of work, and the centimetre-gram-second (c.g.s.) unit of energy is the erg.

A foot-pound is equal to 1.356×10^7 ergs. Other units of energy in common use are the following: the joule, or watt-second,

equals 10^7 ergs; the kilowatt-hour equals 3,600,000 joules or 3.6×10^{13} ergs; the kilogram-metre equals 9.807×10^7 ergs.

Forms.—It will be seen from that which has been stated that energy may become manifest in many ways; or, in other words there are many forms of energy.

The forms of energy which are most readily recognized are of course those in which the energy can be most directly employed in doing mechanical work; and it is manifest that masses of matter which are large enough to be seen and handled are more readily dealt with mechanically than are smaller masses. Hence when useful work can be obtained from a system by simply connecting visible portions of it by a train of mechanism, such energy is more readily recognized than is that which would compel us to control the behaviour of molecules before we could transform it into useful work. This leads up to a fundamental distinction, introduced by Lord Kelvin, between "available energy," which we can turn to mechanical effect, and "diffuse energy," which is useless for that purpose.

Potential Energy.—If a pound weight be suspended by a string passing over a pulley, in descending through 10 ft. it is capable of raising nearly a pound weight attached to the other end of the string, through the same height, and thus can do nearly 10 foot-pounds of work. The smoother we make the pulley the more nearly does the amount of useful work which the weight is capable of doing approach 10 foot-pounds, and if we take into account the work done against the friction of the pulley, we may say that the work done by the descending weight is 10 foot-pounds, and hence when the weight is in its elevated position we have at disposal 10 foot-pounds more energy than when it is in the lower position. It should be noticed, however, that this energy is possessed by the system consisting of the earth and pound together, in virtue of their separation, and that neither could do work without the other to attract it. The system consisting of the earth and the pound therefore possesses an amount of energy which depends on the relative positions of its two parts. In most mechanical systems the working stresses acting between the parts can be determined when the relative positions of all the parts are known; and the energy which a system possesses in virtue of the relative positions of its parts, or its configuration, is classified as "potential energy," to distinguish it from energy of motion which we shall presently consider. The word potential does not imply that this energy is not real; it exists in potentiality only in the sense that it is stored away in some latent manner; but it can be drawn upon at will for mechanical work.

Kinetic Energy.—Let a force of F dynes act upon a body initially at rest and having a mass of m grams. If v is the velocity acquired by the mass in t seconds under the action of the force, F is defined as mv/t. In displacing the body through d centimetres, F does, by definition, Fd ergs of work. But d is $\frac{1}{2}vt$, the average velocity $\frac{1}{2}v$ multiplied by the time. The product Fd is therefore $\frac{1}{2}mv^2$. Being the work done by the force in accelerating the body, it is the energy now resident in the moving mass. This energy of motion is called kinetic energy.

A simple example of the transformation of kinetic energy into potential energy, and vice versa, is afforded by the pendulum. When at the limits of its swing, the pendulum is for an instant at rest, and all the energy of the oscillation is static or potential. When passing through its position of equilibrium, since gravity can do no more work upon it without changing its fixed point of support, all the energy of oscillation is kinetic. At intermediate positions the energy is partly kinetic and partly potential.

Available kinetic energy is possessed by a system of two or more bodies in virtue of the relative motion of its parts. Since our conception of velocity is essentially relative, it is plain that any property possessed by a body in virtue of its motion can be effectively possessed by it only in relation to those bodies with respect to which it is moving. If a body whose mass is m grams be moving with a velocity of v centimetres per second relative to the earth, the available kinetic energy possessed by the system is $\frac{1}{2}mv^2$ ergs if m be small relative to the earth. But if we consider two bodies each of mass m and one of them moving with velocity v relative to the other, so that each has a velocity $\frac{1}{2}v$ with respect to the earth, only $\frac{1}{4}mv^2$ units of work are available from this system alone. Thus the estimation of kinetic energy is intimately affected by the choice of our base of measurement.

Conservation of Energy.—When the stresses acting between the parts of a system depend only on the relative positions of those parts and are directed along the lines joining those parts, the sum of the kinetic energy and potential energy of the system is always the same, provided the system be not acted upon by anything outside it. Such a system is called "conservative," and is well illustrated by the swinging pendulum referred to above. But there are stresses, such as friction, which depend on the relative motion of the visible bodies between which they appear to act. When work is done against these forces no full equivalent of potential energy may be produced; this applies especially to frictional forces, for if the motion of the system be reversed the forces will be also reversed and will still oppose the motion. It was long believed that work done against such forces was lost, and it was not till the 19th century that the energy thus transformed was traced; the conservation of energy has become the master key to unlock the connections between the various aspects of physical nature. The conception of work and of energy was originally derived from observation of purely mechanical phenomena, that is to say, phenomena in which the relative positions and motions of visible portions of matter were all that were taken into consideration. Hence it is not surprising that, in those more subtle forms in which energy cannot be readily or completely converted into work, the universality of the principle of energy, its conservation, as regards amount, should for a long while have escaped recognition after it had become familiar in pure dynamics.

It was pointed out by Lord Kelvin and P. G. Tait that Sir Isaac Newton had divined the principle of the conservation of energy, so far as it belongs purely to mechanics. But what became of the work done against friction and such nonconservative forces remained obscure, while the chemical doctrine that heat was an indestructible substance afterward led to the idea that it was lost. There was, however, even before Newton's time, more than a suspicion that heat was due to the motions of the ultimate parts of which bodies are built. Francis Bacon expressed his conviction that heat consists of a kind of motion or "brisk agitation" of the particles of matter. In the *Novum Organum*, after giving a long list of the sources of heat, he says, "From these examples, taken collectively as well as singly, the nature whose limit is heat appears to be motion. . . . It must not be thought that heat generates motion or motion heat (though in some respects this is true), but the very essence of heat, or the substantial self of heat, is motion, and nothing else." It should be pointed out, however, that early writers make no distinction between heat and temperature. No definiteness therefore can be attributed to such anticipations as this. It was only when Joseph Black showed that heat was something (distinct from temperature) which could be measured that the way was cleared to bring it into the mechanical scheme.

Heat as a Form of Energy.—In a similar way, other forms of energy were shown to be interconvertible by later investigations. In fact, the energy concept has proved so fruitful that the whole field of physics is sometimes characterized as the science of energy and its transformations.

The first vigorous effort to restore the universality of the principle of conservation of energy, threatened by the realization that heat could be lost or created, was made by Benjamin Thompson, Count von Rumford, and was published in the *Phil. Trans.* for 1798. Rumford was engaged in superintending the boring of cannon in the military arsenal at Munich, and was struck by the amount of heat produced by the action of the boring bar upon the brass castings. In order to see whether the heat came out of the chips he compared the capacity for heat of the chips abraded by the boring bar with that of an equal quantity of the metal cut from the block by a fine saw, and obtained the same result in the two cases, from which he concluded that "the heat produced could not possibly have been furnished at the expense of the latent heat of the metallic chips." After a series of similar but more quantitative experiments, Rumford reviewed all the sources from which the heat might have been supposed to be derived, and concluded

that it was simply produced by the friction, and that the supply was inexhaustible. "It is unnecessary to add," he remarks, "that anything which any insulated body or system of bodies can continue to furnish *without limitation* cannot possibly be *a material substance;* and it appears to me to be extremely difficult, if not quite impossible, to form any distinct idea of anything capable of being excited and communicated in the manner that heat was excited and communicated in these experiments, except it be *motion.*"

About the same time, Sir Humphry Davy showed that two pieces of ice could be melted by rubbing them together in a vacuum, although everything surrounding them was at a temperature below the freezing point. He did not, however, infer that since the heat could not have been supplied by the ice, for ice absorbs heat in melting, that this experiment afforded conclusive proof against the substantial nature of heat.

Though we may allow that the results obtained by Rumford and Davy demonstrate satisfactorily that heat is in some way due to motion, yet they do not tell us to what particular dynamical quantity heat corresponds. For example, if it can be shown that, however the duration and all other conditions of the experiment may be varied, the same amount of heat can in the end always be produced when the same amount of energy is expended, then, and only then, can we infer that heat is a form of energy, and that the energy consumed has really been transformed into heat. This was left for Joule to achieve; his experiments conclusively prove that heat and energy are of the same nature, and that all other forms of energy can be transformed into an equivalent amount of heat.

The quantity of energy which, if entirely converted into heat, is capable of raising the temperature of the unit mass of water from 0° C. to 1° C. is called the mechanical equivalent of heat.

In 1842 J. R. Mayer, a physician at Heilbronn, published an attempt to determine the mechanical equivalent of heat from the heat produced when air is compressed. He made the assumption that the whole of the work done in compressing the air was converted into heat, and neglected the possibility of heat being consumed in doing work within the air itself or being produced by the transformation of internal potential energy. Joule afterward proved (*see* below) that Mayer's assumption was nearly in accordance with fact, so that his method was a sound one as far as experiment was concerned; and it was only on account of the values of the specific heats of air at constant pressure and at constant volume employed by him being very inexact that the value of the mechanical equivalent of heat obtained by Mayer was very far from the truth.

The science of heat owes much to James Prescott Joule of Manchester, to whom we are indebted more than to any other for the establishment of the principle of the conservation of energy on the broad basis on which it has since stood. The best known of Joule's experiments was that in which a brass paddle consisting of eight arms rotated in a cylindrical vessel of water containing four fixed vanes, which allowed the passage of the arms of the paddle but prevented the water from rotating as a whole. The paddle was driven by weights, and the temperature of the water was observed by thermometers which could indicate $\frac{1}{200}$ of a degree Fahrenheit. Special experiments were made to determine the work done against resistances outside the vessel of water, which amounted to about .006 of the whole, and corrections were made for the loss of heat by radiation, the buoyancy of the air affecting the descending weights, and the energy dissipated (*i.e.*, converted into heat) when the weights struck the floor with a finite velocity. From these experiments Joule obtained 772.692 foot-pounds in the latitude of Manchester as equivalent to the amount of heat required to raise 1 lb. of water through 1° F. from the freezing point. Adopting the centigrade scale, this gives 1,390.846 foot-pounds.

With an apparatus similar to the above, but smaller, made of iron and filled with mercury, Joule obtained results varying from 772.814 foot-pounds when driving weights of about 58 lb. were employed to 775.352 foot-pounds when the driving weights were only about 19½ lb. The close agreement between the results indicated that "the amount of heat produced by friction is proportional to the work done and independent of the nature of the rubbing surfaces." Joule inferred from them that the mechanical equivalent of heat is probably about 772 foot-pounds, or, employing the centigrade scale, about 1,390 foot-pounds.

In 1844 and 1845 Joule published a series of researches on the compression and expansion of air, which led to values for the mechanical equivalent of heat slightly at variance with those previously determined and depending on the air pressures employed. The explanation of these discrepancies brought to light the existence of an internal energy of a gas, an energy ascribable to the interactions between the molecules constituting the gas.

Joule was not content with assuming that when air is compressed or allowed to expand the heat generated or absorbed is the equivalent of the work done and of that only, no change being made in the internal energy of the air itself when the temperature is kept constant. To test this, two vessels were placed in the same calorimeter and connected by a tube with a stopcock. One contained air at a pressure of 22 atm., whereas the other was exhausted. On opening the stopcock no work was done by the expanding air against external forces, since it expanded into a vacuum, and it was found that no heat was generated or absorbed. This showed that Mayer's assumption was practically true. The subsequent researches of Joule and Lord Kelvin (*Phil. Trans.*, 1853, p. 357, 1854, p. 321 and 1862, p. 579) showed that the statement that no internal work is done when a gas expands or contracts is not quite true, but the amount is very small in the cases of those gases which, like oxygen, hydrogen and nitrogen, can only be liquefied by intense cold and pressure. (*See* Low-Temperature Physics.) This internal energy accounts for the slight variations in the experimental results previously observed.

For a long time the final result deduced by Joule was accepted as the standard value of the mechanical equivalent of heat. Later determinations by H. A. Rowland and others, necessitated by modern requirements, have shown that it is in error, but by less than 1%. The best value of the mechanical equivalent of heat is probably 778.028 foot-pounds. It is more common, however, to express this constant in c.g.s. units: the number of ergs required to raise the temperature of 1 gram of water by 1° C. is 4.187×10^7. The writings of Joule, which thus occupy the place of honour in the practical establishment of the conservation of energy, have been collected into two volumes published by the Physical Society of London. On the theoretical side the greatest stimulus came from the publication in 1847, without knowledge of Mayer or Joule, of H. L. F. von Helmholtz's great memoir, *Über die Erhaltung der Kraft,* followed immediately (1848–1852) by the establishment of the science of thermodynamics (*q.v.*), mainly by R. Clausius and Lord Kelvin on the basis of "Carnot's principle" (1824), modified in expression so as to be consistent with the conservation of energy. The general equation of thermodynamics consistent with the above investigations may be written: Heat entering a system = Increase of energy of the system plus the external work done by the system during the entry. This equation is one of the basic laws of thermodynamics.

This law, however, is not the only one which regulates the transformation of heat into mechanical energy and its reverse. There is furthermore a general tendency for mechanical and other forms of energy to be transformed into heat on account of the friction of rough surfaces, the resistance of conductors or similar causes, and thus to lose availability. In some cases, as when heat is converted into the kinetic energy of moving machinery or the potential energy of raised weights, we observe an ascent of energy from the less available form of heat to the more available form of mechanical energy, but in all cases this is accompanied by the transfer of other heat from a body at a high temperature to one at a lower temperature, thus losing availability to an extent that more than compensates for the rise. This fact is usually stated as the second law of thermodynamics. (*See* also Calorimetry; Heat: *Mechanical Theory of Heat.*) (W. Gar.; J. La.; H. Ma.)

Electromagnetic Energy.—In 1840 Joule showed that when an electric current was produced by means of a dynamo the heat generated in the conductor, when no external work was done by the current, was the same as if the energy employed in producing

the current had been converted into heat by friction, thus proving that electric currents conform to the principle of conservation of energy, since energy can neither be created nor destroyed by them. This was the first significant indication of the universal range of the principle, of its validity even in nonmechanical and nonthermal phenomena. It led subsequently to the establishment of specific formulas for various types of electric and other energies.

A conductor is said to have an electric potential v if v ergs of work are necessary in order to impart a unit amount of additional positive charge to it. The stresses against which this work is done are the forces of electrostatic repulsion between the charged conductor and the charge to be added to it. If the conductor is initially uncharged and hence at zero potential, and is then given a final charge Q corresponding to a final potential V, the process of charging being carried out (for the sake of convenience) by bringing up, successively, small equal increments of charge, then the average potential of the conductor during the process is $\frac{1}{2}V$. The work done is obtained on multiplying this average potential by Q. Hence the electrostatic energy of a charged body is given by the formula $\frac{1}{2}QV$.

In a similar way, it may be shown that the work done by an electromotive agent (battery, dynamo, etc.) in establishing a current I in a single conductor possessing a coefficient of self-induction L, is equal to $\frac{1}{2}LI^2$. The maintenance of such a current against a resistance R requires a supply of energy I^2R per second, and this is drawn from the battery or dynamo furnishing the electromotive force for the current.

The process of charging a body described above may be viewed from another angle. While it is natural to suppose the energy of a charged conductor to have its locus within the conductor, there is no physical fact requiring this assumption. It is indeed equally valid to regard the energy as distributed throughout the space surrounding the conductor and to ascribe it entirely to the electric field which has been established therein. This interpretation is more closely in harmony with the views of Michael Faraday who, through his famous researches from 1820 to 1830, advanced the notion of lines of electric force which, by being capable of supporting stresses, provided a mechanism for the expenditure of work and hence a seat of energy. According to this picture the electrical energy residing within a cubic centimetre of space pervaded by a field of intensity E is $\frac{1}{8}\pi \times E^2$ ergs.

Very similar results may be obtained for the magnetic case. The medium surrounding a magnet, called the magnetic field, contains magnetic lines of force, and if the field strength within a certain region is H, the magnetic energy which it contains is $\frac{1}{8}\pi \times H^2$ ergs per cubic centimetre. This fact allows a remarkable extension to the phenomena of optics. For, as James Clerk Maxwell showed in 1864, a light wave is a combination of rapidly alternating electric and magnetic fields; hence the formulas stated above describe the energy content of a light wave as well. (*See* also Electromagnetic Waves.)

Other Forms of Energy.—Application of the energy concept to other aspects of nature has led to the recognition of specific forms or types too numerous for description. Among them is the energy of sound, the work which a sound wave is capable of doing, or the heat it can generate during absorption. In the subjects of hydrodynamics and aerodynamics (*qq.v.*), use of the idea of energy is widespread and fruitful; in chemistry various forms of "free energy" have been introduced and have clarified many reactions. What is commonly known as chemical energy, *i.e.*, the work a substance may do or the heat it may evolve in undergoing a change of composition, is to be thought of as potential energy due to the mutual interaction of its molecules, or parts of molecules. No transformation has yet been discovered in which a violation of the law of conservation of energy could be seen; it is indeed unlikely that this can be the case. For while the concept of energy, and the law of its conservation, were first discovered as empirical facts, theory during the 19th century so thoroughly consolidated the different features of the energy problem, and so firmly embedded the conservation principle even in the definition of physical quantities, that its breakdown is difficult to conceive.

Equivalence of Mass and Energy.—The expression $\frac{1}{2}mv^2$, which was shown to be a measure of the kinetic energy of a moving mass, loses its validity when the velocity v becomes very great. In that case it is necessary to employ the theory of relativity (*q.v.*) proposed by Albert Einstein in 1905, in the calculation of this quantity. The result for the kinetic energy may then be expressed in terms of the mass at low velocities, m_o, and the speed of light c (approximately 3×10^{10} cm. per second or 186,000 mi. per second) by the formula

$$T = \text{kinetic energy} = m_o c^2\left[\left(1-\frac{v^2}{c^2}\right)^{-\frac{1}{2}}-1\right]$$

If v is small, this reduces to $\frac{1}{2}m_o v^2$. In general, however, the result can be written in the form $T = (m - m_o)c^2$, provided m stands for the variable quantity $m_o\left(1-\frac{v^2}{c^2}\right)^{-\frac{1}{2}}$. This is often called the relativistic mass of the body, in contradistinction to its rest mass m_o. While the latter is constant, the former becomes infinite as v approaches the speed of light.

An interesting fact emerges if the last formula is written in the form

$$m = m_o + \frac{T}{c^2}.$$

for it shows that the mass m of a body increases in proportion to its kinetic energy T. Mass and energy are thus seen to be equivalent; one can be converted into the other in accordance with the relation

$$\text{energy} = \text{mass} \times \text{square of velocity of light.}$$

This relation may be shown to hold not only for kinetic, but for all other types of energy.

Prior to its discovery, the principle of conservation of energy was supplemented by the principle of conservation of mass. Later it was seen that the latter is merely a special case of the former.

The truth of the equivalence relation is confirmed by numerous careful investigations dealing chiefly with atomic nuclei. Forces between ordinary particles, including those between atoms or molecules, are too weak, and energies produced by them too small, to make an appreciable contribution to m because of the very small conversion factor $\frac{1}{c^2}$. But the forces between the constituent particles of atomic nuclei (protons and neutrons) are enormously great and their energies affect m in a measurable way. For example, a proton attracts a neutron so strongly that in the union of the two, that is, in the formation of the nucleus of heavy hydrogen, 3.45×10^{-6} ergs of work are done, and hence this amount of energy is lost. As a result, the nucleus of heavy hydrogen should have a mass smaller than the sum of the masses of a neutron and a proton by $3.45 \times 10^{-6}/c^2 = 3.8 \times 10^{-27}$ g.; this "mass defect" amounts to 0.0023 times the mass of a proton and is indeed observed. By such methods, the equivalence of mass and energy has been tested in a great number of nuclear processes.

Quantization of Energy.—In 1900, Max Planck discovered the solution of the baffling problem of "black-body radiation," the problem of why the radiation emitted by certain incandescent solids possesses the peculiar distribution in frequency that is observed. His celebrated solution involved the postulate of the quantization of radiant energy: the energy of a light wave, instead of being a continuous variable capable of taking on any value, is a constant multiple of some smallest unit called a quantum. In the case of light, Planck found this quantum to be proportional to the frequency; hence quanta of blue light are larger than those of red light. Indeed if ν is the frequency of the (monochromatic) light under consideration, its quantum is $h\nu$, h being called Planck's constant and having the value 6.625×10^{-27} erg. sec. Therefore, the energy of a monochromatic light wave is $E = nh\nu$. In this formula n, an integer, is the number of quanta, or "bundles of energy" possessed by the light wave. A single bundle of energy $h\nu$ is frequently called a photon.

Niels Bohr, in 1913, advanced the first successful theory of atomic structure and showed that quantization is very general in atomic phenomena. His results, brilliantly confirmed by spectroscopic observations, include the thesis that atomic electrons,

far from having any energy they please, obey quantum laws which select for them certain specific energy values and force them to move in definite orbits. The energies in this case are no longer multiples of a single quantum but are still functions of integers called quantum numbers as well as functions of h. When an electron changes from one quantum state of energy E_1 to another of energy E_2, the difference $E_1 - E_2$, if positive, is emitted in the form of a monochromatic light wave of frequency ν such that

$$E_1 - E_2 = h\nu$$

Later developments of the quantum theory (*see* QUANTUM MECHANICS), particularly at the hands of W. Heisenberg, E. Schrödinger and P. A. M. Dirac, though creating a different theoretical foundation, have left the facts of the quantization of energy intact (except for minor modifications). Atomic, molecular and nuclear systems, when in their normal condition, possess a discrete arrangement of energy states, the arrangement depending on the nature of the system. In addition, there is a continuum of unstable states which, if realized, lead to the disruption of the system. These continuous states correspond to ionization of an atom, dissociation of a molecule and radioactive emission by a nucleus.

Quantization implies the existence of zero-point energy; *i.e.*, an irreducible energy remaining in an atomic system even when the temperature of its surroundings is cooled to absolute zero. It occurs because the lowest quantized state of a particle is not of zero but of finite energy. Empirical evidence for zero-point energy is found in the thermal behaviour of gases at low temperatures.

See ATOMIC ENERGY; DYNAMICS; MECHANICS; *see* also references under "Energy" in the Index volume. (H. MA.)

ENERGY CONVERSION, DIRECT. Direct conversion is a rather loosely defined term covering processes in which heat or radiation is transformed direct to electricity rather than first to mechanical energy and then to electricity. Direct conversion has several advantages, among the most important of which is that electricity can be generated by devices employing no moving parts. Some of the phenomena—such as the generation of an electrical potential difference in a circuit consisting of two dissimilar metals with the junctions at different temperatures, and the flow of electricity which occurs when metal plates are illuminated with ultraviolet light—have been known for many years. These phenomena were discovered by T. J. Seebeck in 1826 and by H. R. Hertz in 1887, respectively. However, it was not until the late 1950s that serious consideration was given to the production by direct conversion of significant amounts of electrical power at reasonable efficiencies.

The unifying consideration in all energy-conversion devices is the Carnot cycle, which states that the maximum efficiency cannot be greater than the difference between the temperatures of the heat source and heat sink divided by the temperature of the source. It is often difficult to see the connection between a heat engine and the phenomenon being exploited for the production of electricity, but the analogy to the Carnot engine is nevertheless always present.

Thermoelectric Devices.—Thermoelectric phenomena come into existence because of the tendency of charged particles (both negatively charged electrons and positively charged "holes") to travel from the hot end of the material to the cold end. If one had a material with an initially uniform density of electrons independent of temperature, then—since the electrons in the hot end would be moving faster than those in the cold end—there would be a tendency for more electrons to leave the hot end and travel into the cold end than vice versa. However, as soon as a few electrons had diffused to the cold part, they would give rise to an electrical field which would discourage further flow of electrons to the cold part, thus bringing about a steady-state condition. Another phenomenon that can bring about a flow of charged particles from one end to the other occurs when the density of free carriers increases with increasing temperature, in accordance with the same type of equation that describes the increase in water vapour pressure with temperature. Thus, if there are many more free electrons at the hot end of the material, there is a strong tendency for them to diffuse through the material to the cold end, raising it to a negative potential.

Experiments with a variety of materials show that the cold end could become either negatively charged or positively charged. The reason for this is that in semiconductors or metals the current can be carried either by electrons or by holes. Thus, in order to obtain the maximum voltage or current in a thermoelectric circuit, one leg should be composed of a material in which current is carried by negative carriers and the other of a material in which current is carried by positive carriers. In other words, the temperature gradient or heat flow in both legs is in the same direction but the electrical currents flow up one leg and down the other, thus reinforcing each other.

When current flows from one material to another there is an abrupt change, across the junction between them, in the environment in which the charged particle is moving. This gives rise to what is called a Peltier heat. The process is analogous to the change in energy of an ion when it moves from one solution to another through a membrane. There is a thermodynamic relationship between this Peltier heat, the temperature difference and the thermoelectric potential difference, a relationship known as the Seebeck voltage, generated when the circuit is broken and the electrical potential measured. Charge carriers flowing through a circuit not only provide electrical current but also carry kinetic energy with them. This kinetic energy is a form of heat, and this heat is wasted. In every Carnot engine some heat is transformed and the rest wasted.

The over-all efficiency of a thermoelectric material can be expressed in terms of a dimensionless parameter called the figure of merit:

$$ZT = \frac{S^2 \sigma T}{\kappa}$$

where S is the Seebeck coefficient in volts per degree (centigrade), T is the temperature in degrees (absolute), σ is the electrical conductivity in ohms per centimetre, and κ is the thermal conductivity in watts per centimetre per degree (absolute). The fraction of the Carnot cycle efficiency which can be obtained for cases where ZT is less than unity is $ZT/4$.

Metals have values of ZT of a few per cent, while semiconductors can have values as large as 1, or possibly a little larger. In the early 1960s the best over-all efficiency that had been obtained with a semiconducting device was of the order of 10%.

Thermionic Devices.—Thermionic devices consist of evacuated or plasma-filled cells in which electrons are boiled out of a hot anode and are collected at the cold cathode. Thomas A. Edison applied for a patent in 1883 on a direct-conversion device utilizing thermionic emission, although at the time he was not aware of the physical principles of the device. The Edison tube was essentially one of his carbon-filament light bulbs into which he inserted an extra electrode. It was extremely inefficient.

To improve the efficiency, the loss of heat by radiation from the hot cathode needs to be reduced as much as possible. This can be accomplished by employing cathodes of low optical emissivity and anodes of high reflectivity. A means of avoiding space charge, the self-shielding effect associated with electrons in transit from the cathode to the anode, is also needed. This can be accomplished either by making the distance between the electrodes so small (say, less than $\frac{1}{1,000}$ in.) that the total number of electrons in the region is small, or by filling the space with some easily ionized gas such as cesium that automatically adjusts itself so that on the average there are equal numbers of positive and negative particles in the region between the electrodes.

The theoretical limit on the efficiency of thermionic devices is about 40%. The first cells having efficiencies of over 10% were constructed in 1960 by a number of different groups. Of particular interest was the nuclear-heated cell that was successfully tested at Los Alamos, N.M., atomic energy laboratory in 1958. The cathode consisted of uranium-zirconium carbide and, when it was heated to bright incandescence by placing it in the Omega West reactor, it produced 20 amp. at 0.5 v.

The heat that must be supplied to evaporate an electron is

slightly greater than the work function of the cathode. The heat wasted when the electron enters the anode is the work function of the anode. Thus, the energy available to be transformed into electricity is the difference between the work functions of the cathode and anode. At first sight, it might be thought that by making the work function of the anode very small it might be possible to get an efficiency greater than that of the Carnot cycle, since there would be no waste heat. However, further analysis shows that if the work function of the anode is made too small, electrons will boil out of this electrode and go backward through the system even though the anode is maintained at lower temperature than the cathode.

Both thermoelectric and thermionic converters are low-voltage, high-current devices. A typical 100-w. thermoelectric stage produces 1,000 amp. at $\frac{1}{10}$ v., while a typical 100-w. thermionic cell produces 100 amp. at 1 v. Units must be arranged in series and parallel to produce the desired output voltage and current.

Some Other Direct-Conversion Devices.—The vacuum phototube is very similar to a thermionic device, except that the electrons are liberated from the cathode in a quantum process in which a light photon gives up all of its energy to free the electron and increase its kinetic energy. Those electrons liberated near the surface which have energies greater than the work function can escape. The solid-state analogue to a photocell is P-N semiconducting junction. At a P-N semiconductor junction the extra electrons and holes that are formed during the absorption of light are separated by the internal electrical fields existing in the semiconductor. These internal fields are analogous to the electrical fields brought about by the difference in work functions of the anode and cathode in a thermionic device. The thermodynamic efficiencies of semiconducting P-N junctions are potentially quite high for solar-energy application, since the temperature of the source of radiation is at 6,000° C. and the temperature at which the heat is rejected is around room temperature. The most efficient solar cells used on the early artificial satellites to supply a few watts of electric power were between 10% and 15% efficient.

A thermally regenerative fuel cell represents another class of direct-conversion devices, and a typical cycle is illustrated in fig. 1. Here lithium and hydrogen are burned in the fuel cell at a temperature of 450° C. The lithium hydride formed is then thermally decomposed at a higher temperature, about 850° C., to produce free lithium and hydrogen, which are returned to the fuel cell for recombination. While such cycles are of great theoretical interest they have not yet proved to be of practical importance.

Another class of device that operates at very high temperatures uses the principles of magnetohydrodynamics. A gas is heated to an extremely high temperature, about 2,500° C., so hot in fact that it becomes not only incandescent but partially ionized. In this state it is a very good conductor of electricity. The hot gas is expanded through a nozzle so that the kinetic energy of agitation is transformed into directed energy of motion. The hot conducting gas moves through a transverse magnetic field and gives rise to an electrical voltage and current in a manner exactly analogous to the copper wire in a generator moving through a magnetic field. However, as the gas cools it loses its ability to conduct electricity, and it appears that in practice such devices will be useful only for the very high-temperature part of a cycle.

The last class of devices to be considered is one in which changes of electrical or magnetic properties as a function of temperature

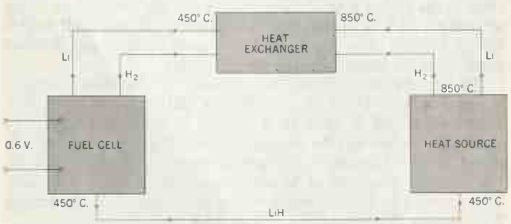


FIG. 1.—THERMALLY REGENERATIVE FUEL CELL

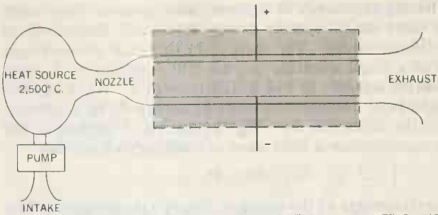


FIG. 2.—MAGNETOHYDRODYNAMIC GENERATION (MAGNETIC FIELD IS DIRECTED AWAY FROM THE VIEWER PERPENDICULAR TO THIS PAGE)

are utilized to generate electricity. When a substance goes through a phase change there are discontinuous changes in physical properties. As an example, consider a condenser containing a dielectric in which there is an abrupt decrease in dielectric constant as the temperature is increased. Suppose that the condenser is charged to a given potential V and removed from the charging circuit and heated. Since the charge Q is constant, then as the capacity decreases the potential must increase (because $Q = CV$ where C is the capacity). If it is considered that there is a discontinuity in dielectric constant with temperature, it appears that very small temperature changes can produce large amounts of work in a reversible manner, thus violating the principles of the Carnot cycle. However, careful analysis shows that the transition temperature is a function of the applied electrical field and that there is no abrupt increase in the energy of the system. Devices using this principle are not very efficient since the heat that goes into heating and cooling the crystal lattice as the device is taken through a cycle is entirely wasted, and this heat is large in comparison with the heat used to change the electrical energy of the system.

See also SOLAR ENERGY, UTILIZATION OF. (P. H. M.)

ENESCO (ENESCU), **GEORGES** (1881–1955), Rumanian violinist and composer, known for his interpretations of Bach and his works in a Rumanian national style. Born at Liveni, near Dorohoiŭ, Aug. 19, 1881, at seven he went to the Vienna Conservatory where he studied violin with Joseph Hellmesberger. In 1894 he became acquainted with Brahms, whose formal symphonic developments he later took as a model. In 1895 he went to Paris where he studied composition with Massenet and Fauré and violin with Martin Marsick. In 1897 his *Poème roumain* was played in Paris and in 1899 he won the first prize for violin at the Conservatoire. He then began his career as a virtuoso violinist and became widely known for his interpretations of Bach. He was also known as a conductor. He died in Paris, May 4, 1955.

Though Enesco spent most of his life in France, his music shows little French influence. His chamber works include three violin sonatas, the last "in the popular Rumanian style," three piano sonatas and two string quartets. Among his orchestral works are three symphonies, two *Rumanian Rhapsodies* and an overture on Rumanian folk themes. National themes are also used in his opera *Œdipe* on a libretto by Edmond Fleg adapted from Sophocles (Paris Opéra, 1936). In his later years he became the leader of the Rumanian school of composers. He was also greatly respected as a violin teacher, and Yehudi Menuhin, one of his outstanding pupils, paid great tribute to his influence.

See F. Brulez, *Georges Enesco* (1947); B. Gavoty, *Les Souvenirs de Georges Enesco* (1954).

ENFANTIN, BARTHÉLEMY PROSPER (1796–1864), eccentric French sociologist, one of the leaders of the Saint-Simonians, was born in Paris on Feb. 8, 1796, the son of a banker from Dauphiné. After studies at the École Polytechnique from 1813, he went to Russia in 1821 to enter a banking house in St. Petersburg, but returned to Paris in 1823 and became cashier in the Caisse Hypothécaire. In 1825, however, he made friends with Olinde Rodrigues, who introduced him to the moribund Claude Henri de Rouvroy, comte de Saint-Simon (*q.v.*). Enfantin was then converted to Saint-Simonism, and he soon became an acknowledged leader of the school. In the 1830s he was contributing Saint-Simonian articles to *Le Globe* newspaper and helping to organize missions to Lyons, Grenoble and Nancy. As it made

wealthy converts, the school was able to move its Paris headquarters to the former Hôtel de Gesvres, rue Monsigny. There Enfantin and Saint-Amand Bazard were proclaimed *"pères suprêmes."* A schism, however, took place when Enfantin persisted in maintaining that free love was legitimate in a collectivist society and should replace marriage for those who found marriage a tyranny, but that society should have a "living" law, dependent on arbitrary decisions, instead of a written one. Bazard's withdrawal left Enfantin as sole *"père"* of the remaining members, who now formed a church, intent on the reform of the moral rather than of the political order. New converts joined, and Enfantin claimed to have 40,000 followers. In May 1832, the government closed the sect's halls and prosecuted him and some of his followers. He retired to his estate at Ménilmontant, near Paris, with 40 disciples, all men, to live a life based on a monastic pattern. In Aug. 1832 he was arrested again and this time was imprisoned. Released in 1833, Enfantin went with some disciples to Egypt, partly to pursue the project of a Suez canal, partly to go in search of "la mère," as the union of east and west would be symbolized in the marriage of the "father" and the "mother." In 1837 he returned to France, having failed in both his aims. In 1839, however, he became a member of a scientific commission on Africa and visited Algeria; and in 1845 his knowledge of railways was recognized when he was made a director of the Paris-Lyons railway. In 1848, he established, with C. Duveyrier, *Le Crédit*, a daily journal (discontinued in 1850), in which he called on the government to promote bank loans for public works and for education and also to enable peasant smallholders to obtain credit on reasonable terms (*crédit foncier*). Enfantin died in Paris on the night of Aug. 31–Sept. 1, 1864. Among his works are his contribution to *Doctrine de Saint-Simon* (1829; republished 1831 and 1854); *Économie politique et politique Saint-Simonienne* (1831); *Correspondance politique* (1835–1840); *Correspondance philosophique et religieuse* (1843–45); *Colonisation de l'Algérie* (1843); and *La Vie éternelle, présente, future* (1861).

BIBLIOGRAPHY.—G. Weill, *L'École Saint-Simonienne, son histoire, son influence, jusqu' à nos jours* (1896); S. Charléty, *Histoire du Saint-Simonisme* (1896) and *Enfantin* (1931); H. R. d'Allemagne, *Les Saint-Simoniens* (1930). (X.; B. So.)

ENFIELD, a municipal borough (1955) of Middlesex, Eng., which under the London Government Act 1963 was to be amalgamated on April 1, 1965, with Edmonton and Southgate to form the new London borough of Enfield (*see* LONDON). Enfield municipal borough, 11 mi. N. of London bridge, extends westward from the River Lea Navigation to Barnet. Pop. (1961) 109,524. Area 19.4 sq.mi. The New river, cut in 1613 as a means of increasing London's water supply by Sir Hugh Myddleton runs south through Enfield. The town is also on the Great Cambridge road. Enfield returns two members to parliament. The district, which includes Enfield Town in the centre, Forty Hill and Bull's Cross in the north, Ponder's End, Enfield Highway and Enfield Lock in the east and the Chase in the west, grew rapidly as a residential area, while retaining in many parts its rural character. There are electrical engineering, cable and gas works, an electric power station and rolling mills. At Enfield Lock is the long-established royal small arms factory. The parish church of St. Andrew, mainly Perpendicular with Early English portions, contains 15th- to 18th-century monuments and brasses. The boys' grammar school was founded in 1557 and there is a county school for girls.

Enfield Chase was for several years the home of Charles and Mary Lamb. Once a royal preserve, it was disafforested in 1777 and was largely enclosed in the years thereafter. Several fine mansions in splendidly timbered parks remain: Trent Park, built by George III's doctor, Sir Richard Jebb, is a training college for teachers; Forty hall, built in 1629–36 by Sir Nicholas Raynton from designs by Inigo Jones, is the property of the district council; and Whitewebbs, suspected as the scene of the hatching of the Gunpowder plot, is a home for old people. The borough council controls more than 950 ac. of parks and open spaces, including Hilly Fields (65 ac.), Bush Hill (27 ac.) and Enfield Town (23 ac.) parks, while 2,508 ac. of the London green belt are in the district.

An Anglo-Saxon derivation signifying "forest clearing" is given for the name Enfield. The principal manor of Enfield, which was held by Asgar, Edward the Confessor's master of horse, was in the hands of Geoffrey de Mandeville at the time of Domesday and passed to the Bohun family in the 12th and 13th centuries. It came by succession and marriage into the possession of the crown under Henry IV and was included in the duchy of Lancaster. Of the other manors, Worcesters came to the crown in the time of Henry VIII, whose children resided at the manor house, Elsynge hall. Elizabeth I was a frequent visitor to Enfield palace (rebuilt for her by Edward VI) both before and after her accession.

ENFILADE, a military term for raking fire directed down the long axis of an enemy position or body of troops. (It is a French word, from *enfiler*, "to thread," and so, to pass through from end to end.) Against a column of troops, enfilade would be frontal; against a line of troops, flanking. This species of fire is most demoralizing since only a fraction of the enemy forces can reply to it, the fire of the rest being masked by their own troops. Against entrenchments or the parapets of fortifications, enfilade is still more effective, as the enemy is deprived of the protection given by his works. Banks of earth (traverses), built at right angles to the line of defense, are usually employed to protect against enfilade.

ENGADINE (Ger. ENGADIN; Italian ENGADINA; Romansh ENGIADINA), the Swiss portion of the upper Inn valley, extends from the source to Finstermünz, in the canton of Graubünden (Grisons).

The length of the valley from the Maloja plateau (5,938 ft.) to Finstermünz (3,406 ft.) is about 60 mi. All the villages except Sils lie at a higher altitude than the plateau; the highest is St. Moritz (6,036 ft.). The upper valley contains several lakes, including Sils, Silvaplana and St. Moritz. Administratively the plateau and the southwestern half of Sils belong to the commune of Stampa in Bregaglia administrative district. Sils (Segl) is the first village in Engadine proper. The latter is divided into two administrative districts—Upper Engadine, with 11 communes and Lower Engadine, comprising the three districts of Sur Tasna, Ramosch and Suot Tasna, with 12 communes. The 1960 population was 16,849, the majority being Protestant and Romansh speaking. The chief town of Upper Engadine is Samedan; of Lower Engadine, Scuol (Schuls).

Four major passes lead into Upper Engadine: Maloja (5,955 ft.), Julier (7,493 ft.), Albula (7,585 ft.) and Bernina (7,621 ft.). Other passes are of local importance. Lower Engadine can be reached from Davos over the Flüela pass (7,818 ft.) and by the Ofen pass (7,050 ft.). From Martinsbrück (Martina) a road leads into the Inn valley.

In Upper Engadine the Inn flows in a straight, almost level trough (26 mi. long by ½–1 mi. wide) which is bounded southeast by the glacier-clad Bernina group (Piz Bernina, 13,284 ft.) and by the range which separates the valley of the Inn from the Livigno. Northeast, and also northwest, its boundary is the Albula group (Piz Kesch, 11,214 ft., Piz Vadret, 10,594 ft.). In the north rises the Silvretta group (Piz Linard, 11,191 ft.), and to the east and south are ranges on either side of the Ofen pass (Piz Sesvenna, 10,515 ft.). Because of high elevation and northerly exposure, the scenery of Upper Engadine is stark. Firs and larches flourish, garden vegetables grow only in sunny spots and there is no tillage; alpine flora is rich and varied. The chief resources are the hay meadows, pastures and extensive forests. In Lower Engadine tillage is more varied and grain is grown. There are important reserves of water power.

Tourism is a major industry. The mineral waters of St. Moritz (*q.v.*) have been famous for centuries; in the 19th century Upper Engadine became fashionable as an "air-cure" and winter-sports centre. Lower Engadine has two major tourist attractions: the mineral springs of Scuol, below Tarasp, and the Swiss National park, a wildlife sanctuary (54 sq.mi.). The villages are built of stone, as a precaution against the destructive fires which were frequent in this narrow wind-swept valley. In Upper Engadine the villages lie on the floor of the valley; in Lower Engadine many are perched on terraces high above the river bed and are

cut off from each other by narrow gorges. The Engadine house is architecturally interesting for its "chaminada" (larder), "balcun tort" (oriel) and graffito decoration with mottoes and heraldic devices.

The Engadine was already settled in the Bronze Age; finds have been made in St. Moritz and isolated bronze artifacts have also been discovered elsewhere. The pillars on the Julier pass between Oberhalbstein and the Engadine indicate the existence of a Roman sanctuary there. During the 4th and 5th centuries, territorial sovereignty passed into the hands of the bishop of Coire. In 1139 he became powerful by the acquisition of vast estates, with the Planta family of Zuoz as his bailiffs. The bishop later became a prince of the Holy Roman empire and allied himself with the Habsburgs. In order to check his power the League of God's House (*Gotteshausbund*) was founded in 1367. The bishop finally entered the league in 1392. Lower Engadine was not entirely an ecclesiastical domain; in the 8th century Charles the Great assigned it with the Vintschgau to the counts of Toggenburg. At that time the area was absorbed into what came to be known as the county of Tirol. In 1363 the Tirol passed into the hands of the Habsburgs.

The Raetian leagues were formed, and joined with the *Gotteshausbund* to curtail the power of the Habsburgs; they fought together and defeated the Habsburgs at the battle of Calven in 1499. Lower Engadine officially then achieved independence but it was not until 1652 that this was finally secured. During this period (1801–03) Raetia was established as the 16th canton of the Helvetic Republic and changed its name to Grisons or Graubünden. Because Tarasp remained in the hands of the Habsburgs until 1803 its inhabitants adhered to Roman Catholicism. Easier communication with the Tirol caused Samnaun to abandon Protestantism and Romansh speech for Catholicism and German speech. The upper Spöl valley (Livigno) is Italian. *See* VALTELLINA; *see* also GRAUBÜNDEN; SWITZERLAND. (P. JU.)

ENGAGED COLUMN, in architecture, a structure taking the form of a column partly embedded in a pier or wall, a half or three-quarter column attached to a pier or wall face. Engaged columns are usually built integrally with the wall of the building of which they form a part, but occasionally, because of the use of the lathe, they are made of separate stones, and in woodwork they are almost always separate pieces glued or nailed in place. Engaged columns occurred rarely in Greek work, as in the temple of Zeus at Agrigentum (*c.* 500 B.C.). They were, however, common in Roman buildings. They were occasionally used decoratively in the walls of a temple that had a colonnade only at the front, as in the temple of Fortuna Virilis, in Rome (*c.* 40 B.C.) and the Maison Carré at Nîmes (time of Tiberius). Most commonly they were used to decorate the piers of arcades, as in the Colosseum at Rome (A.D. 80), and this combination of arch and engaged column is sometimes termed the Roman order. In medieval architecture, engaged columns frequently formed the component parts of clustered piers. They are common in all Renaissance styles.

ENGEL, ERNST (1821–1896), German statistician whose statistical studies were pioneering efforts to analyze economic behaviour systematically, was born in Dresden on March 16, 1821. Engel's lifelong interest in economic statistics was generated initially by the budget studies of the French engineer Frederic Le Play and the views of the Belgian statistician Adolphe Quételet. His name is usually associated with his generalization concerning the relationship between income and expenditure on food. Engel's law states that the poorer a family the greater proportion of its income is spent on food. His conclusion was based on a budget study of 153 Belgian families and later was verified by a large number of studies of consumer behaviour. Engel made a number of other contributions to economic statistics. He made several studies of the cost of rearing children. His study of the Prussian rye harvest and the average price of rye for a number of years prior to 1860 is probably the first empirical study of the relationship between price and quantity that underlies the theory of demand in contemporary price theory. Engel was head of the statistical department of Saxony between 1854 and 1858, and from 1860 to 1882 he headed the Prussian statistical department in Berlin. He died at Radebeul, near Dresden, on Dec. 8, 1896. (D. C. N.)

ENGELS, FRIEDRICH (1820–1895), social philosopher and businessman, the closest collaborator of Karl Marx, was born on Sept. 28, 1820, at the Rhineland industrial town of Barmen, the son of a well-to-do cotton manufacturer. His formal education terminated with the *Gymnasium*, after which he continued to study and write in his free time while he was a commercial apprentice, Prussian military recruit and, from 1849 to 1869, successively an employee, partner and director of a textile firm founded by his father in Manchester, Eng. He retired at age 49 to devote himself to political activity and writing.

Engels' principal contribution to modern thought is the attempted proof that Marxian dialectical materialism, the historical rather than metaphysical interpretation of passing events, is implicit in the theory and practice of modern science. Progress in human understanding, Engels argues, is based on conscious awareness of the interrelatedness and, consequently, the reciprocal determination of all specific forms of organization, natural and social, in an endlessly changing process of "history in the making." Any concrete physical phenomenon or social institution is fixed only in a relative sense; *i.e.*, pending the cumulation of internal tensions to a critical point, at which change occurs. All events being interrelated, no one event or phenomenon can be understood outside the total complex of natural and social evolution.

Uppermost in Engels' mind is the career of capitalism. To accumulate private capital, men of business find it profitable to specialize. The economy-wide effect is an increasingly productive *social* division of labour, not planned, however, to preserve overall balance of expanding production, consumption, aggregate saving and investment. Unco-ordinated outpouring of added productivity creates, for all, incalculable uncertainties of inflation or recession. Characteristic of Engels is his prediction of the obsolescence of war as an instrument of national policy in a world progressively unified through industrialization. In literature and art, Engels calls for a broad realism that records local and temporal stresses of social change but interprets them in the hopeful perspective of the human race in the process of unification.

Engels' effective collaboration with Marx began in 1844 (*see* MARX, KARL HEINRICH), taking the form, in 1848, of the *Communist Manifesto*, prepared for the propagandistic *Bund der Kommunisten*. Engels in 1878 published *Anti-Duehring: Socialism, Utopian and Scientific*, a work that is said first to have clarified the essential bases of Marxist reasoning. At his death, in 1895, Engels was preparing the general scientific treatise *Dialectics of Nature*, posthumously published in English translation in 1940, with an introduction by J. B. S. Haldane.

Engels' publications include: *Herr Eugen Duehring's Revolution in Science* (Anti-Duehring); *Dialectics of Nature; Ludwig Feuerbach and the End of Classical German Philosophy; The Housing Question; The Origin of the Family, Private Property and the State; The Peasant War in Germany; On Marx's Capital*—"synopsis, reviews, letters and supplementary material"; *Karl Marx and Friedrich Engels—Correspondence 1846–1895*. For a complete list of books see the entry, Engels, Friedrich, in *A Catalog of Books* . . . Library of Congress, now superseded by the National Union Catalog. *See* also MARXISM.

See the authoritative biography by Gustav Mayer, *Friedrich Engels, a Biography*, introd. by G. D. H. Cole (1936); Z. K. Coates, *Life and Teachings of Frederick Engels* (1945). (G. W. Z.)

ENGELS, a town of Saratov *oblast* of the Russian Soviet Federated Socialist Republic, U.S.S.R., stands on the low, left bank of the Volga, opposite Saratov. Pop. (1959) 90,000. Founded in 1747 as the peasant colony Pokrovskaya Sloboda (the name was changed in 1931 in honour of Friedrich Engels; *q.v.*), the town became one of the main centres of the Trans-Volga and from 1922 to 1941 was the capital of the former German Volga republic (*q.v.*). It is still a centre for processing the agricultural products of the region, especially meat. Rolling stock and trolley buses, diesel motors, glue, leather goods and saw-timber are made. An artificial textile mill was constructed after 1960. Many workers commute to Saratov. Engels is connected with the railway from

Saratov to Astrakhan and to Orenburg, which with the navigable Volga gives excellent communications. (R. A. F.)

ENGHIEN, LOUIS ANTOINE HENRI DE BOURBON-CONDÉ, Duc d' (1772–1804), French prince whose execution marked the end of all hope of reconciliation between Napoleon and the royal house of Bourbon and was widely condemned as an atrocity, was born at Chantilly on Aug. 2, 1772, the only son of the duc de Bourbon (Louis Henri Joseph; *see* Condé, Princes de) and Louise Marie Thérèse d'Orléans. Emigrating with his father at the outbreak of the Revolution, he served in his grandfather's army of *émigrés* from 1792 till its dissolution after the peace of Lunéville. He then secretly married Charlotte de Rohan-Rochefort and settled at Ettenheim in Baden, near the Rhine.

Early in 1804 Napoleon, then first consul of France, heard news which seemed to connect the duc d'Enghien with the conspiracy of Georges Cadoudal and Charles Pichegru, then being investigated by the French police. The news was that the duc was in company with C. F. Dumouriez (*q.v.*) and made secret journeys into France. This was false; the acquaintance was the harmless Thumery; and the duc had no dealings with Cadoudal or Pichegru. Napoleon, however, gave orders for his seizure. French mounted gendarmes crossed the Rhine secretly, surrounded his house and brought him to Strasbourg (March 15, 1804) and thence to the castle of Vincennes, near Paris. There a military court-martial was hastily gathered to try him. The duc was charged chiefly with bearing arms against France in the late war and with intending to take part in the coalition then proposed against France. The officers hastily and most informally drew up the act of condemnation, on instructions brought by Savary.

Though Josephine and Mme de Rémusat begged Napoleon for mercy, the duc was shot in the moat of the castle, near a grave already prepared, on March 21, 1804. With him ended the house of Condé.

The famous comment on the execution, "C'est plus qu'un crime, c'est une faute" (It's more than a crime, it's a mistake), because of the indignation that it provoked, is often ascribed to Talleyrand, but is also claimed for Joseph Fouché and for A. Boulay de la Meurthe.

See A. Boulay de la Meurthe, *Correspondance du duc d'Enghien et documents sur son enlèvement et sa mort,* 3 vol. (1904–10); B. Melchior-Bonnet, *Le Duc d'Enghien* (1954). (G. de B. de S.)

ENGHIEN (Flemish, Edingen), a town of Hainaut province, Belg., lies 30 km. (18¾ mi.) S.W. of Brussels by road. Pop. (1961) 4,225. It has a dungeon (1166) which serves as a chapel, and a Capuchin church with the alabaster mausoleum (1521) of Guillaume de Croy, archbishop of Toledo and primate of Spain. The title "duc d'Enghien" was long borne by the eldest son of the princely house of nearby Condé. The town is a junction on the Brussels-Tournai railway. It has lace and pharmaceutical manufactures. (R. M. An.)

ENGINE: *see* Aircraft Propulsion; Air Engine; Automobile; Diesel Engine; Internal-Combustion Engine; Marine Engineering; Refrigeration; Rockets and Guided Missiles; Solar Energy, Utilization of; Turbine.

ENGINEERING (ARTICLES ON). Engineering is sometimes described as applied physics, as distinguished from theoretical or "pure" physics. But this tells only part of the story. For example, the particle accelerator, an engineering development made possible by nuclear physics, has in turn multiplied knowledge in the latter field. In many respects engineering, which has created the physical structure of our civilization, holds the key to its future.

To provide insights into the achievements and objectives of engineering for the general reader, basic principles of the sciences of engineering have been emphasized and technical detail reduced to a minimum in a group of introductory surveys. These include Engineering, which defines the scope of the profession, and Agricultural Engineering; Architectural Engineering; Automation; Engineers, Military; Industrial Engineering; Industrial Revolution; Inventions and Discoveries; Machine Tools; Management Sciences; Marine Engineering; Mass Production; Nuclear Engineering; Physics: General Survey; Servomechanism; and Space Exploration.

The articles cited in the following paragraphs, while mainly of a technical nature, contain introductory sections, and in some cases historical sections, in which general principles are outlined in nontechnical terms. Among articles in which various types of motive power are described are Aircraft Propulsion; Automobile; Diesel Engine; Internal-Combustion Engine; Marine Engineering; Motor, Electric; Railway; Rockets and Guided Missiles; Steam; and Turbine.

Articles dealing with applications of engineering to atomic science include Accelerators, Particle; Atomic Energy; Bubble Chamber; Cloud Chambers; and Nuclear Instruments.

Important phases of electrical and electronics engineering are represented by such articles as Broadcasting; Computing Machines, Electronic; Generator, Electric; Electron Tube; Lighting; Radar; Radio; Television; Transistor; etc.

Architectural Engineering is supplemented by such articles as Brickwork; Bridges; Carpentry; Drawing, Engineering; Foundations, Structural; Steel Construction; Tunnel.

Applications of various forms of engineering to waterways and related installations are treated in Breakwater; Dam; Dock; Harbours; Jetty; Lighthouses and Other Aids to Navigation; River and River Engineering; Shipbuildng (Merchant and Naval); Water Transport, Inland; etc.

Articles on hydraulic engineering include Hydraulics, Applied; Hydromechanics; and Mechanics, Fluid.

Individual phases of chemical engineering are described in such articles as Alkali Manufacture; Electrochemistry; Gas Industry; Hydrogenation; Petroleum; Plastics; etc. Industrial applications of chemistry are also described in articles on the individual elements—for example, Antimony; Helium; Neon; Oxygen; Selenium; and Uranium.

Mining Engineering gives references to key articles.

Articles on acoustical engineering include Acoustics of Buildings; High-Fidelity Sound Systems; Loud-Speaker; Microphone; Phonograph; and Tape Recording, Magnetic.

Among articles dealing with sanitary engineering are Incinerator; Plumbing; Sewage Disposal; Sewer Design and Construction; Smoke and Atmospheric Pollution; and Water Supply and Purification.

Great exponents of the engineering sciences are represented by biographical articles.

Additional listings of articles on specialized fields of engineering will be found in Aeronautics (Articles on); Agriculture (Articles on); Architecture (Articles on); Chemistry (Articles on); Geology (Articles on); Physics (Articles on); and Transport.

The articles cited in the second paragraph above are recommended as a minimal reading list by which the engineering sciences can be seen in the perspective of present-day industrial culture. Two additional articles should be consulted for their discussions of engineering in relation to the history of world civilization: Civilization and Culture and Material Culture.

The articles listed above are intended merely as illustrations of the treatment by *Encyclopædia Britannica* of the major branches of engineering. For comprehensive lists of articles and sections of articles devoted to specific topics, the Index should be consulted.

ENGINEERING has been defined by the Engineers Council for Professional Development (U.S.) as the creative application of "scientific principles to design or develop structures, machines, apparatus, or manufacturing processes, or works utilizing them singly or in combination; or to construct or operate the same with full cognizance of their design; or to forecast their behaviour under specific operating conditions; all as respects an intended function, economics of operation and safety to life and property." The term is sometimes more loosely defined, especially in Great Britain, as the manufacture or assembly of engines, machine tools and machine parts, including instruments and associated measuring and control devices. It is also commonly used where the words

engineering science would be more appropriate—for instance, in describing a course of study undertaken by university students.

In its early uses engineering referred particularly to the construction of engines of war and the execution of works intended to serve military purposes. Military engineers were long the only ones to whom the title "engineer" was applied. About the middle of the 18th century there arose a new class of engineers who concerned themselves with such works as the making of roads, which were neither exclusively military in purpose nor executed by soldiers. By way of distinction those men came to be known as civil engineers. Their aims and functions were given in the charter (1828) of the Institution of Civil Engineers (London), which describes civil engineering as the "art of directing the great sources of power in nature for the use and convenience of man, as the means of production and of traffic in states, both for external and internal trade, as applied in the construction of roads, bridges, aqueducts, canals, river navigation and docks for internal intercourse and exchange, and in the construction of ports, harbours, moles, breakwaters and lighthouses, and in the art of navigation by artificial power for the purposes of commerce, and in the construction and adaptation of machinery, and in the drainage of cities and towns." Wide as is this enumeration, the practice of a civil engineer in the earlier part of the 19th century might have covered many or even most of the subjects it contains.

Gradually more specialized branches of engineering emerged. Perhaps the first of these to be recognized was mechanical engineering, which deals with engines, machine tools, millwork and moving machinery in general. Mining engineering, concerned with the location and working of coal, metallic ores and other minerals, soon followed as a distinct specialty. Subsequently, many other subdivisions of engineering became delineated. Some (civil, mining and metallurgical, mechanical, electrical, chemical, aeronautical and industrial) may be readily recognized through appropriately named professional societies and school curriculums. Even more specialized branches have been identified as sanitary, structural, drainage, hydraulic, highway, railway, electric power, electrical communications, steam power, internal combustion, marine, welding, production, petroleum production, fire protection, safety, architectural, nuclear, and sales, management or administrative.

Many engineers assume managerial or administrative responsibility, since their training and experience encompasses handling men, money, methods and materials.

In recent years engineering increasingly has gained respect as a profession. The training of its practitioners is described under ENGINEERING EDUCATION.

Some engineers feel strongly that the art of engineering—the knowledge of actual practice and the skills for its utilization—as well as the application of scientific principles is an essential part of engineering. The need for men who can view engineering, wide and complex as it is, as a single field of operation with relatively few basic laws and methods is increasingly recognized. Such men can work well with people from other disciplines and, when they gain sufficient experience and judgment, can successfully plan and direct vast enterprises.

See also references under "Engineering" in the Index volume.

(L. E. G.; W. E. HD.; J. F. BR.)

ENGINEERING EDUCATION. The education of an engineer extends over a wide range of knowledge: from pure science, and especially what is known as engineering science, to technology. The major portion of the field is covered by the following branches: aeronautical, agricultural, chemical, civil, electrical, industrial, mechanical, metallurgical, mining and geological, and nuclear engineering. There is great variety and a good deal of specialization. A civil engineer may aim at highway engineering, structural engineering or some other branch, and his education will be influenced to some extent by his choice. Similarly a mechanical engineer may aim at automotive, machine-tool, aeronautical or general production engineering; and an electrical engineer at heavy-current work such as power supply, at light-current work such as telephone communications or at work in the field of electronics.

Modern engineering demands a sound training in general science, particularly in physics, mathematics and chemistry. For certain industries a basic knowledge of biology is also essential. The relative importance of the fundamental sciences depends on the branch of engineering; for example, an electrical engineer who wishes to specialize in communications or electronics needs an extensive knowledge of physics and mathematics, whereas for an agricultural engineer, training in chemistry and the biological sciences is more important.

Views on engineering education in the latter half of the 20th century emphasized the need for a broadening of the curriculum and for an understanding by engineers of the social implications of their work. It was recognized that—especially in management posts—the ability to deal skilfully with problems of human relations was sometimes as important as technical knowledge.

History.—Although the engineering profession can be traced back into earliest times, the records of history do not appear to contain any mention of the formal education of engineers until 1747, when the French engineer Jean Rodolphe Perronet was given the task of instructing "designers in the sciences and practices needful to fulfilling with competency the different occupations relating to . . . bridges and highways." The Corps des Ponts et Chaussées, which he headed, had been established in 1716 and the school established in 1747 for the workers in the organization became the École Nationale des Ponts et Chaussées, the first engineering school. It has been said that the French were the leaders in engineering in the 17th and 18th centuries; they were not only the leaders but also the pioneers in engineering education. In 1778 there was founded the school that later became the École Nationale Supérieure des Mines; in 1794 the École des Travaux Publics, which became the École Polytechnique; and in 1788 the École d'Arts et Métiers. The École Centrale des Arts et Manufactures was established in 1829.

The early development of engineering education in America was greatly influenced by these French models. The curriculum of 1849 of the Rensselaer Polytechnical institute, Troy, N.Y., which became an almost universal prototype in the United States, was largely adapted from the course of study in the École Centrale.

In Great Britain, the pioneers of industry were workingmen trained by apprenticeship and self-taught in industrial science, and the leading engineers of the early period were themselves products of the old system of indenture and pupilage, the accepted form of professional training. Only three of the first ten presidents of the Institution of Civil Engineers—the all-powerful British engineering society (founded in 1818)—were university-trained men, and they were Scots. But this organization now places the sanction of the profession clearly behind the system of higher education in engineering, including postgraduate study and research.

In Germany, the polytechnic school at Karlsruhe was created in 1825 from two older institutions; it was the first of its kind in that country. Its curriculum was changed in 1833 to emphasize a "high scientific discipline" and the school became essentially professional.

After 1871 technical schools developed rapidly in Europe and England and in the United States to serve the needs of the growing industries that burst into activity at this time. (W. E. HD.)

UNITED STATES

The first engineering school in the United States was the U.S. Military academy at West Point founded in 1802. However, it was not until 1817, with the coming of Sylvanus Thayer as superintendent, that the academy took on stature in the engineering field. Colonel Thayer patterned the instruction on French models. He brought Claude Crozet, a graduate of École Polytechnique, to be the first professor of civil engineering in America.

Rensselaer Polytechnic institute was founded in 1824; it granted its first engineering degree in 1835. Norwich university in Vermont was established in 1819 and is said to have offered engineering in the early days; the details of its courses, however, are no longer available. As mentioned above, Rensselaer remodeled its curriculum in 1849, basing it for the most part on the course of study in the École Centrale. Union college, Schenectady, N.Y., offered some courses in civil engineering in 1845, and Brown university, Providence, R.I., in 1850. In 1852 the University of Michigan added civil engineering to its curriculum.

All schools in the early years confined their engineering instruction to civil engineering. In 1864 Columbia university founded its school of mines. The Massachusetts Institute of Technology opened in 1865 and offered programs in civil, mechanical and mining engineering and in what was called "practical chemistry."

Three "scientific" schools, namely, Lawrence (1847), Sheffield (1847) and Chandler (1851), were established as adjuncts respectively of Harvard, Yale and Dartmouth colleges. These institutions very carefully kept the "scientific" program separate from their regular collegiate curriculums. Chandler later became Thayer school, the present engineering division of Dartmouth. Lawrence was practically abolished in 1907–08; Harvard university continues to conduct engineering instruction mainly at the graduate level.

The rapid expansion of engineering education in the United States after 1850 is attributable to the vision of a few persons who realized that the country's rich resources demanded engineers and technicians capable of exploiting them fully. The United States congress in 1862 passed the Morrill act, an "Act donating Public Lands to the several States and Territories which may provide Colleges for the benefit of Agriculture and the Mechanic Arts" (*see* LAND-GRANT COLLEGES AND UNIVERSITIES). The half dozen engineering schools of 1860 increased to more than 80 by 1880—some in land-grant colleges, others established as a result of private or community enterprise. These schools trained men who could apply the increasing wealth of scientific knowledge to practical problems, and they had a large part in transforming the industrial, economic and social life of the nation.

Technical Institutes.—An important phase of technical training in the engineering field is the so-called technical institute program. In technical institutes students are trained for immediate employment in a relatively restricted activity, usually in a two-year program. Graduates supervise industrial processes, perform detailed drafting, serve as laboratory assistants, become technical salesmen, or the like. Their training differs from that of engineers in that the emphasis is on the practical rather than the theoretical; the technical graduate knows how to perform a test or operate a machine but not necessarily how to design the test or machine.

American Society of Engineering Education.—The Society for the Promotion of Engineering Education, which was organized in 1893, became a dominant factor in shaping and co-ordinating ideas, policies and practices of engineering educational institutions throughout the United States and also to some extent in Canada. The society began with 70 members, a figure that had increased to about 10,000 by mid-20th century.

From the beginning the society sponsored significant studies of various facets of engineering education, and in 1907 it initiated a practice of requesting the professional engineering societies to collaborate in their major studies and investigations. In that year the societies of the civil, mining, mechanical, electrical and chemical engineers were invited to join it in forming a joint committee on engineering education for the purpose of exploring various aspects of the subject. In 1908 the American Society of Civil Engineers made an appropriation for the operation of a committee to study engineering education, the first recorded grant from an engineering society for a study of this sort. Because of lack of funds to complete the study an appeal was made to the Carnegie Foundation for the Advancement of Teaching. The foundation undertook the task, and in 1918 it issued a report entitled "A Study of Engineering Education" by C. R. Mann.

In 1946 this society was reorganized and its name was changed to the American Society for Engineering Education (A.S.E.E.). Its special committees have prepared and issued reports that have had a marked influence upon engineering education. Principal committees include those concentrating on selection and guidance, research, atomic energy education, recruiting of engineering teachers, and improvement of teaching.

Engineering Council for Professional Development.—Six professional societies (those for civil, mining and metallurgical, mechanical, electrical and chemical engineers, and the American Society for Engineering Education) initiated steps to form the Engineers' Council for Professional Development (E.C.P.D.) in 1932. In addition to these six societies the Engineering Institute of Canada and the National Council of State Boards of Engineering Examiners were invited to participate in the development of the organization. The council is a conference body commissioned, in its own words, "to promote efforts toward higher professional standards of education and practice, greater solidarity of the profession, and greater effectiveness in dealing with technical, social, and economic problems." The committee on education is concerned with accreditation of engineering curriculums, and the committee on guidance has devoted its attention to supplying information on engineering to secondary school students and to the determination of aptitudes for engineering study in the development of testing procedures. A committee on adequacy and standards of engineering education was also established. (G. A. H.)

GREAT BRITAIN

Though lectures on engineering topics were introduced at Cambridge university toward the end of the 18th century and mechanics institutes for the education of craftsmen were established in most of the industrial cities at about that time, it was not until the first quarter of the 19th century that university courses were introduced, first at Edinburgh and then in 1827 at University college, London. In 1838 the first complete university school of engineering was established at King's college, London, and in 1840 the first regius professor of engineering was appointed at Glasgow. The first university examination in engineering was established at Cambridge in 1865; in 1875 there followed the first Cambridge engineering professorship and the department of engineering, which soon became what is still the largest science department in that university. At about this time many of the civic universities were founded; from the start all of them maintained strong departments of engineering. At most of these universities the course takes three years for a bachelor's degree based on a unified engineering curriculum. The bachelor of arts (B.A.) is awarded at Cambridge, but at most universities the degree is that of bachelor of science (B.Sc.), with bachelor of engineering (B.Eng.) at Liverpool and Sheffield. Many universities offer postgraduate courses and some of these lead to the master of science (M.Sc.). It is most usual for the M.Sc. and the Ph.D. to be awarded as a result of research work in the university.

In Great Britain rigorous professional training is supervised by the engineering institutions (civil, mechanical and electrical being the oldest and largest). An applicant is accepted as an associate member (chartered engineer) of one of the professional institutions only after he has passed its examination in engineering science, or has been exempted both by his university degree or other qualification and his practical professional training (usually for two or three years) under a member of the institution. These institutions continually review the needs of engineering education and function as societies to which scientific and professional papers are submitted.

Since the days of the first mechanics institutes Great Britain has had a widespread and excellent system of technical colleges offering evening, part-time and full-time instruction at levels ranging from that of craftsmen up to that of engineers seeking exemption from professional institution examinations. Courses lead to the ordinary national certificate (O.N.C.) or the higher national certificate (H.N.C.) controlled by the ministry of education in collaboration with the professional body.

In the middle of the 20th century a number of colleges of advanced technology were founded providing full-time courses ("sandwich courses") in which the student devotes half the year to full-time study in the college and half to practical training in industry. Both types lead to the diploma in technology (Dip.Tech.), comparable in standard to that of a university bachelor's degree in engineering. (J. F. Br.)

EFFECTS OF TECHNOLOGICAL CHANGE

Even prior to 1940 important scientific developments began to place new demands upon the skill and knowledge of engineers in all industrially advanced countries. In order to cope with the new problems in electronics, for example, the electrical engineer needed a deeper understanding of solid-state physics. An increas-

ingly large body of scientific information was required of the civil engineer in the study of structural problems as related to welding processes, vibration control and newly encountered soil conditions. As a result of increasing interest in gas turbines and jet propulsion, mechanical engineers were confronted with new problems in heat transfer, thermodynamics and fluid mechanics, all of which required new methods of analysis and a better comprehension of physical phenomena. Metallurgy was changing from an art to a science as it came to be based ever more firmly upon the foundations of physical chemistry and solid-state physics. These developments, together with many others, caused a revolution in engineering practice and demanded the evolution of new educational programs.

The advent of new scientific and technological developments, and particularly the increased utilization of atomic energy, resulted in a need for integration of learning in the technical field, an integration beginning in the undergraduate programs of some universities. In the design of nuclear power reactors, for example, information from all areas of engineering must be brought to bear upon the problem. An engineer with a strong foundation in the fundamental engineering sciences is required. Undergraduate specialization in a limited field, as it was practised in the past, cannot equip the engineer to work in these highly complex areas; consequently, marked changes in engineering curriculums began to be introduced in the 1960s.

As modern engineering requires an ever more adequate grasp of basic engineering principles during the undergraduate period, it is inevitable that mastery of specific arts and techniques must, to a greater extent, be left to the engineer's continuing self-education after graduation. It is not possible to teach the constantly proliferating engineering arts to the student while in college. There are too many of them in any branch of engineering. The school must prepare the engineer to master later (often on the job) a good deal of specialized knowledge that the school cannot give him in detail. It becomes the lot of the employer to lead the engineer into the specialized area of engineering art that is related to the task in hand. A great many industries have readily accepted this responsibility. Engineering has arrived at the professional status in which the engineer's first few years of employment really constitute an internship that is an important part of his education.

It is recognized that some specialization is absolutely necessary, no single mind being large enough to absorb all the facts, to say nothing of an understanding of their relationships, in even one field of modern engineering. There are those who ascribe the great progress of technology to this very specialization. But the trend toward specialization reached a point of diminishing returns, and there arose a need for more men whose range of understanding covers at least the main methods of thinking and analysis in the so-called basic engineering sciences. Specialists still are needed, but even more important are men who know enough about the specialties of their associates so that they can effectively co-operate with them to solve problems that require the knowledge of many specialists.

To meet this need, as well as to broaden the scientific understanding of the specialists, various unified programs in engineering education were developed. The program at the University of California at Los Angeles, for example, was so completely integrated that the students were no longer designated as to their areas of study, such as electrical, mechanical or civil engineering.

Many engineering schools initiated programs comprising several avenues of study to suit the various interests of the students. For example, in one institution, at the beginning of the third year, the student is offered two choices: a preprofessional option for students who propose to enter the profession immediately upon receiving a baccalaureate degree and a pregraduate option for those with demonstrated aptitudes who propose to pursue further formal education beyond the baccalaureate level.

The student who plans to attend graduate school thus continues to spend the greater part of his time building a foundation for graduate study and relatively little time in acquiring more practical knowledge. On the other hand the student who is going to terminate his formal education at the baccalaureate level begins as early as his third year to study courses whose object is primarily to show him how to apply his knowledge of engineering sciences so that he may be a useful member of the profession immediately upon graduation.

Nontechnical Education.—To meet the criticism that engineering education has been too exclusively technical, continual efforts are being made to improve the student's ability to communicate his ideas and to help him acquire an understanding of social, economic and esthetic values. Many U.S. engineering schools have adopted new general-education programs to give engineering students some depth of background in the humanities and social sciences.

Growing Demands.—In the second half of the 20th century the demand for engineering education was especially great in countries that had been industrially underdeveloped and in those newly emerging as independent nations. Many students from the far east, Latin America and the middle east and increasing numbers from Africa sought engineering education in the U.S. and the U.S.S.R. (*see* EDUCATION, INTERNATIONAL). The U.S.S.R. cited itself (crediting communism) as an example of what could be accomplished by a country that had been largely bypassed by the Industrial Revolution. Although there had been some development of technical training before the Soviets came into power in 1917, higher education, including engineering education, was available only to a privileged few. By the second half of the century, the Russians were training more engineers than any other country, both in terms of actual numbers and in proportion to population. Also, students from other communist countries and from many noncommunist nations were receiving engineering and scientific training in the U.S.S.R.

In China, because of the growing demand for technically competent personnel in all fields, and because of the unevenness in different aspects of economic development, technical education was being expanded on all fronts. On the highest level, the Chinese Academy of Sciences and its branches, together with other institutions of higher learning, trained an increasingly large number of engineers, scientists and technicians in a variety of areas of specialization. Graduates of these institutions and those trained in the Soviet Union and elsewhere form the high echelon of technical and engineering personnel.

BIBLIOGRAPHY.—American Society for Engineering Education, *Report on Evaluation of Engineering Education, 1952–1955; General Education in Engineering* (1956); *Report on the Engineering Sciences, 1956–1958; Engineering Faculty Recruitment, Development, and Utilization* (1960); *Report on Objective Criteria in Nuclear Engineering Education*, 1960–1961; L. E. Grinter, "A Survey of Current Changes That Are Modernizing Engineering Education," *J. Engng. Educ.*, vol. 49, no. 7 (March 1959); G. A. Hawkins, "A Challenge to Engineering Management," *Mech. Engng., N.Y.*, vol. 81, no. 11 (Nov. 1959); E. J. Holstein and E. J. McGrath, *Liberal Education and Engineering* (1960); J. F. Baker, "Engineering Education at Cambridge," *Proc. Instn. Mech. Engrs.*, vol. 171 (1957). *See* also *J. Engng. Educ.*, vol. 43, no. 4 (Dec. 1952). (G. A. H.)

ENGINEERING INSURANCE. A British term denoting insurance covering the owners of boilers, machinery and power plant equipment. Unlike most other classes of insurance, which are designed to indemnify the insured after he has suffered loss, engineering insurance endeavours primarily to reduce the possibility of a loss occurring. To this end insured machinery is periodically inspected by skilled engineers to ascertain defects before they result in an explosion of, or breakdown in, the plant. A large proportion of the premiums received is expended upon these services. *See* CASUALTY INSURANCE.

ENGINEERS, MILITARY. The role of the military engineer in modern war is to apply engineering knowledge and resources to the furtherance of his commander's plans. The basic requirement of a military engineer is a sound general engineering knowledge directed to the technical aspects of those tasks likely to be encountered in war. Engineer work is influenced by topographical considerations and in battle also by tactical limitations. At times engineering factors will actually govern the choice of the military plan adopted; a military engineer must, therefore, possess a sound military education if he is to give the best technical advice to his commander.

In ancient and medieval times the military engineer was a specialist, a man of ingenuity who made and used engines of war such as catapults, ballistas, battering-rams, ramps, towers, scaling ladders and other devices in attacking or defending castles, fortresses and fortified camps. In peacetime the military engineer built fortifications for the defense of the country or city against foreign invasion, or for internal security. Because they frequently dug trenches or tunnels as means of approaching or undermining enemy positions they came to be called sappers or miners. With the invention of gunpowder and the countless other inventions that came in later centuries, the military engineer was required to have far more technical knowledge. He nevertheless remained a soldier and fought side by side with the infantry in many wars.

The functions of modern military engineers vary among the armies of the world but as a rule they include the following activities: (1) construction and maintenance of roads, bridges, airfields, landing strips and zones for the airdrop of men and supplies; (2) interference with the enemy's mobility by means of demolitions, floods, destruction of matériel, laying of mine fields and erection of obstacles or fortifications of many types; (3) mapping and aiding the artillery to survey gun positions, rocket launching sites and target areas; (4) supplying water and engineer equipment; (5) disposal of unexploded bombs or warheads. In the British army the royal engineers also operate the army postal service.

UNITED STATES

The United States army corps of engineers is both a combatant arm and a technical service. Alone among the arms and services it engages in civil as well as military activities. During the 20th century its civil works activities have centred upon the planning, construction and maintenance of improvements to rivers, harbours and other waterways, and upon flood control. The principal military service performed by the corps of engineers for the army and air force, in the United States and abroad, is the construction and maintenance of buildings and utilities. In theatres of operation in wartime such construction is carried out by engineer troops. In the United States, in peace and war, and overseas in peacetime, such construction, like civil works projects, is usually accomplished by private industry under contract to the corps of engineers.

The corps of engineers administers its civil and military construction duties through the office of the chief of engineers and an extensive field organization, both manned by civilians under the supervision of engineer officers. Improvement of construction techniques and equipment, with emphasis on missile bases and on construction in arctic regions, consumed the greatest portion of the corps's research and development program in the early 1960s. This program also included combat support; mapping and terrain investigations; supply support; and utilization of nuclear energy. The corps also maintained field installations for preparing and distributing maps.

Early History.—The duties and administrative organization of the corps of engineers evolved over a long period of time. On June 16, 1775, the day before the battle of Bunker Hill, the continental congress authorized the appointment of a chief engineer (Col. Richard Gridley) and two assistants for the grand army, and in Dec. 1776 the raising of a corps of engineers. Since at this time there were scarcely any trained military engineers in America, most engineer officers with the Revolutionary forces were French or Polish volunteers. In the construction of fortifications at West Point, N.Y., and of siege works at Yorktown, Va., engineer and infantry troops led by these officers and by Col. Rufus Putnam rendered outstanding service to the American cause.

With the return of peace in 1783 the corps was disbanded. Ten years later the outbreak of war in Europe focused attention on the need for more adequate coastal defenses. A concurrent demand for public improvements made the services of trained engineers all but indispensable. A number of the country's leaders envisioned a single organization capable of performing both military and civil engineering tasks. To this end congress, on March 16, 1802, reconstituted the corps of engineers, directing that it be stationed at West Point, where it was to establish and superintend a military academy. Although the headquarters of the corps was later moved to Washington, D.C., chiefs of engineers continued to exercise control over the academy's affairs until after the American Civil War. (*See* also MILITARY, NAVAL AND AIR ACADEMIES: *United States.*)

Civil Accomplishments.—The federal government soon began to employ the skills of engineer officers on projects of a civil as well as a strictly military nature. Following the War of 1812, numerous coastal fortifications were laid out and constructed. On the civil side, surveys of routes for roads and canals were uppermost at first. After the advent of railroads in the late 1820s, engineer officers rendered similar services to these enterprises. During the 1840s and 1850s explorations of the west became major engineer exploits. The reports and maps of these expeditions, which included scientists and artists, did much to replace legend by fact not only about the topography of the area but also about its plants, animals and natural resources.

The activities of the corps of engineers have ranged from the construction of public buildings in the District of Columbia to the digging of the Panama canal. At one time or another congress or federal agencies have called upon the corps to perform such diverse tasks as surveying routes for transcontinental railroads, erecting lighthouses and building roads in national parks. Over the years the engineers' unspectacular but important work of dredging harbours and channels, constructing and maintaining canals and removing obstacles to navigation has done much to stimulate the growth of water-borne commerce. In times of national emergency the vast network of inland waterways throughout the country has carried large quantities of vital war materials, thereby easing the strain on other means of transportation. The corps's success in preventing floods has reduced loss of life and saved millions of dollars in damage to crops and property. Long-range flood control programs embracing entire river systems are carried out on a mammoth scale.

At War.—Civil works naturally and properly gave way to military activities during periods when the United States was at war. In the Mexican War (1846–48) a number of engineer officers, among them Robert E. Lee, George B. McClellan and Henry W. Halleck, served with distinction. The Civil War, with its large-scale maneuvers through territory cut by numerous rivers and streams, made the Union army particularly dependent upon pontoon bridging provided by engineer troops. Some of the bridges constructed in a matter of hours were more than 2,000 ft. long. It was in World War I, however, that the corps of engineers first emerged as a sizable force. Expanding from a total of 256 officers and 2,000 enlisted men to 11,175 officers and 285,000 men, the corps made up 12% of the army. In France its most important service was the repair of roads and railroads.

The general employment of mobile tactics and the truly global nature of World War II placed a great premium on engineer combat and service operations. With a peak strength of over 700,000 officers and men, the corps of engineers comprised 8% of the army. Engineer troops were able to accomplish their job with relatively fewer men than in World War I because they were equipped with labour-saving equipment. Bridges designed to carry the heavy loads of the modern army, yet more easily transported than those employed during the Civil War, could be erected in a matter of hours (*see* BRIDGING, MILITARY). Specially cut steel mats were fitted together to form landing strips for aircraft, thus obviating the need for time-consuming paving operations. Aerial photography was utilized in conjunction with stereoscopic instruments to compile maps.

For the invasion build-up in Great Britain, engineer troops built airfields, camps, depots and hospitals. They overcame the destruction wrought by the retreating German forces in Italy by clearing ports and roads and by throwing bridges across rivers. At the Normandy landings they cleared mines and other obstacles from beaches and rolled supplies across them. Construction of port facilities on the Persian gulf and of transportation facilities inland through Iran opened a much-needed alternate route through which large quantities of lend-lease supplies reached the U.S.S.R. Opening new supply routes into China also, engineer troops constructed airfields on both sides of the "hump" and pushed the Ledo

road and the longest petroleum pipeline system in the world through the mountainous jungle. In the long advance from Australia to Tokyo, engineers manned landing craft, delivering men and matériel on island after island which they then converted into operating bases.

Before the war could be brought home to Germany and Japan, large forces had to be raised, equipped, trained and maintained in combat around the world. The demand for facilities to house and supply the expanding U.S. army called forth a tremendous construction effort on the part of private industry under contract to the war department. The military construction program was launched under the supervision of the quartermaster corps, the branch of the army traditionally responsible for troop housing and related activities in the United States. In Nov. 1940 the corps of engineers relieved the quartermaster corps of one phase of this program, the construction of airfields. All military construction was transferred to the corps of engineers approximately one year later. Between the summer of 1940 and the end of the war, the two branches carried out a construction program costing nearly $11,000,000,000, with the engineers performing by far the greater portion of the work. All told, nearly 3,000 command installations, including training camps, airfields, hospitals, storage depots, ports of embarkation and harbour defenses, and 300 major munitions plants and other industrial facilities were built during the course of the war. Such projects as the Pentagon building in Washington, D.C., and those under the Manhattan engineer district for the development of the atomic bomb stand high on the list of the corps's achievements in World War II.

After World War II.—Contrary to the usual pattern, the military activities of the corps of engineers did not give way to a predominance of civil activities after World War II. The constant tension engendered by the cold war with the Soviet Union and the open conflict in Korea demanded strategically placed air bases all over the world. The engineers constructed or supervised the building of bases in such diverse locations as France, Turkey, Morocco, Iceland, Greenland, Okinawa and Korea. Among other important duties during the Korean War, engineer troops supported the advance of ground forces by repairing roads and railroads, erecting temporary and permanent bridges across swift rivers subject to rapid flooding and participating in amphibious landings. In this war the engineers were also called upon to engage in large-scale demolitions and employ other means to delay and frustrate the enemy's penetration.

Faced with the possibility of nuclear attack across the polar icecap by jet-propelled aircraft, the department of defense revised its antiaircraft defense system during the 1950s. The corps of engineers was called upon to construct Nike batteries to guard key industrial and population centres. A network of radar warning stations across Canada further extended this protection (*see* AIR DEFENSE; RADAR). The corps also designed and built the U.S. portion of the St. Lawrence seaway. Challenging new assignments undertaken by the corps of engineers in the early 1960s included construction of ballistic missile bases and facilities for the aeronautics and space program. (B. D. C.; J. A. RN.)

GREAT BRITAIN

The modern corps of royal engineers (customarily abbreviated R.E.) came into being in 1856 when the officers and men of the engineer arm of the British army were for the first time amalgamated into one corps. The history of military engineering in England, however, can be traced from the time of the Norman Conquest in 1066. The title of the senior engineer officer has changed from time to time, but the appointment has persisted in direct succession under the following names: 1066–1642, king's (or queen's) chief engineer; 1642–60, parliament's chief engineer; 1660–1707, chief engineer of England; 1707–1802, chief engineer of Great Britain; 1802–1904, inspector general of fortifications; 1904–43, director of fortifications and works; since 1943, engineer in chief.

The first recorded overseas operation in which English artillery and engineers were employed was the siege of Calais in 1347 when engineers, including miners from the forest of Dean and smiths from the City of London, formed part of the army of Edward III. In 1350 the headquarters of the engineers was established in the Tower of London, and the board of ordnance was set up there in 1414 to administer all matters connected with fortifications, artillery and stores. Throughout the wars of the 16th and 17th centuries, permanent engineer officers were often in command of ordnance trains. Each train usually consisted of cannon, pontoons, entrenching and artificers' tools, scaling ladders, mining timber, tentage, and stores and transport; the men were formed into gun detachments, companies of bridgemen, artificers, storemen and transport drivers. The Restoration of Charles II in 1660 marked the birth of the modern British standing army and in 1698 the first ordnance train to be maintained in peace was formed. This included ten engineer officers in addition to the permanent king's engineers. In 1716, by warrant of George I, the artillery was formally constituted as a separate establishment. At the same time a corps of engineers was formed consisting of officers only; the rank and file continued to be raised, as required, for each campaign. Civilian tradesmen and artisans were engaged for defense works in peacetime.

In 1772 a military company of artificers, which was almost immediately renamed the soldier artificer company, was raised at Gibraltar for the construction and maintenance of its defenses. The company was gradually increased in size and a second company was formed in 1786. Officers for these companies were provided from the corps of engineers. The title corps of engineers was changed to that of corps of royal engineers in 1787, and the corps was authorized to take precedence, together with the royal regiment of artillery, on the right of the line when parading with other corps of the army. Also in 1787 six companies of the corps of royal military artificers were formed; these units were distinct from the soldier artificer companies serving at Gibraltar. Administered by the board of ordnance and commanded by officers of the corps of royal engineers, they were raised for service at home and in garrisons overseas.

In 1795 the headquarters of the corps of royal engineers was moved from the Tower of London to Woolwich and the depot of the royal military artificers was also established there. In 1797 the two companies of soldier artificers at Gibraltar were finally incorporated into the corps of royal military artificers, renamed in 1813 the royal sappers and miners. The increased importance of field works and military engineering in all its forms during the Peninsular War gave rise to the formation at Chatham, Kent, of the royal engineers establishment in 1812 for the professional and technical training of officers and men. The headquarters of the corps was moved from Woolwich to Chatham and in 1869 was retitled the School of Military Engineering. The royal sappers and miners were renamed the royal engineers in 1856 and the rank of private was replaced officially by that of sapper. The board of ordnance was abolished and the amalgamated corps of royal engineers came directly under the command of the commander in chief.

In 1832 King William IV granted to the corps of royal engineers permission to wear the royal arms and the mottoes *Ubique* ("Everywhere") and *Quo fas et gloria ducunt* ("Where right and glory lead").

Experiments in diving and underwater demolitions were first started by the royal engineers in 1825, and in 1871 submarine mining companies were formed to lay and man controlled mine fields and operate torpedoes for the defense of harbours and naval bases at home and overseas. These activities were taken over by the Royal Navy in 1905. By this time, however, the royal engineers were developing searchlights to illuminate the approaches to defended harbours. With the development of military aircraft, engineer-manned antiaircraft searchlights were introduced and an elementary method for training the lights on targets by means of sound locators and visual spotting was used. By 1940 most searchlights, both coast defense and the antiaircraft, were operated by the royal artillery.

The first field telegraph line used in war was laid and operated by the royal engineers in the Crimea in 1856. During the years that followed many different types of signal units of the royal engineers were formed. By the end of World War I military signaling had developed enormously and a separate corps of signals was

formed in 1920. (*See* SIGNAL COMMUNICATION.)

The engineers began experiments with the use of man-carrying kites and captive balloons for military purposes in 1878. To develop the technique of military flying an experimental balloon school and factory was set up at Chatham, and in 1890 the first balloon section in the British army was formed at Aldershot, Hampshire. With the introduction of the airplane royal engineer officers became the first pilots and navigators of military aircraft. In 1911 the balloon school at Aldershot was reorganized as the air battalion, royal engineers, and trained officers and men in the handling of military aircraft until the royal flying corps was created in 1912.

During the South African War the royal engineers operated steam transport companies with traction engines pulling trailers. With the introduction of gasoline-powered vehicles, engineer officers did much to develop mechanical transport in the army before it was taken over by the army service corps.

Engineer officers have been most widely used on the construction, operation and management of railways, particularly in India and in Africa. In World Wars I and II many thousands of transportation engineers were employed on railway work and in the operation of ports and inland waterway traffic in all theatres of operations.

Military survey had its origin in the mapping of the Highlands of Scotland after the 1745 rebellion, and in 1791 the ordnance survey was set up whose initial task was the production of a one-inch-to-the-mile map of the country, primarily for military purposes. The staff of the ordnance survey was drawn almost entirely from the corps of royal engineers. The development of the British empire necessitated the accurate mapping of vast areas and in nearly all cases the preliminary work, and in many cases the whole work, was done by royal engineers. They were responsible for the production of military maps of all kinds and for their storage and distribution in bulk. The provision of maps for the air force and accurate mapping on the scale required for intercontinental missiles produced new and difficult problems for the military surveyor in the 20th century.

The royal engineers used to be responsible for all works services, including barrack construction, required by the army at home and abroad. In 1959 the responsibility for works services in peace passed to a civil works service organization under a civilian director-general of works at the war office, but the royal engineers retained the responsibility for works services in war.

The strength of the corps of royal engineers at the beginning of World War I in 1914 was about 26,000; by 1918 it was 340,000. These figures give an idea of the magnitude of the engineer work that had to be undertaken. After the first German gas attack a special brigade of engineers was raised to develop antigas measures and make preparations for the offensive use of poison gas. The invention of the tank was largely due to an engineer officer (Lieut. Col. E. D. Swinton). To meet the enormous requirement of timber for the trenches and for mining operations engineer forestry companies were raised. In addition, all the R.E. tasks already described above were carried out on the largest scale.

At the start of World War II the strength of the royal engineers was about 54,000. Its wartime strength soon reached a figure of 280,000, excluding Indian and other dominion engineers. After the beginning of the German air offensive against the British Isles the royal engineers took charge of the disposal of unexploded enemy bombs. The mechanisms of the German fuzes were unknown and many bombs had delayed-action fuzes or fuzes fitted with anti-handling devices, but methods of dealing with them were soon developed. During the winter of 1940–41 when the railway and dock systems in and around London were threatened with disruption by enemy air attacks, a special force of engineer construction and dock maintenance companies was sent to help the overwhelmed civilian organizations. In Africa and Italy the royal engineers developed mine laying and mine-field clearance, airfield construction and water and gasoline supply. The engineering problems connected with the Normandy landings and the subsequent breakthrough were enormous. A prefabricated harbour with pontoon landing piers was towed across the channel and placed in position by transportation engineers, and a fuel pipeline was laid on the bed of the channel from England to France. Fighter airstrips and roads were rapidly constructed in the beachhead area. Later the Seine and Rhine rivers were bridged and ports in northern France, Belgium and Holland were cleared and put into operation.

During the Korean War, when the 1st commonwealth division was formed, its engineer regiment consisted of British, Canadian and New Zealand engineers. A mainly Australian "Britcom" engineer regiment was also raised and employed on the provision of base installations in Japan. This was the first time that commonwealth engineers served together in integrated units.

During the nuclear tests carried out on Christmas and Malden islands in 1957 and 1958, the royal engineers were responsible for providing all living accommodations, water and power supply, roads, airfield and port facilities, besides all the technical structures required for the firing of the thermonuclear weapons and for assessing their effects.

COMMONWEALTH OF NATIONS

The following commonwealth corps are allied to the corps of royal engineers: the corps of royal Canadian engineers, the royal Australian engineers, the royal New Zealand engineers, the Indian engineers, the corps of Pakistan engineers, the Ceylon engineers, the federation engineers of the Federation of Malaya, the corps of Fiji engineers, the Rhodesia and Nyasaland corps of engineers, the Nigerian military engineers and the Ghana engineers. The Gurkha engineers are affiliated with the corps of royal engineers and, until the Union of South Africa left the commonwealth, the South African engineer corps was allied to the corps of royal engineers.

The Corps of Royal Canadian Engineers.—The first military engineers in Canada were drawn from the corps of royal engineers. The first two units of Canadian engineers were formed as volunteer companies in 1851 on a nonpermanent basis. In 1903 a permanent Canadian engineer corps was established in addition to these nonpermanent units, and in 1904 it was redesignated the corps of royal Canadian engineers. The strength of the corps was increased in 1906 when Canada assumed responsibility for its own defense. During World War I Canadian engineer units served with great distinction on the western front; their tunneling companies played a prominent part in the greatest mining operation in military history when 500 tons of explosive were fired under the German lines at Messines ridge in June 1917. In 1932 the nonpermanent units of the Canadian engineer corps were redesignated corps of Canadian engineers, and in 1936 they also were awarded the title "royal."

During World War II R.C.E. units took part in the Dieppe raid and in the invasion of Sicily and Italy. They were withdrawn from the Mediterranean to take part in the Normandy landings and served in northwestern Europe. Units of the corps of royal Canadian engineers were employed in Canada on the construction and maintenance of the northwestern highway system and in survey and other works in the northern regions of the country. The Royal Canadian School of Military Engineering is at Chilliwack, B.C.

The Corps of Royal Australian Engineers.—Prior to federation each state maintained its own army, and an engineer corps was first raised in Victoria in 1860. In 1902 all state engineers in Australia were amalgamated into the corps of Australian engineers. In 1907 King Edward VII conferred on the corps its title of the corps of royal Australian engineers. Australian engineers first saw overseas service in the Sudan in 1885. They also served in the South African War of 1899–1902. During World War I they served in Gallipoli in 1915, in Flanders in 1916 and in the Palestine campaign of 1916–18.

In World War II the corps of royal Australian engineers were employed in Greece and Crete in 1941, in the western desert in 1942 and in the southwest Pacific operation. The headquarters of the corps is located at the School of Military Engineering, Liverpool, New South Wales, with an affiliated Transportation Training centre at Chowder Bay, Sydney. The corps consists of regular army units and the engineers of the citizen military forces who provide the backing of specialist engineer units. Units of the royal

Australian engineers were engaged in the construction and maintenance of the Woomera rocket range during the 1950s.

Royal New Zealand Engineers.—The military forces in New Zealand were first constituted in 1909, consisting of a small permanent force and a territorial force, both containing engineer units. For the outstanding services of the engineers in World War I in the Australian and New Zealand army corps (A.N.Z.A.C.) they were awarded the title royal New Zealand engineers. In World War II a very considerable expansion of the corps took place and units of the royal New Zealand engineers distinguished themselves in Greece, Crete, the western desert, in Italy and in the southwest Pacific.

The Corps of Engineers in India and in Pakistan.—The military engineers of the Indian and Pakistan armies can trace their history back to the East India company. The company formed settlements in the three presidencies of Madras, Bengal and Bombay and raised independent military forces to defend them. In this way the Madras, Bengal and Bombay engineers were formed in 1759, 1761 and 1775 respectively. As was the custom at the time in the British army, the officers and the other ranks served in separate corps. The officers were British; and the other ranks, except for a few British noncommissioned officers (N.C.O.'s) were enlisted locally in the three presidencies. These engineer units were originally known as pioneers, but during the period 1819–31 each of the three presidency engineer units was redesignated sappers and miners.

After the Indian mutiny the officers of the East India company engineers and the British N.C.O.'s of the presidency sapper and miner corps were absorbed into the British corps of royal engineers. The Madras, Bengal and Bombay sapper and miner corps were, however, retained in the Indian army. These corps were, until 1947 when India and Pakistan received self-government, officered by the royal engineers, and royal engineer N.C.O.'s were also posted for tours of duty with Indian sapper and miner units. Between 1890 and 1907, five Indian states raised their own sapper and miner units which were officered by special service British officers.

In India itself the three corps of sappers and miners carried out many large civil and military engineering projects in peace, and they distinguished themselves in many campaigns on the Northwest Frontier. In World War I four companies of Indian sappers and miners were almost continuously in the front line in France during 1914 and 1915 where they suffered heavy casualties. The main sapper and miner effort took place, however, in Iraq, Egypt and Palestine.

The strength of the Indian engineers at the outbreak of World War II was about 13,000, and by the end of hostilities it had risen to almost 345,000. In 1941 the sapper and miner corps became groups of the corps of Indian engineers. Indian engineers served in campaigns in Asia, Africa and the Mediterranean. They also carried out large engineering projects in India itself, including the provision of six major military bases. Transportation and other units of the Indian engineers completed the work of rehabilitating the port of Bombay after the explosion of an ammunition ship there in April 1944. In 1944–45 shortage of coal threatened the maintenance of India's war effort and Indian engineer units were employed in coal mining.

In Feb. 1946 the Indian engineers were awarded the title "royal." When the partition of the subcontinent took place in 1947 the Muslims of the royal Indian engineer units were absorbed into the new corps of royal Pakistan engineers, whose headquarters was established at Sialkot. In Jan. 1950 India became a republic and the Indian corps of engineers discarded their royal title. A few years later the Pakistan engineers also dropped the word "royal" from their title.

No record of the military engineers in India would be complete without mention of the work of R.E. officers in the survey of India, including the great triangulation, of their outstanding work in the construction, maintenance and operation of the Indian state railways and in public works.

Burma Sappers and Miners.—In 1887, while Burma was still part of India within the British empire, a company of Burma sappers and miners was formed and affiliated with the Madras sappers and miners.

The corps was disbanded in 1929 but re-formed in 1937. After the Japanese invasion of Burma during World War II the corps practically ceased to exist.

Ceylon Engineers.—Ceylon was granted fully responsible status within the commonwealth in 1948 and raised its own armed forces. The Ceylon engineers were formed in 1952 with headquarters at Homagama.

Federation of Malay Engineers.—The first engineer field squadron was formed in Malaya in 1952, and two years later another field squadron and a field park squadron were formed. These three units were incorporated into a field engineer regiment in 1955 which disbanded in 1957, when the personnel of the regiment were absorbed into two squadrons of the newly formed federation army engineers.

Gurkha Engineers.—A Gurkha field squadron was raised in 1948 and a second field squadron was formed shortly afterward. After the Korean War a divisional engineer regiment was formed as part of the brigade of Gurkhas.

Nigerian and Ghana Engineers.—During World War II two west African divisions were raised which included locally recruited engineer units officered by royal engineer officers. After the war a Nigerian field squadron was formed which, when Nigeria became independent in 1960, formed part of the Nigerian military forces. Similarly a Gold Coast field squadron was raised after the war, and this unit, after Ghana became independent in 1960, formed part of the Ghana army.

Other Commonwealth Engineers.—The Rhodesia and Nyasaland corps of engineers consists of a regular cadre backed by territorials. A detachment of Fiji engineers was raised to serve with the royal engineers serving on Christmas Island.

OTHER COUNTRIES

France.—The engineers of the early French armies were usually civilian artisans engaged for specific operations and commanded by any officers who possessed an elementary knowledge of military engineering. In 1673, however, Sébastien Vauban was appointed *directeur général des fortifications du royaume* and it was largely due to this great designer of fortified places that in 1690 an officer corps of engineers was established. Sapper and miner companies were formed later, although these units were generally attached to the artillery. In 1715 the word "*génie*" was first used to denote the *corps des ingénieurs*, and in 1776 the corps was given the title "royal," to be discarded in 1789 with the French Revolution. In 1801 the officer corps of engineers was integrated with the sapper and miner units and the amalgamated corps served with great distinction throughout Napoleon's campaigns. In 1868 military telegraphists were added to the corps. The first engineer railway battalion was formed in 1876 and a battalion of aeronauts, raised in 1904, was the forerunner of the French air force.

During the expansion of the French north African empire during the early part of the 20th century the *génie* were employed in opening up communications and other developments in the colonies, and in 1911 a unit of Moroccan engineers was raised.

World War I saw a great expansion of the *génie*, who were employed on the western front in a vast number of projects. Between World Wars I and II French military engineer effort was directed mostly toward a study of defenses and the construction of the strongly fortified Maginot line.

After the German army had overrun France in 1940, Free French engineer units were formed in London and fought in the western desert and north Africa. They later took part in assault landings and operations in Europe, being strengthened by volunteers from the French colonies and afterward from within France itself.

A far east expeditionary corps served in Indochina until 1954. Difficult engineer tasks in that undeveloped, wild country engaged more than 7,000 men of the *génie*. French engineers were also actively engaged in the war in Algeria.

The *génie's* first engineering school was established at Mézières in 1748, and L. N. M. Carnot, a former graduate of Mézières,

moved the school to Metz in 1795, where it was renamed L'École Polytechnique. In 1802 the artillery school moved from Châlons to become integrated with L'École Polytechnique. In 1884 a military school to train officer cadets for the artillery and engineers was established at Versailles. L'École l'Application du Génie was established at Angers in 1945, and the following year L'École Supérieure Technique du Génie was opened for the higher education of engineer officers.

Germany.—Before World War I, and even throughout that war, the German corps of engineers and pioneers, unlike their counterparts in most other armies, were considered to be a purely technical corps, and they were denied a say in the planning of operations. After the war the engineer element was limited to the provision of close engineer support to the infantry. Even when compulsory military service was introduced by Hitler in 1935 the main strategic engineer tasks were allotted to a civilian organization headed by Fritz Todt. Before the outbreak of World War II this organization constructed the Siegfried line, defending the frontier with France, and a network of military roads. During the war it constructed a chain of submarine bases along the north coast of occupied France and an elaborate defensive system known as the Atlantic wall to counter any seaborne invasion. The Todt organization was also used to construct the concrete and steel defenses of the Gothic line established by the Germans in central Italy.

The uniformed engineers consisted of engineers and construction troops, the distinction between them lying in their employment rather than their organization. The engineers prided themselves on being fighting troops. Duties not involving combat were carried out by the construction troops, who were regarded as a separate arm similar to the Todt organization. German engineer troops developed the use of antitank mines into a fine art in Africa; throughout the Italian campaign and the fighting in northwestern Europe they developed great skill in demolitions and in the use of booby traps; and they were faced with enormous communications and bridging problems on the Russian front.

The organization of the army engineers in the German Federal Republic is similar to that of the German army in World War II.

U.S.S.R.—Up to and including World War II the engineers of the Soviet army were classified as sappers and engineers. The former carried out simple field works, the more technical tasks falling to the engineers. Their tools and equipment were frequently primitive, but they were usually effective. Many tasks performed by the engineers in western armies, such as road and rail construction, provision of accommodation, etc., were in the Soviet army performed by rear service troops assisted by local civilian labour. The engineer units themselves in the forward areas were often employed as a trained nucleus for the technical direction of engineer tasks carried out by infantry or other arms.

The supply difficulties encountered by the Russians during World War II forced a good deal of improvisation on the Russian sappers. In this respect they showed great ability and ingenuity and their endurance was extraordinary. Their outstanding performance was largely due to their individual skill with hand tools and their efficient employment of their traditional building material, wood. In crossing many formidable rivers they constructed wooden bridges of locally felled timber. On numerous occasions these bridges were built with the deck about a foot below water level to conceal them from enemy observation. As electrical mine detectors were scarce, detection was carried out by groping with the hands or by prodding with sticks, bayonets or pitchforks. Modern earth-moving equipment was unknown, and human labour took the place of the bulldozer. In a withdrawal the Russian engineers were thorough and ruthless in their destruction of everything that might be of value to the enemy.

The great advances in education and technology in Russia after World War II led to a vast improvement in the technical competence of their sappers and in their equipment.

See also references under "Engineers, Military" in the Index volume. (J. H. S. L.)

BIBLIOGRAPHY.—W. S. Holt, *The Office of the Chief of Engineers of the Army* (1923); S. Forman, *West Point: a History of the United States Military Academy* (1950); Office of the Chief Engineer, General Headquarters, Army Forces, Pacific, *Engineers of the Southwest Pacific*, vol. i–iv (1947–59), vol. vi–viii (1949–51); Van Rensselaer Sill, *American Miracle: the Story of War Construction Around the World* (1947); F. G. Hill, *Roads, Rails and Waterways: the Army Engineers and Early Transportation* (1957); B. D. Coll, J. E. Keith and H. H. Rosenthal, *The Corps of Engineers: Troops and Equipment* (1958); William H. Goetzmann, *Army Exploration in the American West, 1803–1863* (1959); Institution of Royal Engineers, Chatham, *History of the Corps of Royal Engineers* (Norman Times to 1948), 9 vol. (1952 *et seq.*) E. W. C. Sandes, *The Military Engineer in India*, 2 vol. (1935), *Royal Engineers in Egypt and the Sudan* (1937), *The Indian Sappers and Miners* (1948), *The Indian Engineers, 1939–47* (1956).

ENGINES OF WAR, a generic term applied to the heavy weapons that were the artillery pieces of the ancient and medieval worlds as distinguished from the lighter arms of the individual soldier. As early as the reign of Uzziah the Old Testament refers to "engines, invented by cunning men, to be on the towers and upon the bulwarks, to shoot arrows and great stones withal." These weapons remained in use for many centuries until superseded by weapons employing gunpowder at the end of the middle ages. (*See* ARTILLERY.)

Although engines of war are most often associated with siegecraft (*see* FORTIFICATION) both the Greeks and Romans used them effectively in the field. The engineers of Alexander the Great, according to ancient chroniclers, carried only the cords and other essential gear; they cut the heavy timbers of the framework from trees near the scene of operations. Alexander's crossing of the river Jaxartes (now Syr-dar'ya) was made possible by a barrage of stones and spears from his war engines that kept the Scythians at a respectful distance on the opposite bank.

Catapult and Ballista.—These were two of the most important weapons of the ancient and medieval worlds, but were known in different lands and ages by a wide variety of names such as onager, scorpion, torment, petrary, mangonel and sling. Variations in translation probably account for the confusion in terminology, for the same functions are ascribed in reliable sources to both the catapult and ballista. One authority differentiates between the two only with respect to the missiles used as ammunition; another states that catapults were merely small ballistae; others make a distinction according to the period of history when they were used, or their principle of propulsion. Sir Charles Oman, the famous military historian, asserted in his *History of the Art of War in the Middle Ages* (1898) that the ballista was the parent of the crossbow, but it would appear that the Roman scorpion was more nearly the crossbow's parent.

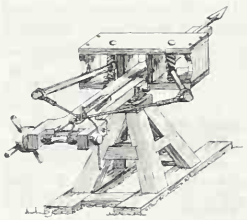


FIG. 1.—BALLISTA OR MANGONEL USED IN ANCIENT WARFARE

Although there is no clear-cut distinction between the various weapons, it is known that two separate principles of propulsion—torsion and tension—were used in the war engines of the ancients. A third principle, counterpoise, came into use during the middle ages. Torsion (the twisting of heavy cords or fibres) generated the force for hurling spears, arrows or stones. Tension (the bending of a bow or long plank) was the propelling force in such weapons as the springal, discussed below. The medieval trebuchet, also discussed below, employed the principle of counterpoise.

In one common form of catapult, retractable arms were entwined in cords secured to a framework of strong timbers. The cords were twisted by winches and the arms were drawn back under heavy pressure. When spears or arrows were used for ammunition, a groove comparable to the bore of a gun channeled the flight with considerable accuracy.

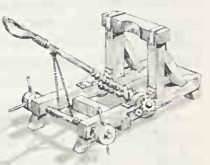


FIG. 2.—COMMON FORM OF THE CATAPULT USED IN ANCIENT WARFARE

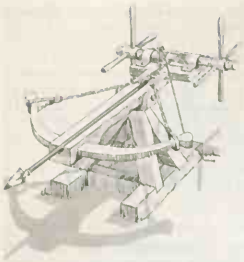


FIG. 3.—SCORPION OR BALLISTA USED BY ROMAN ARMIES

The cords supplying the motive force of both the catapult and ballista were made of rope, horsehair or sinews of animals, but the secret of preserving their elasticity has never been recaptured by modern experimenters. In the 20th century the term catapult gained wide acceptance to denote devices for propelling aircraft from the decks of ships. The word has also survived as a verb denoting propulsion of the slingshot type.

The ballista appeared in a variety of forms, most of them not readily distinguishable from catapult types, though ballistae were generally larger than catapults. Tacitus refers to a huge ballista in a siege of A.D. 69 that wrought terrible execution. Modern experiments have established the possibility of ranges up to 400 yd. with stones weighing 50 lb. Some of the ballistae of the Romans were too large to be moved and could be used only from fixed positions during a siege. On the other hand, the Romans developed in the scorpion a weapon small enough to be operated on the battlefield by two soldiers. In effect, the scorpion was merely a huge bow, the wooden arms of which were retracted by a winch for the motive power. Steel replaced wood in later types.

By the 4th century A.D., according to Vegetius, as many as 10 catapults and 60 ballistae were assigned to each Roman legion. This is a ratio of 1 artillery piece to 100 infantrymen as compared to the 3 pieces of artillery to 1,000 men in the armies of Napoleon.

Springal or Falarica.—This type of war engine was used chiefly in naval engagements of both the ancient and medieval ages. It derived its force from tension rather than from torsion. It consisted simply of a plank bent back by winch-tightened cords until it generated the force, when suddenly released, to strike the missile a sharp blow and propel it in a straight or curving flight, depending on elevation. Not a powerful weapon, as compared to the catapult or ballista, the springal was most effective when employed for hurling firebrands onto the deck of an enemy vessel or over a city wall. Stones were also used for missiles.

FIG. 4.—SPRINGAL OR FALARICA USED IN ANCIENT AND MEDIEVAL WARFARE

Battering Ram.—Though not a weapon of propulsion, the battering ram is grouped with engines of war because it assumed the function in siegecraft later taken over by the artillery of gunpowder—the opening of breaches in the walls of a besieged city. A heavy timber, usually the trunk of a large tree, the ram had at its front end a metal knob or point, often in the shape of a ram's head. Wielded by as many as several hundred men grasping thongs, it dealt blows which only the stoutest walls could withstand. Small battering rams were held by hand between two lines of soldiers; larger varieties were suspended by ropes or chains from overhead beams. The troops who operated these weapons were usually protected by a testudo (*see* below) or other form of shelter because the defenders on the walls above would endeavour to drop heavy stones on them, shoot flaming arrows at them or pour boiling oil on their heads.

Bore.—The bore was a tool, somewhat like a battering ram, used to drill holes through joints in the masonry until large stones could be extracted. Sometimes called the "cat," because it clawed, or the "mouse," because it gnawed, this device often succeeded where the much larger battering ram failed. In effect, it was a drill, one end revolving in the socket of a wooden upright and the other end terminating in a metal point for penetrating the mortar. The bore was turned by soldiers grasping handles, and the size of the crew depended on such factors as the length and weight of the bore.

FIG. 5.—THE BORE

FIG. 6.—BATTERING RAM SLUNG FROM A TESTUDO COVERED WITH GREEN HIDES AS PROTECTION AGAINST FIREBRANDS

Testudo.—The Romans originally applied this word, meaning tortoise, to the tactic of troops locking shields overhead while advancing against the walls of a besieged city. But testudo came in time to mean a structure of heavy timbers built for the protection of working parties exposed to arrows, spears and firebrands hurled down upon them. Covered with damp clay or green hides to resist fire, it served not only to protect the besiegers but also to provide a beam from which the ram was suspended. Mounted on wheels or rollers, some of these mobile fortresses were large enough to shelter several hundred soldiers; additional stories could be superimposed to form a tower overlooking the walls.

Trebuchet.—The only engine of war developed during the middle ages was the trebuchet, sometimes called perrier or petrary. In building trebuchets the artificers of medieval Europe employed the principle of counterpoise. Heavy weights were attached to the short end of a spar, the long end being pulled back by winch-drawn cords until parallel to the earth. The usual missile was a stone held in a sling or a spoon-shaped receptacle at the end of the long arm. When the cords holding the long end were suddenly released, the descent of the weights at the other end supplied the force to hurl the missile in a curving flight over the wall of a besieged city. The distance could be controlled to some extent by varying the weights, but accuracy was out of the question. Counterweights of several tons were sometimes used, so that the trebuchet gradually took the place of the ballista. Even after the introduction of the first weapons of gunpowder, bombards mounted on sledges and firing stone balls, engines of war were still formidable enough in comparison to survive for several decades.

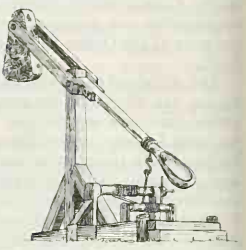


FIG. 7.—TREBUCHET, THE ONLY ENGINE OF WAR TO ORIGINATE IN THE MIDDLE AGES

If legend may be credited, Archimedes invented a war engine at the famous siege of Syracuse in the form of a crane powerful enough to grapple the bows of enemy galleys, hoist them high in air and plunge them to their destruction beneath the sea. Whatever the foundation in sober fact may have been, nothing resembling this fabulous weapon was inherited by later ages.

BIBLIOGRAPHY.—H. Wilkinson, *Engines of War* (1841); Charles Oman, *A History of the Art of War in the Middle Ages* (1898; rev. ed. 1924); J. R. Newman, *The Tools of War* (1942); J. E. Hicks, *What the Citizen Should Know About Our Arms and Weapons* (1941); Lynn Montross, *War Through the Ages,* rev. ed. (1960); Edwin Tunis, *Weapons* (1954). (LN. MS.)

ENGLAND, JOHN (1786–1842), U.S. Roman Catholic prelate, first bishop of the diocese of Charleston, was born at Cork, Ire., on Sept. 23, 1786. He was ordained Oct. 11, 1808, and became an instructor at St. Patrick's seminary, Cork. In 1812 he was made president of the seminary. While parish priest at Bandon, County Cork, he was named bishop of the new diocese of Charleston, which, erected in June 1820, consisted of the states of North Carolina, South Carolina and Georgia. John England was consecrated in Ireland on Sept. 21, 1820. The bishop saw that the first need of his diocese was education. He prepared and printed a catechism and a missal for Americans. He founded the *United States Catholic Miscellany,* the first Catholic newspaper in the country, which continued publication until 1861. He began two schools, the Philosophical and Classical seminary for boys and an academy, conducted by the Ursulines, for girls. For the care of the sick and orphans he founded a religious community, the Sisters of Our Lady of Mercy. To assist immigrants and workingmen he organized the

Brotherhood of San Marino. In 1833 Bishop England was appointed apostolic delegate to Haiti, the first important diplomatic mission given to an American prelate. His efforts to secure a concordat were not successful.

John England was recognized as an eloquent orator, and he was the first Roman Catholic clergyman to be invited to speak before the congress of the United States (1826). He became a United States citizen in the same year. He died at Charleston, S.C., on April 11, 1842.

Bishop England's works were collected and arranged under the advice and direction of his immediate successor, Ignatius Aloysius Reynolds, and published in five volumes at Baltimore in 1849.

See Peter Guilday, *The Life and Times of John England*, 2 vol. (1927). (R. C. Mn.)

ENGLAND, the largest, most industrial and populous unit of the United Kingdom of Great Britain and Northern Ireland, is roughly triangular with sides between 300 and 360 mi. in length, measured in straight lines. The distances along the eastern and southern seaboards are less than twice their straight lengths, for in general these coasts follow sweeping curves rather than sharp indentations. On the west coast the Bristol channel penetrates deeply into the mainland, so that the curved length (including Wales) is approximately 1,200 mi. The area of England is 50,871 sq.mi. Continental Europe is 21 mi. from the south coast at the Straits of Dover, and Cornwall is the only part in direct contact with the Atlantic ocean. The remaining coasts are washed by the partially enclosed North sea, Irish sea and St. George's channel. The land is territorially divided into 40 counties.

This article contains the following sections and subsections:

I. Physical Geography
  1. Geology and Structure
  2. Physiography
  3. Climate
  4. Soils and Vegetation
  5. Animal Life
II. Geographical Regions
  1. Highland England
  2. Lowland England
III. The People
  1. Early Ethnic and Social History
  2. Social Development: Middle Ages to Modern Times
  3. Religion
IV. Archaeology
V. Population
  1. Early Population Patterns
  2. 19th-Century Growth
  3. 20th Century
VI. Administration and Social Conditions
  1. Local Government
  2. Housing and Planning
  3. Welfare Services
  4. Police and Fire Protection
  5. The Judiciary
  6. Education
VII. The Economy
  1. Agriculture
  2. Forestry
  3. Fishing
  4. Mining
  5. Power Resources and Utilization
  6. Iron and Steel Industries
  7. Engineering Industries
  8. Textile Industries
  9. Textile-Using Industries
  10. Chemical Industries
  11. Glass and Pottery Manufacture
  12. Rubber Manufacture
  13. Paper Manufacture
  14. Transport and Communications

For history *see* BRITAIN and ENGLISH HISTORY. For subjects concerned with the United Kingdom as a whole (*e.g.*, constitution, legislature, defense, finance, trade) *see* GREAT BRITAIN AND NORTHERN IRELAND, UNITED KINGDOM OF.

I. PHYSICAL GEOGRAPHY

1. Geology and Structure.—In a general account of the geological history of England, it is convenient to include Wales, since the two are virtually one unit from this point of view. The positions of the outcrops of the various rocks and the physiography

PHYSICAL FEATURES OF ENGLAND AND WALES

of the area are the combined result of the geographies of past times, the formation of the sedimentary and other rocks, the changes made to them by earth-movements, and the varying amounts of erosion which they have experienced. In what follows, a chronological sequence will be adopted so far as is possible. Perhaps no other area of comparable size possesses such a varied sequence of rocks, for in England and Wales there are representatives of every geological age excepting the Miocene (when the whole area was land), and if a traverse is made from Anglesey to London most of the succession is crossed in order from Pre-Cambrian to Recent.

Pre-Cambrian.—This rock formation occurs on the surface in parts of Anglesey, the Lleyn peninsula between Nevin and Bardsey island, the Longmynd, the Wrekin ridge, the Great Malverns, Charnwood forest and in other areas. In Anglesey and the Malverns, these rocks include gneiss as the basal rocks. It would appear probable that considerable earth-movements and subsequent denudation occurred within the Pre-Cambrian, for some of the volcanic and pyroclastic rocks of Charnwood and the Malverns, although still belonging to the same period, rest on an irregular and weathered surface of the underlying gneiss. Later, the grits and shales of the Longmynd and other places are often indicative of sedimentary deposition, and may well be of the same age as the Torridonian of Scotland.

No doubt the Pre-Cambrian underlie many parts of England and Wales now covered with later rocks, and they form the foundations upon which the rest are deposited. They were probably the south-eastern limit of the shield known as North Atlantis (the Old Red continent, or North Atlantic continent; *see* DEVONIAN SYSTEM), which survived until quite late geological times and exerted a dominating control over the succeeding events. Britain owes its complex history and the variety of its present scenery to the fact that it lay on the periphery of this great, stable continent.

Cambrian.—Long-continued subaerial erosion reduced the Pre-Cambrian land to relatively low relief, and negative crustal movements converted it into a shallow sea. The usual succession of sedimentary deposition followed. Thick layers of coarse sand

were spread over considerable areas, in some parts pebbly beaches led to the formation of conglomerates, in favourable areas of clear water limestones were formed, and, as the whole area sank still farther, with the consequent removal of coastlines to a greater distance, the general deposit became silt or mud, which ultimately changed into a shale. This sequence is repeated many times in the evolution of England and Wales, and in the Cambrian the area of deposition appears to have covered all Wales and most parts of England. The maximum total, original thickness of the sediments was about 16,000 ft., and this occurred in Merionethshire and Denbighshire. The largest area where rocks of this age now outcrop on the surface is in the Harlech dome in Merionethshire, where their lower layers comprise coarse beach and offshore deposits, while occasional outcrops farther south and east are of rather finer grain-size, as would be expected with a land-mass to the northwest. In the Nuneaton area, the Cambrian is represented by about 1,000 ft. of quartzites, followed by 3,000 ft. of varied shales (Middle and Upper Cambrian), with occasional calcareous conglomerates and flags.

Ordovician and Silurian.—The broad geosynclinal sea still persisted with the North Atlantis continent to the northwest and Fennoscandia, probably with a southwesterly extension, forming the southeast flank. The depth of water and the proximity or remoteness of the land whose rivers brought sediment, varied from time to time, so that the texture of the resulting rocks changed accordingly. The Arenig series, the lowest part of the Ordovician, is first a coarse grit and then becomes in turn a sandstone, a sandy-mudstone and a dark shale—all indicative of a nearby coast, which gradually retreated farther and farther away as the geosyncline deepened. During Middle Ordovician times there was a sudden uplift and coarse grits were deposited, to be followed in Bala times by shales or fine sands once more.

It is possible that parts of midland England were then land, and when the Silurian succeeded the Ordovician, the Caledonian orogeny, or mountain-building period, had already begun. New mountain ranges lay along the old North Atlantis coast. Their remnants now form the Highlands of Scotland. A triangular block occupied what is now the Midlands of England. The geosyncline still persisted between these two land areas, although it had probably divided in two, with a land ridge in a line from Berwick through Solway to Wexford. The deposits were often fine-grained, especially during the first part of the period. Later, when the Midland block became submerged and the northern half of the syncline became land, the deposits became sandy. In the Cardiganshire area, the Llandovery series often consisted of alternating beds of gray sandstone (graywacke) and fine mudstone, and it is suggested that these may be the result of slips down the slope of the sea-bed from the Pembrokeshire area.

The rocks which form the core of the Lake district, the Borrowdale Volcanic series, were formed as lavas, tuffs and agglomerates probably during the later part of the Ordovician period, and these had followed mudstones and grits, which later became metamorphosed into the Skiddaw and other slates. It is not definitely known whether these are of Upper Cambrian or Lower Ordovician age.

There was some slight folding and erosion before a negative movement brought the area below sea level again at the end of the Ordovician, with the result that the Ashgillian and Silurian rocks of northwestern England were deposited.

The Caledonian orogeny, already operative in the north, spread into England and Wales at the end of the Silurian. The Lake district dome came into existence for the first time, and a series of northeast to southwest trending synclines and anticlines were produced in northern and western Wales.

Devonian.—The Lake district and most of north Wales remained land during the Devonian period which followed, and the whole geography of Britain had been changed. North Atlantis still remained, but with great new mountain chains, and Fennoscandia had increased its area to cover the southern part of the North sea. From it there was now a much greater peninsula than before across northern England and Wales and into central Ireland. During Devonian times, that peninsula changed its area and plan considerably, influenced by the dying movements of the Caledonian orogeny. There were thus very varying embayments and straits between that peninsula and North Atlantis, and these influenced the deposits.

In south Wales the Lower Devonian sandstones were formed as a delta on the south side of that peninsula. As time went on into the middle and upper sections of the Devonian, the land area of northern and central Wales appears to have extended northward to join with North Atlantis and, for a period, southward to cover south Wales. It was during this phase that most of the Devonian rocks of Devon itself were deposited, and there is still controversy as to whether these and other rocks of the same age are marine, lacustrine or even of wind-blown desert origin. It is, however, almost certain that some of the rocks of Upper Devonian age include material obtained by erosion of the Lower Devonian rocks.

Carboniferous.—This has been studied in more detail than any other British group. North Atlantis had been extensively eroded during Devonian times, but its remnants still lay between central Scotland and Iceland. Probably most of the North sea and certainly Scandinavia were a land area, but much of continental Europe was below water, although there were a few scattered islands. Two of these were St. George's land and East Anglia land. The former included the Wicklow mountains, St. George's channel and most of Wales, except Glamorganshire. The name of the latter adequately indicates its locality and it may have extended across the southern North sea to join the Brabant island in Belgium.

In and around these a great variety of sediments were deposited. The sea was quite shallow, although subsidence continued at a rate which in general was only a little less than the rate of deposition. In England, the Lower Carboniferous, or Avonian, contains the main deposit of limestone. It has been subdivided into several zones by using distinctive fossils, but it suffices to refer to the Lower and Upper Avonian, which generally are known as the Tournaisian and Viséan respectively. In both these limestone was extensively deposited in northern England, in the Lake district area, in northeastern and southern Wales, and in Kent. Coral reefs occurred in some localities, while by contrast during the Viséan, an area in northeastern England, because of intermittent subsidence, became alternately a delta-swamp and a clear sea. This gave rise to sandstones, with vegetative remains making thin coal seams, alternating with limestones and shales. The resulting rocks are known as the Yoredale series.

The Upper Carboniferous, or Namurian, was separated from the Lower, or Avonian, by relatively slight earth-movements, which, however, greatly altered the conditions of deposition in those shallow seas. St. George's land and East Anglia land became united, then certainly joined the Brabant island, and became larger, so that they covered more of southern England. There was a gulf where the Bristol channel now lies, but of considerably greater width, and in this area, especially in Devonshire, the Culm measures, mostly shales and sandstones, were laid down as basin deposits. In the Pennine area, a great delta grew from the coast of North Atlantis and gradually spread until it covered the whole of northern England and northeast Wales. This must have been of much the same size as is the present-day Ganges-Brahmaputra delta. Its composition naturally varied greatly from point to point and from time to time. The coarse sand deposits are generally known as the Millstone Grit, and the intricate sequence of grits, that followed one another as the delta evolved, have been worked out.

Parts of these deltas frequently became covered with a luxuriant tropical vegetation of forest and peats. Small downward crustal movements often drowned these swamps, overwhelmed them with sand and so gave rise to the occurrence of seams of coal. Frequent repetition of these events produced the succession of seams one above the other, separated by other rocks, generally sandstones. This necessitates mining rather than quarrying, excepting where particularly rich, shallow deposits make opencast working economic.

Toward the end of Carboniferous times vigorous earth-movements began once more, and the orogeny variously known as Armorican, Hercynian, Altaid and Variscan changed the geography.

The alterations were greatest in continental Europe, but from that time until it finally disappeared, probably before the Pliocene, North Atlantis was to the west of Britain, rather than to the northwest, and occasionally it became linked with land over France and Spain.

Permian and Triassic.—After the Carboniferous and during the orogeny, Britain experienced hot desert conditions, and the sand rocks of the Permian and Triassic periods were formed, sometimes as wind-blown sediments, sometimes as sand-deltas in lakes or in seas. Often the lakes were centres of inland drainage and so salt accumulated, as in Cheshire and other places. The seas were linked with the Tethys sea, which covered much of central Europe and was the forerunner of the Mediterranean. Under such conditions, small changes of level made substantial geographical and depositional alterations. The Triassic rocks belong to the beginning of the Mesozoic era, and they now lie on the surface, although often covered with the relatively thin glacial deposits of the Pleistocene Ice Ages, on either side of the Pennines, in the Midland plain, in the vales of Eden and Clwyd, and in a complex pattern from the mouth of the Severn to Sidmouth.

Jurassic.—These deposits of sand, limestone or clays which followed, stretch as a long belt from the Dorset coast to form the Cotswolds and pass through Northamptonshire to the Humber, northward of which they occupy only a narrow belt until they swing round the Vale of Pickering as the Hambleton fells, Cleveland hills and North Yorkshire moors.

The synclinal hollows, which had formed in Cheshire and in mid-southern England, became infilled, and more and more of England was dry land—an area of erosion rather than of deposition. Peneplanation continued until the beginning of the Cretaceous period.

Cretaceous.—The general level had become so low that the reverse process set in. Tethys gradually enlarged itself, due to slight crustal down-warping, and the deposits of the Lower Cretaceous, or Wealden, spread over southeastern England as the coastline, lying roughly northeast to southwest, slowly retreated toward the Welsh borderlands. Among the silts and sands of this time are the Weald clay and the Lower Greensand. By the time of the Upper Cretaceous, most of Europe, including practically the whole of Britain, was submerged. This is known first as the Albian, and then as the Cenomanian, transgression, and in this shallow sea great layers of Gault clay and then Chalk were deposited.

Gault now forms the water-tight floor on which the Chalk lies beneath the lower Thames basin, and it extends a little beyond the Chalk escarpments to outcrop as a clay vale at their foot. The Chalk has become the North and South downs, Salisbury plain, much of Dorset, and the Chiltern hills and their northward extensions as far as the Yorkshire wolds. It is believed that it originally covered almost the whole of Britain, thinning considerably toward the west, but it has been wholly removed from most places, although a little Chalk may still be found in Northern Ireland.

At the end of the Cretaceous the surface of England was thus a fairly level sheet of chalk in a shallow sea. Beneath it lay the remnants of the past. The Alpine orogeny began at this time and continued through much of the Tertiary period. Only the outermost ripples of this great series of earth-movements, which built up most of the mountain ranges of modern continental Europe, reached England. There was the gentle warping of southeastern England into the elongated dome of the Weald, and of the lower Thames basin into the syncline between the North downs and the Chiltern hills.

Eocene and Oligocene.—This syncline became infilled with the various sands and clays that constitute the Eocene, and similar conditions prevailed in the Hampshire basin. There, however, the sometimes marine, and sometimes lacustrine, conditions persisted through Oligocene times.

Miocene.—The land area gradually extended, either due to crustal uplift or to a fall in sea level, so that during the Miocene the whole of England was above water, and there were no deposits. Subaerial erosion was the dominating feature. In the now higher west the whole of the Chalk cover was removed by denudation, and the Welsh, Pennine, Lake district and other mountains were exhumed. They had been in existence since early Triassic times, but were reduced almost to a peneplain by the mid-Cretaceous. Further differential erosion then removed the upper surfaces of the Triassic and other rocks, and so restored the uplands to the rank of mountains and led to the present-day relief.

Pliocene.—Small areas of England, especially in East Anglia and the lower Thames basin, became submerged during Pliocene times. There appears to have been very little tilt or warping connected with these submergences, which at their maximum brought the sea up to what now has an altitude of between 600 and 700 ft. This therefore may well be a change of sea, rather than of land, level. During this episode the various shelly, sandy and silty deposits usually known as "crags" were laid down as shore or estuarine material in East Anglia. They may well be correlated with extensive level areas at about the same height, which are found in many parts of England and Wales.

Pleistocene.—At the end of the Pliocene there was marked cooling of the climate, which led to the glaciation of most parts of England. This period, known as the Pleistocene, may well be considered to extend to the present-day, although the word Recent is sometimes used with reference to the latest deposits. Since the events and results of the Pleistocene glaciation are so intimately connected with the relief of England as it now exists, they will be considered in the following section.

2. Physiography.—The present landforms of England are the direct result of the geological history, which has just been recounted. Differential uplift has kept some of the older rocks at relatively high elevations, and differential erosion has exposed some, while leaving others still covered with later deposits.

There are three upland regions: Cumbria or the Lake district, the Pennines or central uplands with their flanking coalfields, and the southwest peninsula. The Lake district began with Caledonian folding, there was additional folding during the Armorican orogeny, and the local doming occurred during the Alpine period of movement. The Pennines and the southwest peninsula (together with eastern Wales, which lies between and links them) were uplifted in Armorican times and owe their origin to these earth-movements. Geologically, therefore, the uplands came into existence at the end of Paleozoic times, although it should be observed that, in the southwest, parts have since been so denuded that the present surface of some of the highest areas takes the form of intrusive masses of igneous rock, mostly granite, which forced their way among the uplifted sedimentary rocks soon after they had been folded. Now only the cores of these old mountains remain.

Thus to the northwest of two lines drawn roughly from Torquay through Cardiff, Chepstow and Great Malvern to Newport (Cheshire), and from Stoke-on-Trent through Nottingham, Doncaster and Tadcaster to Darlington, the rocks are of Carboniferous or earlier age and were folded into their present forms by either the Armorican or Caledonian earth-movements or by both. In later times they experienced substantial erosion and were submerged at least once, but any movement they have experienced has been purely vertical and has produced in them little or no further warping or tilt.

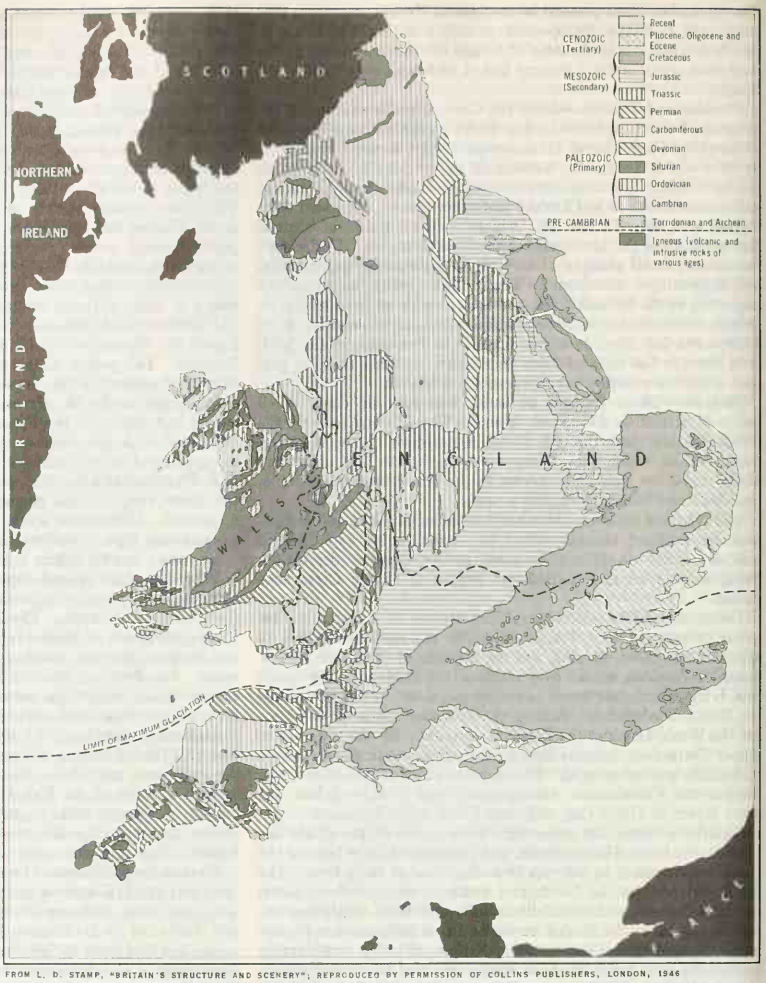
There is a gap between the Welsh borderland and the Pennine "oldlands" occupied by the Triassic sandstones and marls of the Lancashire, Cheshire and Shropshire plains and by the Midlands. These relatively nonresistant rocks were mostly deposited whilst the Armorican uplifting was proceeding, and similar rocks occupy most of the Vale of Eden and the Carlisle plain in the northwest.

The remainder of England, the east, southeast and south to Torquay, is formed of rocks from Jurassic to Pliocene age, and could well be designated the Mesozoic scarplands. There is the Oölitic escarpment (Jurassic), which closely follows the boundary with the older formations. For part of its length it forms the Cotswold hills, famed for their farmlands and scenery. Farther north, it becomes the Lincolnshire heights. The same rocks form the Cleveland hills and the northern half of the Yorkshire moors, but in this area the Corallian group of the Upper Jurassic is the resistant rock and the chief scarp former.

Parallel to the Cotswolds, and 10–15 mi. to the southeast, is the flat-floored Vale of Oxford clay (Jurassic), and along its east-

ern edge, beyond the Kimeridge and Gault clays, the Chalk escarpment of the Marlborough and Berkshire downs and the Chiltern hills rises steeply from about 300 to 800 ft. The dip-slope leads across the Vale of St. Albans southeastward to the lower Thames basin, and south of the river the land rises as the dip-slope of the North downs. While the lower Thames (or London) basin is a synclinal hollow and thus owes its origin to tectonic geology, the Weald (*q.v.*), the area between the North and South downs, is an erosive feature. Structurally it is an elongated dome or pericline (the eastern end is in France across the Strait of Dover). The arch of Chalk, which once stretched over it, has been eroded away, as also have several of the underlying rocks. A traverse of the Weald from north to south begins with a sharp step down the North downs Chalk escarpment, often from 600 to 300 ft., a short passage across the narrow, low and generally wet Gault clay, a gentle ascent up the Lower Greensand to nearly 1,000 at Leith hill, and a drop to 250 ft. on to the sticky Weald clay. This forms a fairly level surface along which almost all the exceptionally straight railway from Redhill to Ashford (London-Folkestone route) is built. The Hastings beds, about 20 mi. wide, occupy the middle area of the Weald as its core. In the centre these rocks have been completely removed, and for a narrow strip the Purbeck beds of the Upper Jurassic penetrate the otherwise entirely Cretaceous area.

The remainder of the traverse crosses the same features in the reverse order, and generally with rather narrower outcrops, until the south coast with its Chalk cliffs is reached.



FROM L. D. STAMP, "BRITAIN'S STRUCTURE AND SCENERY"; REPRODUCED BY PERMISSION OF COLLINS PUBLISHERS, LONDON, 1946

GEOLOGIC FEATURES OF ENGLAND AND WALES

Another important low-lying area is the Hampshire basin. This is an elongated, inverted dome, with much of its southern limb removed by present-day marine action. The only parts of this southern section which still remain are the Isle of Purbeck (not an island) and the Isle of Wight.

The Pleistocene was initiated on these landforms. It took the form of a series of glaciations with warm interglacial periods between. The present may be such an interglacial, for it is not known whether or not there are further glaciations to come. Geologically, the Pleistocene lasted a very short time, a little more than 500,000 years, and as such it scarcely merits the distinction of a special name of "period" rank. Its proximity to the present, the influence that its effects have on the human use of England, and its unusual nature, all serve to give it special importance.

There was no sudden climatic deterioration at the onset of the Ice Ages and the growth of the glaciers and ice sheets required several millenniums for their full development. Nevertheless, at its maximum, most of the mountains and virtually all the valleys and plains of England were occupied by glaciers or were covered by ice sheets as far south as a line roughly from Gloucester, through Stratford-upon-Avon and Luton to southern Essex.

After at least one interglacial period, which lasted for several thousand years, the ice spread again. This "Newer" or Main Irish sea glaciation extended from the Irish sea across the Cheshire-Shropshire plain probably to Wolverhampton, and on the eastern side of the Pennines as far south as Worksop. Later glaciations were smaller in area and in England scarcely extended beyond the Lake district and the Welsh borderlands.

As the ice melted away it deposited the morainic material that it carried under, within, or on it as extensive sheets of boulders and clay, much of which had been brought from distant rock outcrops. This blanket varied from zero to 200–300 ft. in thickness, but probably the average value over much of lowland England was 10–30 ft. At the same time, the meltwater from the ice, by its fluvial action, was able to redistribute and re-sort a considerable quantity of this boulder clay, so that stratified patches of sand and small pebbles are commonly found in all localities. These Pleistocene deposits are of the utmost importance to the economic life of England, for, north of the limiting line described, they

constitute almost the whole of the ground on which the farmers grow their crops, and thus they affect the drainage, texture and nutriment of the soils.

South of the ice limit, conditions were periglacial during the Ice Ages. Snow lay for several months during the winters and the ground was frozen to a considerable thickness. In the summers the snow melted for a few months and surface thaw, over a subsoil that still remained frozen, encouraged such processes as solifluction and mass wastage. It has produced head and coomb deposits, again with a great influence on modern agriculture.

Changes in sea level have resulted in the partial drowning of old river valleys, such as the Bristol channel, the Thames estuary, the Humber and, on a smaller scale, along the Essex, Cornish and Devonshire coasts. The Solent, which separates the Isle of Wight from the mainland, is a drowned valley which previously extended westward to beyond Poole harbour. The southern side of the middle portion of this valley has been completely eroded away by marine action, so that Bournemouth is exposed to the open sea.

In an opposite sense there are numerous examples of raised beaches, such as those around Morecambe bay, which are indicative of sea levels in post-glacial times that were slightly higher than now.

Drainage.—The drainage system of the country has existed whenever the area has been land, and with such a complex history the evolution of the rivers has been equally varied. Parts of the present rivers are inheritances from the past, while other parts only existed after the end of the glacial period. Wherever rivers flow from the mountains on to a moraine-covered plain, the routes by which they cross the latter may bear no relationship to the preglacial courses.

The longest river wholly in England is the Thames (*q.v.*), 210 mi. It rises in the Cotswolds and flows eastward along the wide vale of Oxford clay to Oxford. There it abandons the vale, ceases to be a subsequent stream and turns south to plunge through the Chiltern hills by the Goring gap. This brings it into the infilled syncline containing London, through which it flows to the North sea, which it enters by a partially drowned estuary, choked with sand and mud banks.

The Severn (*q.v.*) is said to be about ten miles longer, but it is difficult to decide where its mouth lies and the river becomes estuary, and in any case its upper course is in Wales. In the Pliocene it flowed northward from near Shrewsbury to accept the Dee as a tributary and enter the Irish sea by the present Dee estuary. At the approach of one of the interglacial periods there was a stage when the ice had melted in central Shropshire but not in Cheshire. Thus the meltwaters and waters which may have come down by the upper Severn valley from the retreating glaciers were unable to make their way to the Irish sea. A lake accumulated and overflowed so as to cut the Ironbridge gorge and thus flow southward to the Bristol channel by what is now the lower Severn valley. In time the gorge became sufficiently deep to drain the lake, and the deposition of great mounds of moraine across the plain in the Wem area prevented the Severn from resuming its preglacial course, even when all the ice had melted. Thus a permanent change of route occurred.

Something similar happened in connection with the Yorkshire Ouse (*q.v.*), where glacial deviation and normal river capture have together turned several Pennine streams into this one river. There are many other examples of the same type.

Other important rivers are the Tyne and Wear from the eastern Pennines; the Trent and Don from the southern Pennines; and the Weaver, Mersey, Ribble, Lune and Eden from the western Pennines. The drainage of the Weald is effected by a number of relatively small rivers which, excepting for the more easterly ones that enter the sea direct, cross the North or South downs by deep gaps. These they have cut as erosion of the Weald dome has progressed, the downward cutting in the chalk keeping pace with the lowering of the surfaces of the Weald itself. Chalk, like limestone, naturally produces a narrow gorgelike valley, such as these gaps exemplify. They are of great importance for communications.

Coasts.—The coastal scenery is quite varied and examples of most types, except fjords, may be found. There are high cliffs along most of Northumberland and Durham, but south of Flamborough head to Suffolk the coast is almost entirely in glacial deposits. From Suffolk to the Thames, it is formed of unconsolidated Pliocene crags or Eocene clays and sands. All these are materials which are easily eroded, and so these coasts regularly retreat landward. The Fens and the Wash interrupt this with lowlying marshes.

Most of the south coast consists of fairly high, to high, cliffs. The Chalk cliffs of Dover (North downs), Beachy head (South downs) and Dorset (Dorset downs) are particularly well-known. The finest cliff scenery of England is undoubtedly to be found in Devon and Cornwall, especially where the resistant granite faces the full onslaught of the Atlantic rollers.

The mouth of the Severn, with its great tidal range, has very extensive mudflats at low-water, and wide beaches occur also in Lancashire, where the tides are unusually high for an open coast. These beaches are most often of sand, and an extensive area of sand dunes occurs along the south Lancashire coast. Solid rock coasts are quite rare between Dee and Solway; St. Bees head (Triassic sandstone), in Cumberland, stands out as one of the few outcrops. The coast is thus mainly of boulder clay, with some dunes, such as those already mentioned, and others north of Barrow, and extensive raised beaches, whose margins form salt marshes in the Ribble estuary, Morecambe bay and Solway firth. So weak a coast can survive with scarcely any erosion because the enclosed nature of the Irish sea prevents the occurrence of big waves.

3. Climate.—England is situated in the low-pressure belt between the Arctic High, which commonly lies over northern Greenland and Spitsbergen, and the Tropical High over the Azores, and for parts of the year over southern Europe. Winds from both these meet in the intervening area and the combat between the two results in the development of local centres of low pressure, which most often originate either south of Iceland or over the Bay of Biscay. The former move in a semicircular path, first southeast, then east across southern Scotland, and northeast toward Scandinavia. The latter move in a general northeasterly direction, crossing the Midlands of England. A procession of such depressions commonly occurs, coupled with secondary effects which often modify the weather considerably.

As a depression passes, south of its centre there are rain showers as the warm, humid, southerly air arrives to replace the cool, dry, northerly. The temperature rises, the winds change from southeasterly to southwesterly and the sky more or less clears as the warm front passes. Behind it there is the cold front with sudden, often heavy rain, a fall in temperature and a change in wind direction to northwesterly. Often the warm air has been lifted completely off the ground by the time the depression reaches England so that part of the effect is missed. Such a depression is said to be occluded.

The high-pressure areas often expand to cover England, sometimes with a centre forming an anticyclone, and sometimes as air masses with a fairly even barometric distribution. These masses may arrive from the northwest, as polar-maritime air; from the east, as polar-continental air; from the Azores, as tropical-continental air. The polar air is naturally cool or cold; the tropical, warm. The maritime is humid, and thus may lead to precipitation. This especially applies to the polar-maritime, which is unstable, and thus likely to produce good visibility. Fog is common with tropical-maritime masses.

Anticyclones often produce heat waves in summer and cold, frosty weather in winter, but under certain conditions fog may occur. This, like most high-pressure weather, may persist for several days, but sometimes it is raised off the ground as a blanket of low cloud, producing the dull, still, dark weather.

England experiences weather rather than climate, for the factors are so variable that mean values, even monthly means, have little significance in connection with the conditions at any specific moment. Nevertheless, they have some value, especially for comparison with other climates. Perhaps the outstanding characteristic of the English climate is this extreme variability. Brilliant sunshine can rapidly change to rain, winds can blow up or alter

their direction, as a warm or cold front passes by. One day may be very different from the next, and prediction cannot be made for more than a very short time ahead. On the whole, however, it is a climate which is exceptionally equable for a temperate one, the winter temperature is high for the latitude, the rainfall is well suited to maintain soil moisture at the temperatures experienced and is evenly distributed through the year. On the other hand, periods of settled weather are rare and gales on the coasts and fogs inland are fairly frequent. It is a climate of few extremes.

Temperature.—In winter there is little difference in temperature between north and south, and Carlisle, Newcastle upon Tyne, Gloucester and London all have a January mean (reduced to sea level) of about 40° F. The northerly towns are no cooler than the southerly because the dominating influence as regards warmth at that time of the year is the westerly air which has traveled over the Gulf stream drift and been heated by it. For that reason Cornwall has a very mild winter (Falmouth has a January mean of 44° F.) and the temperature falls eastward. In July the south at 62° is warmer than the north at 60°, since direct solar heat and therefore latitude is of the greater importance.

These variations result in a more equable climate in Cornwall, with a difference of 17° between summer and winter, than in Essex, where the difference is 24°.

As one result of cloudy skies, which prevent both hot days and cold nights, the mean daily range of temperature is low. At Holyhead, on the Anglesey coast, it is 4° in winter and 8° in summer, although inland it is greater, as for example at Oxford, where the corresponding figures are 10° and 18°.

Frost.—Frost is most frequent in January, February and March, but occasional "killing" frosts may occur as late as May and thus cause considerable damage to fruit blossom. The average number of days on which at least some frost occurs is about 100 a year in most lowland localities, with fewer at coastal stations (Cornwall 50, Lancashire 80) and many more in the uplands. The incidence of frost is one of the most local of weather phenomena and two neighbouring sites may have very differing frequencies.

Sunshine.—Possible sunshine is greater in the north than in the south in summer, but less in winter. In December, northern England could have about 7¼ hours of sunshine each day (sunrise to sunset), and southern, 8 hours. Because of cloud, the actual amounts received are much less: 1½ hours a day in northern coastal areas; ½ to 1 hour in northern inland areas; up to 2 hours a day along the south coast. In June, the possible amounts are 17½ hours and 16½ hours respectively and the actual figures are about 6½ to 7 hours a day on the northern coasts; 6 hours or a little less, inland; and 7½ to 8 hours on the south coast.

Rainfall.—Humid air, and thus the possibility of rain, mostly comes from the westerly maritime winds, and precipitation occurs partly because of the passage of depressions and other pressure systems, and partly because of orographic effects in upland areas. From the former cause the amount of rainfall is greater in the west than in the east and in the south than in the north; and from the latter cause it is greater over highland areas, such as the Lake district, the Pennines, the southwest peninsula and to a lesser degree over the north Yorkshire moors. To the east of the upland masses, there is often a rain-shadow, so that the Vale of York and the lower Trent valley are much drier, as is also the part of England immediately east of the Welsh uplands, and even Somerset east of Exmoor.

The average annual rainfall in low-lying areas on the west coast of England is 30 to 40 in., on the east coast 20 to 25 in., in the Midlands 30 in. In much of the Lake district it exceeds 60 in. and often reaches 80 or 100 in.; on the higher Pennines it is more than 60 in., as also on much of Dartmoor, Exmoor and Bodmin moor. Rainfall is evenly distributed throughout the year, although, due to the increased frequency of thunderstorms, its duration is less during the summer months. Annually most upland areas have more than 200 days with rain (more than 0.01 in. in the 24 hours) and very few areas have fewer than 150 such days.

Snowfall.—It is uncommon for snow to lie on the ground in the lowlands of central England on more than ten mornings a year and the figure is half that on the western and southern coasts. It may be found for much longer periods on the higher mountains, and the highest roads are normally closed to traffic for several days each winter, although this is often due more to icy conditions than to actual snow.

Wind.—Because of the passage of depressions, wind blows from all directions. In general it is most often from the west or southwest, and from each of these directions in most parts of England there is about twice as much wind as from the east. Gales are almost confined to the coasts. There are on an average only two a year in the Midlands, whereas the figure is between 30 and 40 on the west coast and 15 to 25 on the east coast. (R. K. Gr.)

4. Soils and Vegetation.—The soils of England belong principally to two types: podsols and brown forest earths. The former are generally acid and develop where leaching by soil water carries lime and other minerals downward, away from the surface. The latter develop where downwash and evaporation approximately balance, giving a more even balance of soil minerals usually accompanied by a greater concentration of humus. Within this generalization governed by climate, there are many variations chiefly dependent on the parent material of the soil. Again there are two main types, those which have developed from the underlying rock and those which have originated in areas of drift and are not related to the underlying geology. Among those which have developed *in situ* are the acid peat and moorland soils, the alkaline peat or fenland soils, the siliceous soils of highland England, chiefly along the Pennine ridge, and the marls and clays of southeastern England. Of those soils developed from glacial drift covering much of eastern England and the Lancashire-Cheshire plain, the chief are the stiff clays derived from Boulder Clay. Where this clay has been mixed with chalk, as in East Anglia, or with blown sand, as in south Lancashire, it forms a soil of considerable agricultural importance. Glacial drift soils derived from Old Red Sandstone and wind-borne brick-earths are particularly fertile. Gravel and alluvial soils are found in river valleys and estuaries and in some coastal areas. (X.)

The flora of England is relatively poor when compared with that of many other countries. This is mainly because of the destruction brought about by the increasing cold which culminated in the Pleistocene Ice Age. Although the hardier plants survived in southern England and in restricted areas elsewhere, a considerable proportion of the present flora migrated into the country after the period of maximum glaciation. Examination of plant remains in peat and other deposits has made it possible to trace the major changes in the vegetation and some phases of the restocking of the flora. During and after the last retreat of the ice a tundralike vegetation was widespread. Later, birches, Scots pine and hazel became common woody plants. Average temperatures increased and other tree species became established including oak, elm, lime and alder. A climate optimum resulted in the dominance of these broad-leaved deciduous trees, and beech also entered into the English flora. A drier period followed and the influence of Neolithic and Bronze Age man becomes discernible, especially in the increase of weed and ruderal (wasteland) species. A cooler and wetter climate, with a considerable increase in bogs, was followed by the recent period of generally warmer and drier conditions.

England has now an oceanic climate without marked seasonal extremes. The natural climax vegetation over most of its surface is deciduous forest, but because of man's continued activities there is little natural unmodified forest left. It is, however, possible to distinguish differences in vegetation between the warmer south and southwest and the colder north, and between the drier east and the wetter west and northwest. Many of the soils of England are immature, since the older soil layers were removed by the action of land ice and replaced by other deposits or by rock exposures. As a consequence the soils are often differentiated largely on the nature of the subsoil or parent rock. Edaphic (soil) factors are thus of great botanical importance and, in particular, there are striking differences between the vegetation of calcareous and noncalcareous soils. Man has been the chief biotic factor modifying and changing the plant life. By forest destruction, drainage, arable cultivation, livestock raising and the introduction

of weeds and exotics he has changed the botanical face of England. It is, indeed, difficult to say how many "wild" plants are strictly native and perhaps, at most, not more than 1,200 seed-bearing species should be so designated.

If the flora is poor in numbers of species it is systematically varied in that about 90 families and 400 genera are represented among the seedbearing plants. Geographically, the following groupings can be distinguished: widely ranging species, such as broad-leaved pondweed (*Potamogeton natans*), bog myrtle (*Myrica gale*), grass of Parnassus (*Parnassia palustris*) and bulbous buttercup (*Ranunculus bulbosus*); general European species, including European wood anemone (*Anemone nemorosa*), catsear (*Hypochaeris radicata*) and creeping jenny (*Lysimachia nummularia*); some montane to high-mountain plants such as alpine lady's-mantle (*Alchemilla alpina*), moss campion (*Silene acaulis*) and purple saxifrage (*Saxifraga oppositifolia*); central European (continental) species, as milk-vetch (*Astragalus glycyphyllos*), dwarf or burnt orchid (*Orchis ustulata*) and yellow archangel (*Galeobdolon luteum*); west European (oceanic) species, as hemlock wort or water dropwort (*Oenanthe crocata*), ivy-leaved bell flower (*Wahlenbergia hederacea*) and species of gorse (*Ulex*); southern species, as sea spurge (*Euphorbia paralias*), wild madder (*Rubia peregrina*) and tree mallow (*Lavatera arborea*); northern species, including dwarf cornel (*Cornus suecica*), buckbean (*Menyanthes trifoliata*) and *Pyrola* species; and species whose range is centred in or limited to England, as rice-grass or Townsend's cord-grass (*Spartina townsendii*), English monkshood (*Aconitum anglicum*) and a forget-me-not (*Myosotis brevifolia*).

The extensive and varied coast line allows the development of strikingly different coastal communities. Salt marshes occur where mud and silt accumulate, as in many estuaries, and these show zonation from samphire marsh (*Salicornia* marsh), through sea grass marshes, mixed marshes, red fescue marsh and sea rush marsh to dry land. Sand stretches and sand dunes have, in the earlier stages of colonization, an open covering of marram grass, lyme grass and couch or quack grass, and various annuals and perennials able to grow in the superficially dry and more or less mobile surface. Shingle beaches have open communities with sea campion, yellow horned poppy and a dock. Cliff vegetation varies with exposure and the nature of the rock. Aquatic and marsh vegetation is particularly well developed in the Broads and Fens (*q.v.*) of East Anglia. Stages of succession can often be studied in well-defined zones. The flora varies with the depth of the water, its dissolved food content and its degree of acidity or alkalinity. Bogs, dominated by bog-moss, are characteristic of the wetter parts of the north and west. The flora is restricted but includes interesting carnivorous plants, such as sundews and butterworts.

Most English grasslands are artificial, although much of the flora is more or less native on the downlands. Heaths and moors are common on acid substrata with heather, heaths, gorse, bracken, bilberry, heath and moor grasses, and other calcifuges. Brushwoods are, for the most part, stages in succession to forest. In England chalk scrub, limestone scrub and heath scrub show differences in structure and composition. Woodlands, in so far as they are at least seminatural, are of oak and ash over much of the country, sometimes associated with hornbeam and with undergrowth and field layer. On chalk and limestone in the south, beech is often dominant. Alder is common in wetter parts, as along streams. Other English trees are rowan, elms, whitebeam, limes, yew, holly and maple, but a great many species now common, as sycamore and horse chestnut, have been introduced by man. (*See* also *Forestry* below.) (W. B. T.)

5. Animal Life.—The character of the wild fauna of Britain has been determined by the climatic changes that occurred during the Pleistocene period covering the last 600,000 years, and it is doubtful if any animal life was able to survive the second (Mindel) glaciation which lasted 55,000 years and ended about 425,000 years ago. Little of the fauna therefore can have been in continuous occupation for much more than 400,000 years, but some of it probably has been for at least 20,000 years. The fauna has derived from the continent of Europe by migration which ended for terrestrial forms about 7,000 years ago when the Straits of Dover were finally opened. There are thus only about 50 species of land mammals, though many of these have differentiated into subspecies differing from their nearest relations on the continent. Several species such as the bear, wolf and beaver were exterminated in historic times, but others such as the fallow deer, rabbit and rat were introduced by man. The birds are much more numerous in species but though there are many subspecies peculiar to Britain there is only one full species, the red grouse (*Lagopus scoticus*). Many migrant species visit the country for the winter or summer and others pass through but do not stay on their spring or autumn migrations. Although the increasing human population has destroyed many of the former breeding places of some species, the bird fauna is still numerous because of the varied habitats available. There are few kinds of reptiles and amphibians—about half a dozen species of each—but they are nearly all plentiful where conditions suit them. Fresh-water fish are numerous; the char and allied species of the lakes of Westmorland and Cumberland are of interest because they probably represent an old group related to the trout, that formerly migrated to the sea until the tectonic changes that formed these lakes cut off their outlets. The sea fishes are abundant in species and in absolute numbers. The invertebrate fauna is similar to that of the neighbouring continental countries and almost equally immense; some, such as the native subspecies of the large copper butterfly (*Lycaena dispar*), have been exterminated, and others, such as the black-veined white butterfly (*Aporia cratagi*), have disappeared since the mid-19th century, but introductions accidental or deliberate have been added to the fauna. The great diversity of shores produces habitats for a numerous variety of invertebrate animals. (L. H. M.)

II. GEOGRAPHICAL REGIONS

The most fundamental division in the geography of Britain is that between highland and lowland Britain. Rooted in the geological history of the country this division is manifest in geology and climate, in vegetation and mineral resource, in agriculture and industry and in race, language and culture.

The highland areas lie in the north and west of Britain and constitute the major parts of Wales and Scotland; in England the lowland predominates and accounts for the distinctive character of England which contrasts so markedly with the rest of the United Kingdom. Movement from north and west to south and east, from upland to lowland, involves a series of pronounced changes. Sometimes abrupt and sometimes gradual, in many areas these changes mark a transition from old hard and often metamorphosed rocks to younger less disturbed sediments; from relatively high land to undulating lowland; from cool, wet and cloudy climates to rather drier, sunnier conditions; from thin, leached soils difficult to till to deeper and richer soils. These changes in physique are accompanied by vegetational changes from moorlands which provide but rough grazing mainly for sheep to lowlands formerly thickly wooded but now cleared for agriculture, fenced and hedged into the domesticated patchwork of rural England. Lowland England, with one of the highest population densities in the world, presents in all these aspects a major contrast with the lightly populated, largely uncultivated highlands of western and northern Britain.

Within its bounds England contains, however, three major portions of highland Britain: Cumbria in the northwest, the Pennine upland of the north centre and the southwest peninsula. England possesses a richness and variety of landscape rarely equalled in an area of similar size.

1. Highland England.—The uplands of England each possess a geographical individuality though sharing in a broadly similar pattern of rugged relief and moist climate with a stress on pastoralism rather than tillage. Where flanked by Coal Measures their peripheries have become scenes of intense manufacturing activity.

Cumbria.—In the counties of Cumberland and Westmorland and the Furness district of Lancashire lies a denuded and glaciated dome of ancient rocks surrounded by a broken ring of younger sediments. The lakes which occupy many of its ice-deepened valleys have given this upland its name and its character as the

Lake district (*q.v.*). Windermere, Coniston water, Derwent water, Ullswater and Bassenthwaite are among the larger lakes. Relief is strong, waterfalls abound and though altitude is modest (Scafell Pike the highest peak reaches 3,210 ft.) the individual hill masses tower magnificently over the lake-filled valleys and present a scene of mountain grandeur unequalled in England. Slope, soil and climate set close physical limits to agricultural opportunity and the rearing of sheep wintered in valley bottoms and ranged in summer on the moorland grazing of the hills is the main farm economy. The settlement of the farming community is sparse and scattered and the largest towns are the product of tourism which thrives in this area of great scenic beauty, which has been designated as a national park. The climatic and topographic suitability of the Lake district for the construction of reservoirs has made this area of high rainfall an important source of water for the industrial area of southeast Lancashire 80 mi. to the south.

Around the highland core, on the younger rocks, farming is more varied and tillage more extensive though the emphasis as in most of western Britain is on livestock and in particular the production of milk. To the east the fertile Vale of Eden forms the eastern lowland of Cumbria and provides a north-south route along the eastern edge of the Lakeland dome. In the mouth of the valley Carlisle, a fortress border town, is the focal point and servicing centre of northern Cumbria. On the coastal lowlands to the west of the hills, Coal Measures and hematite deposits have given rise to mining towns and iron works. The coalfield is becoming increasingly expensive to mine and the local iron ores are now approaching exhaustion; iron and steel making is carried on with the use of imported ores. Barrow-in-Furness is a shipbuilding centre. A post-World War II development in this somewhat isolated industrial area has been the establishment of the atomic plants at Winscales (begun in 1947) and Calder Hall (opened in 1956).

Pennine Region.—The Pennine region consists of the Pennine upland of central northern England, and the Pennine flank coalfields which though distinct from the upland are linked to it in geology and economy. The carboniferous rocks of the Pennine anticline have been eroded in varying degrees and only along the dipping flanks of the range have coal deposits survived in economic quantities.

In the north, the trough of the Tyne valley separates the Pennines proper from their northerly extension the Cheviots, which stretch to the Scottish border. The Cheviot country is largely of carboniferous sandstones; its bleak flat summits are covered by peaty moorlands which surround the rounded granite mass of Cheviot itself (2,676 ft.). The economy as in so much of the Pennine upland is of sheep farming with farmsteads in the valley bottom. In the Kielder area along the border an extensive plantation of conifers has been established by the Forestry commission.

South of the Tyne the northern Pennines form a massif upland block of limestone dipping gently eastward from the high fault scarp of Cross Fell (2,930 ft.). Population is slight and sheep farming the mainstay of the economy. The headwaters of the Eden and the Tees separate the northern Pennines from the central which in turn are bounded on the south by the Aire gap. This central upland is made up of massif limestones in the west which dip beneath Millstone Grits to the east. In both areas the field boundaries are of local stone built into walls which lie like great nets over hill and dale alike. The upland grazings thus divided again carry sheep but in the valleys the deeper alluvial soils sustain better grass, and cattle become more important than sheep. The dales in the limestone of the West Riding have a scenic attraction of their own compounded of the green turf of the fescue pastures, the white network of walls, the simple dignity of the gray stone buildings, the occasional grove of beech and the bold face of limestone crag. South of the Aire the gritstone uplands present a less attractive landscape but are invaluable sources of soft water for the textile industries of both Lancashire and Yorkshire, while their valleys form suitable sites for strings of reservoirs. In the south in Derbyshire the Pennine anticline is more pronounced and, as in the north, the Coal Measures and gritstones have been stripped away to expose the limestone beneath. The southern edge of the continuous gritstone forms the peat-covered plateau of the Peak district and ends abruptly to overlook the limestones of Derbyshire which are similar in scenery and economy to the Yorkshire dales.

Pennine Coal Fields.—Coal represents the most important British mineral resource and its presence in quantity on both flanks of the Pennines has led to great industrial developments and concentrations of manufacturing towns.

Southeast Lancashire is the cotton manufacturing region of Britain, more than 95% of all cotton workers working within the county. Favoured by an ample supply of soft water for textile processing, of water and then later coal for power, by a humid climate and by the facilities of the great port of Liverpool, cotton manufacturing became the dominant industry of the area. The cotton towns are concentrated into two main groups. In a great semicircle round Manchester towns such as Bolton, Bury, Heywood, Rochdale, Oldham and Stockport concentrate upon carding and spinning processes though some weaving is found in each town. Specialization occurs among the spinning towns, some such as Bolton spinning fine yarns, others like Oldham dealing more in medium and coarse cloths. Manchester itself, while carrying on some cotton manufactures, serves more importantly as the commercial centre of the whole area, while Liverpool is the main source of raw cotton imports. The second important area lies in the valleys of Rossendale forest and on the small coalfield to the north of it. This is the main weaving area, though again other processes are also found. The chief towns are Blackburn, Burnley, Nelson, Colne, Darwen and Accrington.

In addition to the cotton industry a wide range of manufactures has developed, some once ancilliary to the cotton trade have developed beyond their original connection with cotton. The manufacture of textile machinery and mill fittings is now only part of a widespread engineering complex, and chemical manufacture ranges beyond dyes and bleaches. The cotton industry suffered a great decline during the 20th century as the increased output of cotton cloth overseas reduced the exports of cotton goods upon which the prosperity of Lancashire had so long depended. Many former cotton mills house other industries that were attracted to Lancashire by reason of the cheaply available buildings and the pool of skilled labour. The industrial structure of the cotton area is much more diverse than formerly.

The industrial area based on the small north Staffordshire coal field has likewise developed as a specialist region. Named after the staple of their industries, the chief towns of the area are known collectively as the Potteries (*q.v.*), the five most important now being united into the administrative unit of the city of Stoke-on-Trent. More than three-quarters of the pottery workers of Britain live in this area.

On the eastern flanks of the Pennines the coal field in Northumberland and Durham has also given rise to vigorous industrial and urban development. North of the Tyne the Coal Measures outcrop on the coast, their accessibility giving rise at an early date to a coal export trade first to London and later abroad. South of this essentially mining area, upon the lower Tyne, the conurbation of northeastern England has grown up round Newcastle. Born of coal and the coal trade, these towns grew with the development of iron and steel manufacture, shipbuilding, engineering, metal refining and chemical manufacture. Many of these activities suffered acutely during the economic depression of the 1930s and the area became one of chronic unemployment. Though the export of coal was much reduced, the area recovered with a more broadly based industrial structure. The industrial estate in the Team valley south of Gateshead was located specifically to attract light industry to Tyneside and so diversify its economy. South of the Tyne between the Wear and the sea, the Coal Measures are covered by Permian limestone and are worked by large collieries. West of this limestone the exposed Coal Measures have been long and intensively worked and reserves much depleted. The area is still important despite its high costs and approaching exhaustion, for it possesses the best metallurgical coking coal in Britain. At Consett is situated a large integrated iron and steel works using this local coal and imported ore. Sunderland at the mouth of the Wear exports coal and has shipbuilding yards. On the Triassic

rocks south of the coal field the area round the estuary of the Tees is the country's largest producer of pig iron and second, to South Wales, of steel. This industry based originally on Durham coke and Cleveland iron ore now depends to a very large extent on imported ores. Engineering, particularly structural engineering, shipbuilding and chemical manufacture are other important industries; the last established at Billingham and since World War II at Wilton uses local salt and gypsum deposits and Durham coal.

Between the Tees and the Wharfe, Coal Measures are largely absent but from Leeds in the north to Nottingham to the south they occur at the surface along the Pennine foothills dipping gently eastward to disappear beneath Permian limestones as in eastern Durham. More than one-fifth of British coal is mined in this field at a cost considerably below the national average; its industries are various but the most famous and extensive is the wool textile industry of the West Riding of Yorkshire. Favoured by features similar to those pertaining in Lancashire, Yorkshire has become the leading wool textile region of Britain. The numerous towns form a conurbation with a population exceeding 1,500,000 persons. Worsted cloths tend to be the most important fabrics produced in the area round Bradford, Halifax and Keighley, while woolens are more characteristic near Wakefield and Dewsbury, the latter being the centre of shoddy and mungo manufacture. Leeds, the largest city, is of secondary importance as a textile centre but of supreme importance in the manufacture of mass-produced clothing; in addition engineering, including the construction of locomotives, is important.

Farther south in the coal field, the dominant industry changes and in the valley of the Don between Sheffield and Rotherham the manufacture of iron, and more particularly steels, is of overwhelming importance. The area is celebrated for cutlery and machine tools and is the main producer of high quality alloy steels in Britain.

In the counties of Derby and Nottingham, in the southern part of the coal field, Coal Measure iron ores led to the early growth of the iron industry which persists, now however using ore brought into the district; such works using local coal are to be found near Chesterfield and to the west of Nottingham. Nottingham, the largest city of the southern coal field, famous for its lace, hosiery and knitwear, has a wide variety of industries whose products include pharmaceutical chemicals and bicycles. Derby, a little to the west of the coal field, is an engineering centre.

Southwest Peninsula.—The third of the upland areas of England comprises the southwest peninsula. It is lower than Cumbria and the Pennines and the hill land is less continuous, furthermore its position in the southwestern seas of Britain makes its climate moist but mild, especially along the coasts. The rocks are old, hard Paleozoics and over wide areas of north Devon and Cornwall of the same age as the Pennines. Coal is however absent and the mineral resources of the area are the result of the intrusion of granite masses which metamorphosed and mineralized the surrounding rocks. Dartmoor, the largest and highest (High Willhays, 2,039 ft.) of these exhumed granite masses has little mineral wealth but further west smaller intrusions between St. Austell and Land's End resulted in the formation of tin and copper ores which were famous for centuries. The discovery of richer and more cheaply worked deposits overseas led to the depression of mining, and abandoned workings litter the moors. Tin mining has, however, been resumed on a small scale. The most important mineral resource is kaolin, the product of the decomposition of granite. This raw material of the porcelain industry is quarried in great open pits on Hensbarrow north of St. Austell, Bodmin moor and, to a lesser extent, on southern Dartmoor; the white spoil heaps are one of the striking features of the landscape. For the rest, the peninsula is very largely devoted to farming in which livestock products are of greatest importance. In sheltered bays along the coast, market gardening and fruit farming take advantage of the mildest winters and earliest springs on the English mainland, and supply the London market with early produce. The fine cliffed coast line is scenically attractive and provides sheltered harbours; most are small but some, such as those of Falmouth, Dartmouth and Plymouth, are deep and large and could have developed into important commercial ports but for the distance from the main population centres of the country. Plymouth, the largest town of the peninsula, is a naval base and a calling point for liners. Fishing, once the staple economy of the villages, has declined, but the picturesque appearance of the settlements, the attractiveness of the coastline and the mild and sunny climate have made coastal Devon and Cornwall popular tourist areas.

2. Lowland England.—The lowlands of England which surround the highland masses reach the west coast only in the vicinity of the Bristol channel and along the shores of Lancashire and Cheshire. The lowlands can be subdivided into the Midland plains which are largely of Triassic sandstones and marls but include older rocks such as Coal Measures, and the scarplands of eastern England where rocks younger than the Trias are arranged in long sweeps of scarp and vale stretching from Dorset in the southwest to the moors of Cleveland in the northeast.

Midland Plains.—The Midland plains are made up of the core region of the Midlands proper, a low plateau in the heart of England; of northerly reaching arms of lowland which become the vales of Trent and York in the east and the plain of Lancastria in the west; of the Severn lowlands and the Bristol area; and lastly of the marchlands along the edge of the Welsh massif.

The central region of plain and low plateau is one of internal contrasts. Within a rural setting lies the great conurbation of the west Midlands with its 2,250,000 people. This conurbation is the product of manufacturing originally based on the Coal Measures exposed within the Triassic plain. The Midland coal fields have yielded both coal and iron and their exploitation particularly in south Staffordshire has led to the growth of Birmingham and the Black Country (*q.v.*), where in the 19th century the sky was "lit up with the glare of a myriad fiery furnaces." During the Industrial Revolution this area became important for its production of iron and brass and the subsequent working up of these metals into hardwares. Coal and iron resources are now largely exhausted but the skill and tradition of metalworking persists. Metalworking led to engineering and during the 20th century the manufacture of motor vehicles has become important, especially in and around Birmingham and Coventry. The Midland conurbation produces a wide range of goods from engineering, rubber, clothing and armament factories and is a producer of finished articles rather than a processor of primary raw materials; this befits its inland position in a country where many raw materials are imported. Away from the conurbation the countryside is a fertile land of mixed husbandry, the heavier soils concerned largely with dairying, while sandy areas such as that round Kidderminster engage in market gardening.

To the northeast the lowland reaches up between the coal fields in the west and the scarplands to the east as far as the valley of the Tees. The drainage of this lowland is centred on the Humber which, having pierced the limestone scarp at its lowest point, has drawn to itself the drainage systems of the Trent in the south and the Ouse in the north. The Vale of Trent and its tributary valleys carries some of the finest beef-fattening pasture in England. In Leicester on the Soar, a tributary of the Trent, the ancient wool textile industry has developed into hosiery and knitwear manufacture while the town is also an important centre of boot and shoe making. The middle to lower Trent valley is largely agricultural, its emphasis shifting from grazing to cash crop farming as the alluvial lands of rich silt at the head of the Humber are reached. The Vale of Trent with its water supplies for cooling, its space for development and its nearness to the lowest-cost coal field in Britain has become the site for the construction of many large coal-fired electricity generating stations.

The master stream flowing into the Humber from the north is the Ouse; its valley, the Vale of York, is the northernmost limb of the red-rocked plain of the midlands. The southern portion round Selby is a silt-soiled area of cash cropping while farther north, near York, mixed farming with a stress on livestock prevails. York, with its magnificent minster, owes its early importance to its central position in the vale on a morainic ridge which facilitated east-west movement over the once marshy lowland.

The northwesterly arm of the Midland plains, stretched up be-

tween the southern Pennines and the Welsh massif, widens out into the Cheshire plain and in Lancashire forms an ever-narrowing apron to the Pennines until it peters out along the shores of Morecambe bay where the ancient castle town of Lancaster guards its northern approaches. The plain is largely rural; Cheshire with its lush pastures is the most famous of English dairying counties but has industrial towns on the salt fields round Northwich and along the Mersey shores of the Wirral peninsula. Liverpool, on the Mersey estuary, is second only to London among the ports of Britain and is a great manufacturing centre; the industries include not only those typical of a great port, such as sugar refining and flour milling, oil seed crushing and oil refining, shipbuilding and soap manufacture, but in addition cover a wide range of engineering products, clothing, chemicals and man-made fibres and many foodstuffs. North of the port the Lancashire plain is a rich arable area of peaty and sandy soils; the coast is lined by sand dunes with a gently sloping beach which continues north to Morecambe bay and has become a holiday playground for the industrial north of England with such resorts as Southport, Blackpool and Morecambe.

The valley of the lower Severn is the main component in the extension of the central lowlands to the southwest. Its major tributary, the Warwickshire Avon, flowing down from the heart of the Midlands, in its lower portions forms the Vale of Evesham celebrated for its fruit and market-garden crops. The estuary of the Severn, reaching deep into the country, is an obstacle to land travel. It is crossed at its head by a railway and farther downstream by the Severn railway tunnel.

Bristol is a major port and industrial centre. Long an important centre of commerce, its growth was checked in the 19th century by the unsuitability of the Bristol Avon for modern ships. An outport has been developed at Avonmouth on the Severn estuary. The limestone Mendip hills separate Bristol from the isolated lowland of the plain of Somerset. These peaty and once marshy levels, now largely drained, carry great numbers of dairy cattle; Cheddar cheese and cider apples are noteworthy products.

Between the English lowland and the mountains of Wales the border counties of Monmouthshire, Herefordshire, Worcestershire and Shropshire partake of the physique and culture of both. Celtic and Anglo-Saxon place names occur together, lowland gives way to highland, and tillage and dairying to cattle rearing and sheep farming. Worcestershire is more largely a county of the Vale of Severn and together with Herefordshire is known for its orchards, its hops and its beef cattle. The light soils of the Old Red Sandstones of Herefordshire are particularly noted for hops, while the county's cattle are justly famous having been exported as breeding stock to all parts of the world. The towns of the border country are all market centres handling in particular the livestock products of both the marchlands and the Welsh hills. From Monmouth and Hereford in the south to Oswestry, Wrexham and Chester in the north, the border towns are centres for sheep and cattle sales. The varied topography, the entrenched rivers, the half-timbered buildings of town and village, the mixture of wooded hill and cultivated field, of hopyard and meadow all combine to give the Welsh Marches a rare and rustic beauty.

Scarplands of Eastern England.—England to the east of the Midland plains is made up largely of alternate belts of soft and hard rocks etched into scarps and vales and orientated southwest to northeast (see *Physiography* above). This cuesta landscape contains several distinctive subdivisions: the Scarplands from the Tees to the Exe; the Fenland; East Anglia; the London basin; the Weald of southeast England; and the Hampshire basin.

The rocks of eastern England dip, for the most part, gently to the southeast, the oldest sediments abutting on the Triassic rocks of the Midlands. The first major scarp feature is formed by Jurassic rocks and, stretching from Dorset to the North Riding, is known variously as the Cotswolds, Northampton uplands, Lincoln edge and North Yorkshire moors. Characterized by arable farming with sheep, except in Yorkshire where thin podsolized soils on sandstone give rise to heather moorland fit only for rough grazing, the Jurassic scarps also contain England's most important iron deposits. Though lean they can be worked cheaply in open pits and have given rise to iron and steel works, notably in Northamptonshire, north Lincolnshire and on the Tees (see *The Economy* below).

Behind the Jurassic scarp is a wide vale of soft, often clayey rocks which in turn dip beneath the second major scarp, that of the Chalk. The clay vale can itself be subdivided. In the southwest, in Wiltshire, the Vale of Pewsey forms a fertile valley tributary to the Bristol Avon which drains the western end of the clay vale. The central part of the vale, the Oxford clay vale, is drained by the Thames, which pierces the Chalk barrier of the Chilterns to form the Goring gap and flows into the Thames basin. In the east, the vale is drained to the Wash by such rivers as the Nen and the Ouse. The whole of this valley tract is predominantly agricultural with much of the heavy land and river alluvium in pasture; in the east, however, the lower rainfall and more varied, lighter soils encourage arable farming in Huntingdonshire and Bedfordshire and some horticulture, particularly around Biggleswade. The three towns of size within the vale are Swindon in the west, a railway centre with engineering workshops, Oxford in the centre, the largest of the towns, and Bedford, another engineering centre, in the east. Oxford has a dual personality; the medieval town with its university is one of the great cultural centres of western Europe, while Cowley, its 20th-century suburb, is a major producer of motor vehicles.

In Huntingdonshire the vale widens out to become the great levels of the Fenland interrupting alike the northward sweep of clay vale and chalk down. The Fenland, an area of peat and silt deposits, is one of the most distinctive in England. Once a desolate area of malarial swampland, the peats were drained as a result of many centuries of reclamation. The clay islands which arise from this flat bog provided safe sites for early settlements; Ely with its great cathedral stands on one such mound. Once drained, the peats and silts have become the most fertile soils in England yielding heavily of a wide variety of cash crops. Potatoes, wheat, sugar beet and vegetables are the main crops on the peat, while the silts support in addition a number of specialist economies such as soft fruit production near Wisbech and flower and bulb growing round Spalding. Production of peas and other vegetables under contract for the major canning and packing companies is particularly important in the Fenland. On the southern edge of Fenland is Cambridge with its great and ancient university; industrial development has been on a much smaller scale than in Oxford, so that the town retains to a greater degree the appearance and character of a university city.

To the north of the Fens the clay vale reappears in Lincolnshire and in landscape and economy is similar to that in Buckinghamshire. Beyond the Humber the outcrop of Jurassic rocks is too narrow to be significant until along the northern edge of the chalk wolds the clays reappear in the Vale of Pickering, where, together with the lacustrine deposits of the proglacial Lake Pickering, their soils support mixed farming.

The western edge of the Chalk layer which once covered southeastern England, cut back by the expansion of the clay vales, forms the second major cuesta of the English lowlands. It is marked by a line of rounded hills which presents a steep face to north and west and slopes gently to the east and south from Flamborough head in Yorkshire to the western downs of Dorset. On these low hills surface drainage is limited to major valleys, soils are thin and dry, and woodland, frequently of beech, is restricted to higher summits and steep scarp faces. At one time characterized by a cover of short fescue turf, the chalklands are now largely in arable. Sheep have become less important than dairy cattle fed on the arable crops. Villages are often sited at the foot of the scarp slopes where springs emerge and where soils are more varied and deeper. The Chalk outcrop is more complex than the Jurassic. One cuesta runs from the Yorkshire wolds through Lincolnshire to the Wash; in East Anglia the Chalk, largely covered by drift, sweeps in a broad belt into Cambridgeshire where it swings to the southwest to become the Chilterns. The central hub of the chalkland is in Berkshire and Wiltshire from which the line of hills continues to the coast in Dorset. Southeast of the main cuesta the Chalk dips beneath the Tertiary clays of the London basin and rises again into the denuded anticline of the Weald where erosion

has stripped away the Chalk cover to leave but a ring of inward facing scarps which form the North and South Downs. South of the Salisbury plain the Chalk also dips under later deposits to form the Hampshire basin, re-emerging again as a steeply tilted narrow outcrop in the Isle of Wight and Purbeck Downs.

The Wolds of Yorkshire are largely drift free and have an arable sheep economy. To the east, where the Chalk dips below sea level, Holderness exists solely by reason of the thick Boulder clay cover on the Chalk. It is drained by the river Hull at the mouth of which, on the Humber estuary, is sited the city of Hull (Kingston upon Hull), a major port, industrial and fishing centre. South of the Humber the dip slope of the Chalk covered by drift and marine silts forms the cattle-grazing area of the Lincoln marsh. The sandy coast favours holiday resorts but its most important town is the fishing port of Grimsby. The Chalk interrupted by the Wash and the Fenland reappears again in Norfolk, but over much of East Anglia it is covered by glacial drift and in the east by Pliocene sands and gravels. Given cohesion by its somewhat isolated eastern location, East Anglia does exhibit internal variety: from the good sands of the northwest, the home of the New Husbandry of the agrarian revolution, to the sandy wastes of Breckland now largely afforested, and from the rich loamy wheatlands of the northeast through the marshes and meres of the Broads to the dry "sandlings" of east Suffolk. Predominantly agricultural, it is one of the most extensive arable areas in England. It has ancient market towns, such as the cathedral city of Norwich and Bury St. Edmunds, minor ports such as Harwich, fishing centres in Lowestoft and Yarmouth and a score of quiet villages where names such as Worstead and Kersey are the only indications that East Anglia was once the major wool textile area of Britain.

The London basin occupies a triangular depression filled by Tertiary clays, sands and gravels between the converging Chalk ridges of the Chilterns and North downs. Widening eastward, it is drained by the Thames and its tributaries and reaches the sea in a wide marsh-fringed estuary. Heavy London clay predominates, though sands and gravels have been important in locating early settlements, in facilitating market gardening and in giving rise to heathlands formerly despised but now prized as open spaces. The whole basin is dominated by Greater London, a man-made region where more than 9,000,000 people live in one of the greatest settlements on earth. (*See* LONDON.) Still expanding, the metropolis has overlapped the Chalk rim of the basin and spilled into the country beyond. It dominates the life of Britain; road and rail focus upon it; it is the chief port and an important manufacturing centre; about one-fifth of England's population is concentrated there, and it is the country's most important market and commercial centre.

The dissected dome of the Weald, rising to the south of the London basin, is likewise influenced by the presence of the metropolis; its northern rim has already been penetrated by the outer suburbs while the once rural villages have swollen into dormitories of the capital, particularly along the electric railways which cross the area. The resorts of the Channel coast, such as Brighton and Eastbourne, owe their prosperity to London, while the Channel ports, such as Dover and Folkestone, are closely linked by rail. The successive outcrops of Lower Cretaceous rocks which are exposed within the anticline provide a variety of landscapes. The encircling downs are classic chalk country, the clay vales once heavily wooded are now largely pastureland, while the Upper Greensand at the base of the Chalk is settled by a whole string of villages working its fertile loams watered by scarp-foot springs. The more massif Lower Greensand is less fertile and still carries wide expanses of heath and woodland, such as upon Leith hill and Hindhead, and it is a fashionable residential area for people working in London. The central Wealden Sandstones with their expanses of oak, birch and pine still further contribute to the woodlands of the Weald, so named because of its once extensive woodland cover. The grassy claylands of the Weald support herds that supply milk to London, but in other areas, notably in Kent where soils are more loamy, hops, wheat and orchard fruit are products of a rich and varied husbandry. Along the coast where the Wealden anticline has been truncated by the sea, the white Chalk cliffs of Dover, which gave England its early name of Albion, give way in turn to the flats of Romney marsh at the mouth of the clay vales, to the low cliffs of Wealden sandstone near Folkestone, to marsh again in the Pevensey levels, until finally the Chalk reappears in the fine high cliffs of Beachy head.

Beyond Beachy head the Chalk dips to cradle the Tertiary deposits of the Hampshire basin. With its apex in the downland round Salisbury, the basin forms a triangle with its base along the south coast. As in the London basin, clays predominate, but sands and gravels give rise to heath and woodland such as the New forest. The port of Southampton on the common estuary of the Test and Itchen, which, with the Avon, drain the basin, is the focal point of the region. Blessed by deep water and double tides, Southampton is Britain's leading passenger port. Portsmouth, with another fine harbour, is an important naval base. Inland the ancient ecclesiastical towns of Salisbury and Winchester, which grew up at nodal points on the chalkland routeways, have lost their former importance but retain much of their urban attractiveness. (E. S. St.)

III. THE PEOPLE

The British Isles, off Europe's west coast, have links to south and north and have received immigrants and stimuli thence for thousands of years. Successive cultures have overlaid one another in southeastern England and have struggled for acceptance in the hill and moorland country while, along the Atlantic coast, another series of influences, again from north and south, have operated in centuries long ago. An incomplete amalgam of these diversities has made the English people.

1. Early Ethnic and Social History.—The prehistoric contributions include many physical characters. Long-headed skulls of people of moderate stature from the days before metal can be matched among brunet elements, often with gray eyes, in modern populations. Some of these early people came from the south (Mediterranean) and built great tombs of unhewn stone, especially in the Atlantic coast lands from about 3000 B.C. onward (dating by carbon). Sites of early cultivation in Ireland and England have now been dated somewhat before 3000 B.C. In the early age of metal in Britain, about 1900–1800 B.C., makers of beaker pottery, often tall and broad-headed, came, chiefly from the Low Countries, and in some association with the great tomb builders constructed such great circles of standing stones as Avebury and Stonehenge (*qq.v.*). Apparently they were absorbed into the general population through intermarriage but their physical type still persists in various families. (See *Archaeology* below.) The immigrant Belgic lords, just before Roman times, brought an improved plow and germs of urbanism (Colchester, Verulamium, Winchester, perhaps London) and local military units (Iceni, Cantii, Regni), typically with Belgic leaders and indigenous underlings. Roman power made London a commercial city of perhaps 20,000 people, and towns and roads were built in the English plain, with rural villas (farms with Romanized owners and indigenous labourers). (*See* BRITAIN.) Half-learned Latin displaced kitchen Celtic and, when the Romans withdrew (A.D. 410–446) and towns and villas decayed, a leaderless people without a vital language of their own, and without organization, were at the mercy of conquerors. The case was very different among the Welsh hills, where Celtic bards sang the praises of their princes.

At this time, Slavonic speakers with iron axes were cutting forests on the north German plain as they expanded westward in a period of temporarily increasing sunshine. This pressed Germanic people on and some came to Britain, often as migrant families, sometimes as military bands. Family migrants, to judge from early post-Roman cemeteries with cremation burials, came up the Humber to east Yorkshire and up the southern rivers draining to the Wash, and here in modern populations is still found a rather over-average percentage of blood group A, a feature of northern Germany. Anglians from Schleswig, apparently with a tradition of small farms, and other peoples (including Saxons) with a tradition of communal working of long plow strips in the open fields of the village brought their diverse contributions, and surplus young

men spread westward, marrying Romano-British girls and taking control of land often worked by indigenous serfs. The lack of a vital native language allowed the newcomers to spread their dialects as the basis of modern English. A Celtic king still held Lincoln in the 6th century and Elmet, west of the Yorkshire Ouse, also long remained Celtic. Anglo-Saxons in southern England were not always victorious and the Chiltern Hundreds remained an area needing special arrangements. In Elmet and the Chiltern Hundreds the short, brunet element in the local population is still noticcable. English speech triumphed before long, east of Offa's Dyke, and the written language became important under Alfred (848–899).

The early English, mainly in self-contained villages away from brigands on ex-Roman roads, only slowly attained a measure of unity under Egbert and Alfred and his descendants, with some growth of towns in the 10th century. The early English were probably less brunet than the average Romano-British.

Tall blond Norse raiders from A.D. 793 added a new element to coastal populations, spread Norse place names in Cumbria and so divided Celtic Scotland from Celtic Wales, and developed trade between Norse Dublin and Bristol. In this temporary phase of a rather warmer climate, Denmark, as well as Norway, became a centre of activity and Danish military men invaded eastern England; they were interested in roads and towns as well as in village life. The Danelaw, including East Anglia and parts of the northeast Midlands, Lincolnshire, Yorkshire and the Lancashire Fylde, eventually amalgamated with Anglo-Saxon England as the invaders were assimilated into the population through intermarriage.

Celtic Christianity in Atlantic Britain, isolated from Roman influence, had sent missions as far as Basel and Vienna, but weakened in face of the Romanist new-beginning under St. Augustine of Canterbury from 597. Bishoprics were established with, mostly, abbey-cathedrals manned by regular clergy necessarily exempted to some extent from monastic ritual because of their administrative tasks. In Gaul, on the other hand, many cathedrals were under secular clergy, the bishopries having survived from Roman imperial times. In England St. Dunstan effected reforms which strengthened church and, to some extent, civic life in the 10th century.

2. Social Development: Middle Ages to Modern Times.— The Norman conquerors (1066 onward) soon built castles and many churches, spread water mills for grinding corn, compiled the Domesday survey to systematize dues on land, and developed towns and trade with the continent, in connection with which bishoprics were established at the ports of wool export (Lincoln, Norwich, Chichester and Exeter). Cistercian monks (1190 onward), as sheepmasters, brought remote rural valleys into the active national economy and Hanseatic merchants (1250 onward) promoted London's commerce and their "pound of the easterlings" is the parent of the pound sterling. Flemish weavers (14th century) helped cloth manufacture, especially in Norfolk, to supersede export of raw wool.

Though Norman French speech dominated for a while at court, English was well established among the people and was strong enough to borrow from Norman French without being submerged. Intermarriage told in its favour, especially after King John lost Normandy (1204), and, in the 14th century, Chaucer, Langland and Wycliffe gave the language splendid expression (*see* ENGLISH LANGUAGE; ENGLISH LITERATURE). Henry V's expropriation (1415) of properties of continental religious houses further promoted national self-consciousness. Church architecture developed specifically English features from about 1200 leading on to the idea of stone as a frame for glass—so notably illustrated at King's college, Cambridge—a scheme appropriate to England's cool and cloudy climate contrasting with southern small windows and mosaics.

Justices in eyre (circuit court) from Henry II's reign onward developed the legal system, but customary law was too strong to be ousted, and what is called the "common law" is still the most characteristic national feature. Magna Carta (1215) strengthened the legal processes against arbitrary interference. *Placita de quo warranto* (pleas by what warrant), especially under Edward I (1272–1307), effected some systematization but local custom remained strong. A slight feature, but for that reason almost the more notable, is the general use of the solar calendar, with dates in March, June, September and December for many payments of rents and other dues in most of England, while in parts of the northwest and in Wales and Scotland there are survivals of the old vegetational calendar, emphasizing May and November.

In the 13th century the universities of Oxford and Cambridge took definite shape, with groups of teachers and students in halls and soon, more organized, in colleges. This development, away from royal and ecclesiastical capitals, gave them the character they have maintained ever since, and distinguishes them from the centrally controlled and more authoritarian university of Paris.

Throughout the middle ages (*c.* 1150 onward) a money economy was undermining feudal schemes of enforced personal service, and this tendency became stronger after the Black Death (1348–49), because heavy mortality had made labour scarce. Anglo-Welsh archers defeated feudal knighthood at Crécy, Poitiers and Agincourt, and then French cannon brought England to final defeat in the Hundred Years' War and also made the old castles and feudalism out of date. Meanwhile, brickmaking had made fireplaces and chimneys more practical, and residences with more family privacy reshaped social life.

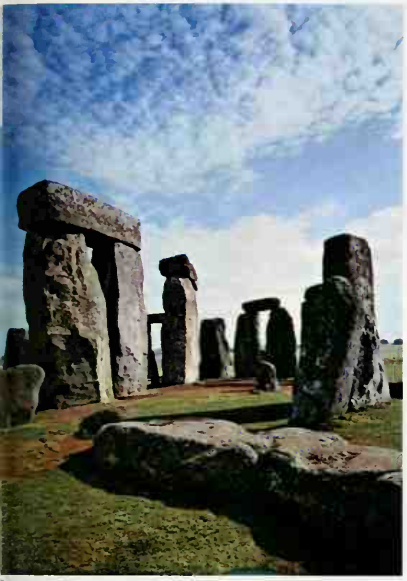
Under the Tudor sovereigns (1485 onward) came a marked growth of royal power and of national self-consciousness. This found expression as a result (not a cause) of the separation of the English Church from the Roman Church, with the consequence of the literary triumphs of the Book of Common Prayer and the vernacular translations of the Scriptures. To all this was added Elizabeth I's abolition of the Hanseatic power over English commerce and the foundation of the maritime trading companies. Alongside these more or less official schemes there came the major blossoming of English literature, with the supremacy of Shakespeare.

Huguenot refugees from persecution contributed to England's industry and commerce, and also diversified the country's religious life by supporting Calvinist doctrine in Congregationalist, Presbyterian and Baptist dissenting churches. After 1688 toleration developed and in the second half of the 18th century the Methodist Churches, based on Arminian and therefore anti-Calvinist teaching, led to further religious diversification. The Methodist Churches gave birth to the Salvation Army, as an expression of social religion, in the latter part of the 19th century under William Booth, previously a Methodist minister. In the 19th century, also, Roman Catholic disabilities were rescinded. In the 20th century new streams of refugees again enriched English life.

From the 17th century the principle of free inquiry made its influence felt in several ways, notably in the Royal society (1660) and, in the religious sphere, through the Society of Friends, small in numbers but great in its effect on social life.

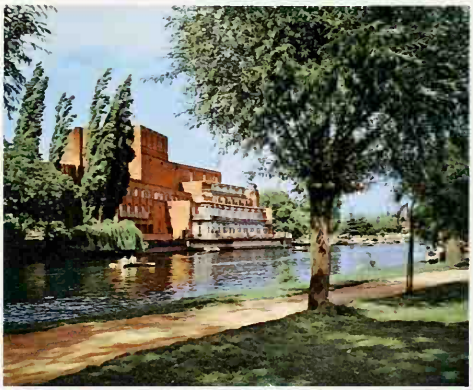
It has long been a social feature that the wife, in general, takes her husband's status. Members of the peerage families frequently married commoners. In such cases, the daughter of a peer married to a commoner retained her courtesy title, but a peer married to a commoner conferred his title upon her. It is possible that this social custom has contributed to the survival of the peerage. At various crises, also, new families have acquired rank, and sometimes titles, as landowners. Illustrations are the granting of former monastic lands to laymen under Henry VIII, the creation of peerages given to merchants under William Pitt, and the Honours list of modern times. These recruitments of the, until recently, more powerful social grades have weakened the practical influence of class distinction, while maintaining it in a milder form, indicated, for instance, by peerages for trade-union leaders and by life peerages. (*See* PEERAGE.)

The English people are a polymorphic mixture of many elements at many periods of a long and unusually continuous history of slow unification, without too persistent imposition of uniformity. England north of Humber and Mersey still has a feeling of distinctness, without separateness, from the southern lowlands, in which East Anglia and the southwest, at least, still have a marked individuality, whatever the immense growth of industrial towns and



Stonehenge on Salisbury plain, Wiltshire, the earliest part dating from about 1800 B.C.

Bodiam castle (14th century) surrounded by a broad moat, near Robertsbridge, E. Sussex



The Royal Shakespeare Memorial theatre at Stratford-upon-Avon, Warwickshire

Salisbury cathedral, Wiltshire, a striking example of Early English architecture

Hawkshead church, 16th century, in the Lake district, Lancashire

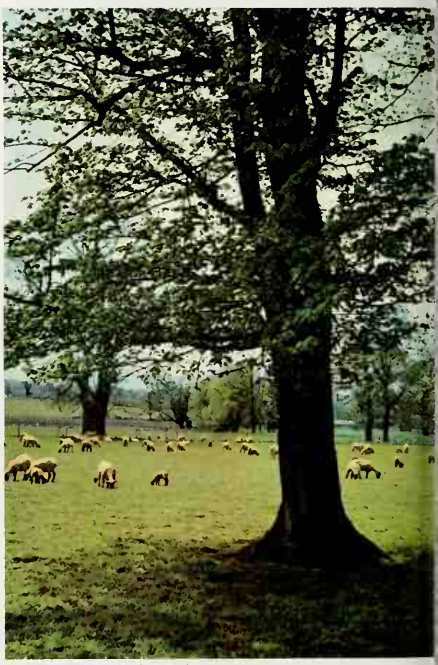
ARCHITECTURAL SETTINGS

PHOTOGRAPHS (TOP LEFT, BOTTOM LEFT) GEORGE AUSTIN FROM SHOSTAL, (TOP RIGHT) RAY MANLEY FROM SHOSTAL, (CENTRE RIGHT, BOTTOM RIGHT) ADE WILLIAMS FROM SHOSTAL

ENGLAND

Ashness bridge, above Derwentwater, Cumberland, with Skiddaw in the background

Arlington Row in Bibury, a noted example of stone-built cottages (15th century) in the Cotswold country, Gloucestershire



Sheep grazing in a Suffolk pasture

A harvest scene near Lewes, East Sussex

RURAL AND AGRICULTURAL SCENES

PHOTOGRAPHS, (TOP) J. ALLAN CASH, (CENTRE LEFT) FRED BOND—PUBLIX, (BOTTOM LEFT) GEORGE AUSTIN FROM SHOSTAL, (BOTTOM RIGHT) JOE BARNELL FROM SHOSTAL

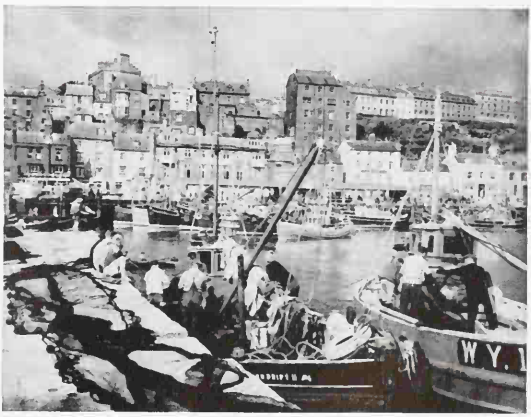
The chalk cliffs of Dover surmounted by Dover castle, an ancient coastal fortress

Southampton docks, seen from the deck of R.M.S. "Queen Elizabeth"

Clovelly, a picturesque fishing village situated in a narrow wooded coomb that descends abruptly to the sea at Barnstaple bay, Devon

ENGLISH PORTS AND COASTAL SCENES

Wharfside scene on the tidal river Thames, Port of London, near Tower bridge



Harbour at Whitby, a market town and seaside resort in the North Riding of Yorkshire

PHOTOGRAPHS, (TOP LEFT) © 1962 BY CHARLES E. ROTKIN FROM "EUROPE: AN AERIAL CLOSE-UP," LIPPINCOTT, (TOP RIGHT) J. ALLAN CASH—RAPHO GUILLUMETTE, (CENTRE LEFT) JOSEPH MUENCH—PICTORIAL PHOTOGRAPHY, (BOTTOM LEFT) F.I.P. BY BRUCE MAYNE, (BOTTOM RIGHT) NOEL HABGOOD F.R.P.S.

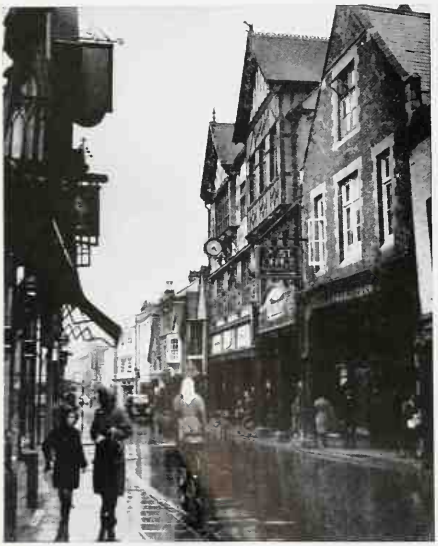
ENGLAND

Typical narrow street, a survival of medieval times, in York. The towers of York minster are in the background

An aerial view of London along the Thames. Prominent landmarks include Tower bridge (centre) and St. Paul's cathedral (right background)

ENGLISH TOWNS AND CITIES

Open market in the centre of Harlow New Town, Essex



The High street in Winchester, Hampshire

The civic centre (opened 1939) of Birmingham, Warwickshire

PHOTOGRAPHS, (TOP LEFT) NOEL HABGOOD (COLEMAN & HAYWARD); (TOP RIGHT) © 1962 BY CHARLES E. ROTKIN FROM "EUROPE, AN AERIAL CLOSE-UP," LIPPINCOTT, (CENTRE RIGHT) W. SUSCHITZKY, (BOTTOM LEFT) THOMAS HOLLYMAN—PHOTO RESEARCHERS, INC., (LOWER RIGHT) CAMERA PRESS—PIX FROM PUBLIX

Pottery works at Longton in Staffordshire

Production line at an aircraft factory in Chester, Cheshire

Power station at Coleshill, Birmingham, Warwickshire

Modern electric furnace for producing high-grade carbon and alloy steel at the Tinsley Park plant in Sheffield

INDUSTRY AND POWER

Miners coming off shift at Manvers Main colliery near Rotherham, Yorkshire

PHOTOGRAPHS, (TOP LEFT) AERO PICTORIAL LTD., (TOP RIGHT) REUTER PHOTOS LTD., (CENTRE LEFT) CENTRAL PRESS PHOTOS LTD., (BOTTOM LEFT) CAMERA PRESS—PIX FROM PUBLIX (BOTTOM RIGHT) NATIONAL COAL BOARD

ENGLAND

The beach and Tower at Blackpool, a popular Lancashire resort on the Irish sea

Watching a cricket match on the village green at Bearsted, Kent

Browsing at an open-air bookstall in Farringdon road, London

LEISURE ACTIVITIES

Gold Cup day during Ascot week, the race meeting held in June on Ascot heath, Berkshire. Queen Elizabeth II and Prince Philip are driving down the course

Weekend cyclists in west Cumberland. Cycling is a popular English recreation

PHOTOGRAPHS, (TOP LEFT) ADOLPH MORATH, (TOP RIGHT) THOMAS HOLLYMAN—PHOTO RESEARCHERS, INC., (CENTRE RIGHT) SUSAN MCCARTNEY—PHOTO RESEARCHERS, INC., (BOTTOM LEFT) CENTRAL PRESS FROM COMBINE, (BOTTOM RIGHT) "ILLUSTRATED" FROM BLACK STAR

universal schooling may be doing to alter this. (H. J. F.)

3. Religion.—*Church of England.*—England is officially a Christian country, with an established church (*see* ENGLAND, CHURCH OF). The country is divided into the provinces of York and Canterbury, each under an archbishop. The provinces are composed of dioceses, each under a bishop, assisted in most cases by one or more suffragan bishops assigned to subsidiary territorial sees, or by one or more assistant bishops not having such sees. The dioceses are subdivided into parishes (the smallest ecclesiastical units). The parishes are grouped for certain administrative purposes into rural deaneries. The cathedral churches are governed by chapters, consisting of a dean (or, in most modern foundations, a provost), canons and (sometimes) additional prebendaries (*see* CATHEDRAL). The present civil parishes do not coincide with ecclesiastical parishes, which have lost their old local governmental significance. Each parish has its church, with its priest who is known as a vicar, rector or perpetual curate, according to his relation to the temporalities of his parish. All parishioners in England, whether Anglicans or not, are entitled to the ministrations of the parish priest. The two archbishops and the bishops of London, Durham and Winchester sit in the house of lords. There are 21 other episcopal seats which are filled from among the remaining 38 diocesan bishops (except the bishop of Sodor and Man), who take their seats as vacancies occur in order of seniority of consecration. The nominal membership of the Church of England is about 27,000,000 (the number of persons baptized into it), but it is estimated that the active adult membership is less than 3,000,000.

Roman Catholic Church.—The modern Roman Catholic hierarchy was established in England in 1850. Although under the Ecclesiastical Titles act, 1851, Roman Catholic bishops were prohibited from assuming territorial titles, the act remained a dead letter. However, most of the titles chosen did not duplicate the names of English dioceses then existing. The Roman Catholic Church in England is under the primacy of the archbishop of Westminster.

There are three provinces, Westminster, Liverpool and Birmingham, each under an archbishop-metropolitan (Westminster being subject directly to the primate), and 16 dioceses, including the three archdioceses subject to the provincial metropolitans. The archdiocese of Cardiff contains the English counties of Monmouthshire and Herefordshire. The Roman Catholic population of England, less these two counties, was estimated in the early 1960s to be more than 3,500,000.

Other Christians.—Many English Christians belong to one or other of the evangelical bodies that were formed either by separation from the established church or more or less under the influence of continental Protestantism. These are termed Nonconformist in considering their relationship with the Church of England, but themselves prefer to be called collectively the Free Churches, in the sense that they are free from the trammels of establishment. The following Free Churches are linked by the Free Church Federal council, formed in 1940: Baptist, Congregational, Methodist, Presbyterian, Moravian, Independent Methodist, Wesleyan Reform union and Countess of Huntingdon's Connexion.

The Methodist Church, as opposed to the smaller Independent Methodist and Wesleyan Reform bodies, comprises the three branches of Methodism (Wesleyan, Primitive and United) that combined in 1932 as the United Conference of Methodist Churches. The Presbyterian Church of England (about 70,000 members) has the same theology and form of government as the established Church of Scotland but is not legally connected with it. Most of the chapels of the Countess of Huntingdon's Connexion are in practice affiliated to the Congregational union. The various Free Churches generally have a resident minister (pastor) where the churches have large congregations. The small congregations are visited periodically by ministers but otherwise have lay preachers. The Independent Methodists and the Wesleyan Reform union particularly rely on lay preachers. The Congregational and Baptist Churches have a less elaborate central organization than the Methodist and Presbyterian and rely more on the principle of self-government of each congregation.

Estimates available show that in the early 1960s there were in England about 730,000 Methodists, 218,000 Congregationalists (including English in Wales), 248,000 Baptists, and 2,800 Moravians. The most notable Christian bodies in England other than those described above are the Brethren (Plymouth Brethren; 80,000 members), Churches of Christ, Swedenborgians (New Church or New Jerusalem Church; 4,300), Friends (Quakers; 21,000, including Welsh members), Christian Scientists (about 340 churches in the British Isles), Unitarians (about 19,500), Salvation Army (more than 5,500 officers and cadets in England and Wales), Seventh-Day Adventists (6,800 in the British Isles) and Jehovah's Witnesses (33,000 in the U.K.). The Society of Friends occupies a special position, as fixed dogmas and (as with the Brethren) an ordained ministry are against the principles of the society; though a small group numerically, the influence of the society upon social and religious thought is great. The Salvation Army, with a more or less military organization, is famed for its work among the less fortunate strata of the community. (*See* also the separate articles on the religious bodies named above.)

Jews.—The Orthodox Jews of England are in two communities: the large Ashkenazi composed generally speaking of Jews of German and eastern European origin; and the small Sephardi, consisting mainly of Portuguese and Spanish Jews. The difference between the communities is rather of ritual use than, as between Christian denominations, of doctrine. The Ashkenazim comprise three large federations of synagogues and a number of independent congregations. The distinction between the various federations and congregations is here historical, administrative and financial, and in no way one of doctrine or practice. The oldest, largest and most influential of the Ashkenazi federations is the United Synagogue, founded in 1870 by the union of four congregations with the Great Synagogue in Aldgate, London (established *c.* 1690). The first Sephardic community was established in the City of London in 1657 and in 1701 it built the famous Bevis Marks synagogue, the oldest synagogue in England still in use.

Besides the Orthodox there are smaller communities of Reform Jews (originating in 1840 with the West London Synagogue of British Jews) and Liberal Jews (who began to organize themselves early in the 20th century).

The head of English Ashkenazi Jewry is the chief rabbi of the Commonwealth of Nations, whose office arose out of the rabbinate of the Great Synagogue. His jurisdiction strictly extends only to the "united congregations" that contribute the chief rabbi's fund; in practice, however, other congregations acknowledge him as leader and representative of the Jewish community. He does not exercise any governance over the internal religious and administrative affairs of the Sephardim, who have their own chief rabbi (*haham*), board of elders and laws. The principal legislative body of English Jewry is the chief rabbi's court (*beth din*), and the main representative body is the board of deputies. It is estimated that there are more than 380,000 Jews in the towns and cities of England. (X.)

IV. ARCHAEOLOGY

England's earliest inhabitants were primitive hunters like their continental neighbours. The Paleolithic (Old Stone Age) began before 500,000 B.C. Its earliest (Lower) phases are represented by implements occurring (mainly in the south) in fluvio-glacial deposits of the Pleistocene Ice Age. Swanscombe man (*see* MAN, EVOLUTION OF) was found in gravels which contained Acheulian implements, so named from St. Acheul near Amiens, the French type site. In the Middle and Upper phases the hunting economy continued but caves were increasingly occupied in limestone areas (Creswell Crags, Cheddar, Torquay, etc.) in response to colder conditions toward the end of the Ice Age. English cave cultures at this time were relatively backward as compared with France and Spain. No cave painting is known and there appears to have been a breakdown in communications which led to the divergent development of a native flint industry known as the Creswellian.

The Mesolithic people were hunters and food gatherers like the Upper Paleolithic, from whom some of them were directly de-

scended. Others were newcomers. From the north came communities of forest dwellers, equipped with axes: they were widely distributed in southeastern and eastern England (Horsham in Sussex; Starr Carr in Yorkshire) and related particularly to the Maglemose culture of Denmark. Others settled on open sandy or upland sites which remained "islands" in the forest. They were widely distributed, of southern origin, and belonged mainly to the Tardenoisian or Sauveterrian cultures of France.

Agriculture, the essential element of the Neolithic period, was introduced *c.* 3000 B.C. along two main routes. From continental Europe (east and south) primary colonizers (Neolithic A or Windmill Hill people) were widely dispersed in the southeast and were in line with the spread of Neolithic civilization by way of the Danube, the Rhineland and the Mediterranean. Their western counterparts (Western Neolithic) were the collective-tomb (Megalith) builders, composed of a number of groups but all associated with the same long-headed ("Mediterranean") physical type. The descendants of the Mesolithic population developed their own cultures (secondary Neolithic or Neolithic B) as a result of contact with the newcomers. They were mainly eastern in distribution, the chief culture being known as the Peterborough culture, and retained some part of their original food-gathering and fishing economy.

The knowledge and use of metal (copper in the first place) began to appear about, or a little after, 2000 B.C. Associated in some way with this were the broad-headed Beaker people, who arrived in a succession of groups mainly from the low countries and northern Europe during the following centuries. The Early Bronze Age developed against a background of trade and prospecting involving not only metals (copper and gold from Ireland and tin from Cornwall) but also amber (from the Baltic), jet (Yorkshire) and, later, faïence (near east). The Wessex culture dominated southern England during this time (*c.* 1600–1400 B.C.) and was responsible for Stonehenge (*q.v.*) in its most advanced form. The Middle Bronze Age (*c.* 1400–900 B.C.) was by contrast a time of relatively peaceful internal development. The Deverel-Rimbury culture appeared in the south *c.* 1200 B.C. and in the Late Bronze Age (*c.* 900–500 B.C.) there was a renewal of movement, both of trade and settlement, from the continent which was continued into the Iron Age.

The Iron Age cultural groups of Britain are generally labeled A, B and C from the order in which they arrived, but in fact they overlap each other in time. Although they came mainly from the direction of Gaul, other influences, particularly from the southwest, were also present. The great hill forts of Wessex and elsewhere belong to this period. The A people belonged to the earliest Iron Age culture of the continent, called Hallstatt (*q.v.*), and were widely distributed throughout southern England. They were Celts, like the B people, and their later groups already possessed elements of the La Tène art which was a feature of the B cultures, first appearing in the east and south about 300 B.C. The C peoples were the Belgae, of mixed Celtic and Teutonic stock, first appearing in Kent and Essex in the early 1st century B.C. The expansion of the Belgae was arrested by the Roman invasion of A.D. 43. (*See* also ARCHAEOLOGY; EUROPE: *Archaeology.*) (WI. F. G.)

Within a few decades of the invasion by the emperor Claudius in A.D. 43, Romanization in the lowland zone of Britain had made good progress. By the turn of the century there were young natives who prided themselves on their Latin. Roman goods had reached Britain before that, even prior to Caesar's first invasion (55 B.C.). It should be emphasized that Britain was never settled by large numbers of Romans or Italians, who were represented probably only by senior officers and officials; the soldiery might have come from anywhere between Spain and Mesopotamia. The main body of the population was still "British," though with growing admixture. Latin was the official language but probably at home Celtic or other languages were spoken. A tile found in London (now at the Guildhall museum) had scratched on it when wet "Austalis has been wandering off by himself for the last thirteen days." Written probably by an outraged fellow workman, this graffito is in passable Latin and indicates a reasonable standard of literacy in Latin. Many an excavated item of the period of Roman occupation was made by Britons according to native traditions and the term "Romano-British" is therefore aptly applied to them. (For a broader definition of the term "Romano-British" *see* BRITAIN.)

The economy hinged on agriculture. Roman "villas," many of them still unexcavated, were agricultural units ranging from small farmhouses to the mansions of magnates with huge properties. The great colters found, for example, at Silchester indicate obviously enough working of the land, but much has still to be learned about day-to-day organization of the farm.

Mining, whether by contractors or by the military, is recalled by finds of pigs of lead (by-products of silver extraction) deriving from the Mendip hills, the Peak, Yorkshire and elsewhere. Copper and even gold were mined in Wales. Iron mining and smelting went on in many parts of England while some tinworking continued in Cornwall. Much archaeological evidence is buried or lost: this can be seen, for instance, from a literary reference to waterproof cloaks that were exported of which no example has been discovered. Wine, certain foodstuffs and luxury goods were imported: excavators on Roman sites find thousands of pieces of the red-glazed Samian or *terra sigillata* ware. This ornate crockery, much of it imported from Gaul, is more prized for its dating value than for its aesthetic merits. A boatload of it, sunk off Kent, still occasionally puts into nets a shell-encrusted Samian dish.

Many of the aspects of daily life revealed by excavation are general knowledge: the public and private bathhouses, the elaborate ground plans of villas, mosaic pavements, oysters, of which probably millions of shells have been thrown on the spoil heaps of excavations. It is perhaps better to choose at random some less known facets of life in Britain: the occasional use of coal fires; the existence of glass window panes; table legs of shale; the many ways of disposing of the dead whether in glass jars or in lead canisters, or in barrows or in linen shrouds and plaster. All this indicates an elaborate pattern of existence which of course varied with nearly four centuries of economic and military vicissitudes and, probably, with the locality.

Few of the theatres of the towns have been excavated as was the one at Verulamium (close by St. Albans). The spade and trowel have disclosed much about the private life of Romano-Britons. An example would be dicing and gaming equipment which show how Rome's northernmost citizens and soldiers whiled their time away. Religions there were in great variety: besides the official "emperor worship" there was a continuation of local cults together with the importation of exotics—a vessel found in south London has "London, by the temple of Isis" scratched on it. There is excavated evidence in several parts of Britain for Mithraic worship, while finds and literary references hint at the presence of Christians in Britain, and even heresy, long before Augustine's mission.

Many local museums in England, especially on or near Roman sites, display or store Roman remains. Examples are the Verulamium museum at St. Albans, the London museum and the Guildhall museum in London, the museums at York, Bath, Reading, Chester, Colchester, Corbridge and elsewhere. *See* also BRITAIN; LONDON. (X.)

V. POPULATION

1. Early Population Patterns.—A sequence of clearly defined cultures is apparent in Britain stretching back in time for at least 2,000 years before the Roman occupation. The chief limiting factors to prehistoric settlement were the widespread existence of heavy forest and the determining soil conditions and climatic changes. As population grew, aided by successive Celtic migrations, man's increasing control over environment enabled him to utilize not only land which was not densely wooded but to begin the attack on "damp" oakwood forest.

The greater part of the plain of southeastern England was occupied by the Romans in their first four years of conquest. The post-Roman centuries saw an extension of lowland rural settlements above water meadows and in woodland clearings. Municipal life was a marked feature of Roman civilization and London replaced Colchester as the leading mercantile centre of the country.

After the Anglo-Saxon conquest great stress must be put on the

importance of the river systems as a key to determining the distribution of settlement. The Anglo-Saxons seldom cultivated the lighter soils, which had formed the great extent of arable land up to the end of the Roman period. Instead, they settled by watercourses cultivating nearby land and controlling only gradually the upland "waste." They preferred deep meadow to hill pasture and worked to bring order into the valley bottoms.

During the 9th and 10th centuries Scandinavian settlers seized not only East Anglia but land to the south in Essex, Hertfordshire, Bedfordshire and perhaps Buckinghamshire. Danes also held land in Yorkshire, Lincolnshire, Leicestershire, Nottinghamshire, Northamptonshire and Derbyshire. The western shores of England and Wales were also subject to Scandinavian influence where there was widespread Norwegian settlement.

Medieval Period.—The Domesday Book provided an important but very inadequate basis for population estimate during the Norman period, because whole classes were omitted from its returns or dealt with summarily. However, F. W. Maitland (*Domesday Book and Beyond*, 1961) estimates the population in 1086 as 1,375,000 and J. C. Russell (*British Medieval Population*, 1948) estimates it as just over 1,100,000. A characteristic of the 11th century is the existence of a great amount of devastated land particularly to the north, much natural wilderness and of land converted to "forest law" to preserve the king's hunting. Therefore, much of Norman England was uncultivated compared with the England of later centuries. Nevertheless, despite the famine and anarchy of Stephen's reign, colonization of waste continued. It has been estimated that at the time of the Domesday survey about 5,000,000 ac. were arable which is half as much as the land returned as arable in the 1879 census. The period from Domesday Book until the 14th century saw the establishment of an English nation through a blending of Norman and Anglo-Saxon groups. By the end of the 12th century it was becoming difficult to distinguish the two groups in the free population. Population during the 12th century ranged at perhaps around the 2,000,000 mark according to A. L. Poole (*Domesday to Magna Carta*, 1955). The vast majority were agriculturalists living in villages.

Town and community life which lapsed after the Roman evacuation of England revived with the coming of the Scandinavians and a strong group of Danish boroughs was formed in the Midlands. The Domesday supplies over 100 names of places containing burgesses. It must be noted, however, that of 43 markets recorded in the Domesday survey, only 11 were held in boroughs. Throughout the middle ages as before the Norman Conquest all facts attest the supremacy of London both in size and importance over other English towns.

Estimates of population in England before the Black Death (1348–49) vary from 2,500,000 to 4,000,000. The visitation of plague killed, however, perhaps a half, perhaps a third of this increasing population and undermined the manor organization as an agricultural system.

Figures based on poll tax returns of 1377 indicate a sparsely populated north and west of England and a more densely populated south, Midlands and east. Density was rather low in Surrey and Sussex because of the Weald and in Hampshire because of the New forest. The following estimates of population are given for the principal towns: London, 35,000; York (the next largest city), 11,000; Bristol, 9,500; Coventry and Plymouth, 7,000; and Norwich, 6,000. There was already some drift toward London and a few other large towns but it is unlikely that the balance of rural population altered greatly between 1307 and 1377. Nevertheless by the end of the 14th century land was going out of cultivation and certain villages were abandoned both as a result of loss of population and because of the increase in sheep farming.

15th–18th Centuries.—During the 15th century the civil wars led to widespread social disorder enhanced by the continuing shortage of labour in the first part of the century. However, land enclosure continued and population redressed the balance by the end of the century.

John Leland and William Camden have left accounts of England during the Tudor period. In the 16th century population was increasing slowly and was still largely rural. By the end of this period a fair number of these rural inhabitants were engaged in some form of industry, either supplying local needs as carpenters, shoemakers and other tradesmen or working for the general market as did miners, clothiers and quarrymen.

During the 15th century Southampton and Sandwich began to grow in importance as ports and new geographical discoveries stimulated trade in the west and south. This movement continued and in early Tudor England certain of the east coast ports declined—notably Boston, Lynn, Lincoln and Stamford.

Enclosure and the upward pressure of population levels created a great problem of poverty in Tudor England. A class of vagrants and rootless labourers was created and town and industrial life was not sufficiently developed to absorb all. This great social problem led to the first Poor Law acts.

The unique and continuing growth of London stimulated agriculture in the home counties (counties over and into which London has expanded) and certain important people already passed part of the year outside London in these localities.

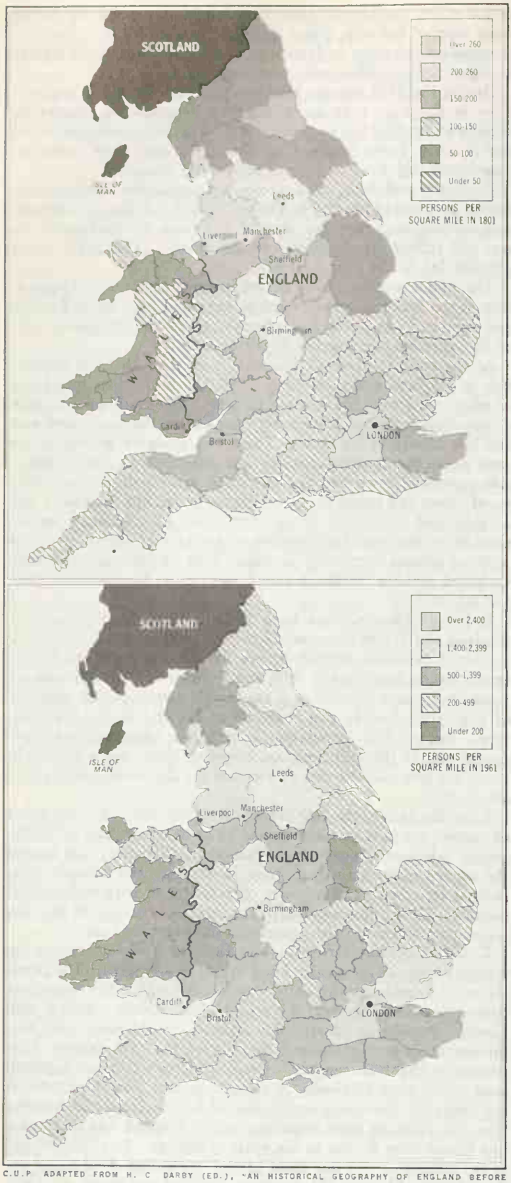
At the end of Elizabeth I's reign it is likely that the population of England was more than 4,000,000, more than four-fifths of whom led a rural life in small parishes. The traditional balance of population remained with high density in the south and east and low density in the north. There was, however, growth in the west where the clothing industry was concentrated. An average-size provincial town had perhaps 5,000 inhabitants. The silting up of rivers and estuaries, the recession of sea, the increasing size of ships and migrations of the wool trade all influenced the decline of certain traditional centres of population. However, towns were in general increasing in size. York, Bristol and Norwich each had perhaps 20,000 in population, while London may have had 200,000.

It is held by some that from 1630 to 1750 the population of England and Wales increased from 5,600,000 to 6,517,000. Estimation is, however, difficult before the publication of the official census figures beginning in 1801, and *Britain: an Official Handbook* (annual) gives 5,500,000 as a reasonable contemporary estimate of population for the two countries at the end of the 17th century. Before 1700 population growth was comparatively slow. At the end of the 18th century, however, there was a decline in the death rate and the rate of increase for the whole century was about 50%.

From the 16th to the 18th century England received a new influx of people, not in large numbers but of great energy and initiative, mainly religious refugees from France and Flanders, who caused the woolen manufactures to grow enormously and brought commercial experience and a power of criticism in both religion and politics. Their influence and that of their descendants on the subsequent growth of British towns and industries was great.

2. 19th-Century Growth.—By 1801 the first census gives the population figure for England and Wales as 9,178,980. The population of London at this time was 864,000 and other important towns were Manchester, Liverpool, Birmingham, Bristol and Leeds. Population distribution shifted considerably and this movement intensified toward the end of the 18th century continuing in the 19th century. Density increased in the highland zone of the west Midlands, in Lancashire and Yorkshire through the migration into these areas affected by the Industrial Revolution. In 1700 the most populous part of England was south of a line drawn from Bristol to the coast of Suffolk. In 1750 districts to the northwest of a line from Bristol to Durham were filling up. At the time of George III, however, England and Wales were still predominantly agricultural. Only two cities had a population of more than 50,000: London had perhaps 750,000 inhabitants and Bristol 60,000.

In the years 1801 to 1811 England showed a population increase of 14% and Wales an increase of 12%. There was in this period a new phenomenon of urban concentrations of unprecedented size fed both by natural increase and rural depopulation. Throughout the whole of the 19th century the rapid population growth was maintained by the high birth rate and decreasing death rate. Urbanization continued particularly during the first half of the 19th century. By 1851 half the total population of England and Wales



C.U.P. ADAPTED FROM H. C. DARBY (ED.), "AN HISTORICAL GEOGRAPHY OF ENGLAND BEFORE A.D. 1800"

(TOP) DENSITY OF POPULATION IN ENGLAND AND WALES, 1801; (BOTTOM) DENSITY OF POPULATION IN ENGLAND AND WALES IN 1961

was living in urban areas. Agriculture was pushed from its central position in the economy. No decade in the 19th century showed less than a 10% increase of total population in comparison with the previous decade. The population doubled in the first 50 years of the period and doubled again by 1911. Only toward the end of the century did a tendency toward family limitation become apparent.

3. 20th Century.—During this period the over-all increase has been maintained at a slower rate. In the 60 years from 1871 England and Wales lost on balance well over 2,000,000 people through external migration. However, there was a net gain from migration after 1931 which to some extent offset the decline in natural increase.

Between World Wars I and II there was a drift through economic reasons from the north, northeast and south Wales to the south and Midlands, and certain coal-producing centres and other traditional areas of heavy industry have declined. This trend is continuing. The magnet attraction of the more southern areas has been enhanced by regional unemployment. Thus, the general picture in the 1960s is of a continued movement of industry to the south with a population drift particularly of young people from the coal-mining and textile areas of the north, and with an accompanying growth of population in the Midlands where iron, steel, brickmaking and chemical industries are thriving. There is increased industrial expansion in southern areas with good communications with London, or round south coast ports. There is a continued movement to dormitory areas from the centres of conurbations, particularly round Greater London, as transport is improved by railway electrification schemes, and an aged population is filling up what have become accepted as the south coast retirement areas. It has been said that Britain is a nation of townspeople and since 1921 nearly 40% of the population of England and Wales has lived in six great conurbations in the following areas: around London, the capital city which lies in the lower Thames basin; Manchester and surrounding areas of southeast Lancashire, Yorkshire and Cheshire; Birmingham, Wolverhampton and surrounding areas of the west Midlands; Liverpool and part of Lancashire and the Wirral peninsula; Leeds and Bradford in west Yorkshire; and Newcastle-upon-Tyne, the natural focus of Tyneside and surrounding Durham and Northumberland.

From about the 1780s the pace of rural migration quickened. By the end of the 19th century rural depopulation was considered a serious problem. This process has continued without abatement in the 20th century and it appears likely that it will continue because of the lack of opportunity for rural employment. Women in particular tend to leave the agricultural areas to seek work in the large towns and especially in London and the southeast. Therefore there is generally a low female–male ratio in rural districts and a high female–male ratio in the urban areas. There were also great movements of population into the industrial areas from Ireland, the Clyde and Wales.

The urban population first exceeded the rural population at the 1851 census. By the end of the 19th century the population of England and Wales was rather more than three-quarters urban. Up to 1939 the urban population slowly increased until four-fifths of the population was included within the boundaries of urban authorities. It would appear that a stable situation was approached in 1939 and since 1951 there has been relatively little change in the rural proportion of the population. Internal migration now appears to consist more of movement from urban area to urban area than a large drift from country to town.

Urbanization has changed during the 20th century with growth tending to take place at the edges of conurbations and in suburbs rather than within the large cities themselves. Since World War II, although the conurbations themselves have in general continued to grow, certain of the large towns contained in them (those of over 250,000 inhabitants) have tended to decline in population while towns of 50,000 and 100,000 inhabitants have grown rapidly, and the population of Greater London has declined.

According to the preliminary census report of 1961 the Greater London conurbation covering 461,882 ac. with a population of 8,171,902 persons was comprised of the administrative county of London covering 74,898 ac. with a population of 3,195,114; the City of London covering 677 ac. with a population of 4,771; and an "Outer Ring" of 386,307 ac. with a population of 4,976,788. This outer ring was composed of suburbs in Middlesex, Essex, Kent, Surrey and Hertfordshire. For the intercensal period 1951–61 there was a 4.6% decrease in population in the administrative county of London and a 10% decrease in the City of London. For the Greater London conurbation as a whole throughout this period

there was a 2.1% decrease. The population of the outer ring, which increased during the 1950s, remained almost stationary in the early 1960s after the area became fully developed. The surrounding southeast England area from 1951–61, however, experienced a 14.2% population increase.

The Local Government act of 1958 recommended that towns of 100,000 inhabitants should be able to support county borough status. Many county boroughs have a smaller population than this and many larger towns are in fact not county boroughs. In 1961 of the 28 metropolitan boroughs in the administrative county of London 13 had populations of more than 100,000. County boroughs with populations under 100,000 include Canterbury with 30,376 (1961), and Great Yarmouth and Burton upon Trent with populations of 52,860 and 50,766 respectively. County boroughs with a population of more than 100,000 are shown in Table III as are the intercensal changes in density. From this table some indication of population shift can be gathered.

After 1801 censuses of population were conducted decennially apart from 1941 when World War II prevented enumeration. According to the general register office *Census 1961, England and Wales: Preliminary Report* the total population for England and Wales enumerated on April 23, 1961, was 46,071,604 of which 22,298,899 were males and 23,772,705 were females.

The 1951 census contained the largest program of inquiries ever included in a United Kingdom census. The 1961 census was somewhat larger in scope; however, because of the introduction of sampling methods at the enumeration stage, nine out of ten people were asked for less information than in any other census of the 20th century.

The 1961 census gives a density for England and Wales of 790 persons per square mile of land and states that only one other major country in Europe, the Netherlands, had a higher density (879 persons per square mile in 1957), although Belgium approaches it (771 persons per square mile in 1958). There was a net gain by migration from other countries of 352,000 in the intercensal period 1951–61.

The register general's *Statistical Review of England and Wales 1959* estimated the total home population for the year as 45,386,000, at June 30, 1959.

Changes in the rate of population growth in Great Britain since 1801 were intensively investigated by a royal commission (Cmd. 7695, June 1949). The rate of increase has slowed down in a most marked manner, although the *Census 1961, Preliminary Report* shows a 2,314,000 increase in England and Wales since 1951. This is a greater absolute increase than at any intercensal period since 1901–11. (The considerable net emigration from England and Wales before 1931 was largely composed of young adults and adolescents and the reversal of this flow after that date was caused by political refugees and the effects of world-wide economic depression.)

The reduction in rate of growth occurred through a decline in the birth rate after 1877 because of an increasing use of contraceptive measures. Nevertheless relatively high birth rates continued into the beginning of the 20th century. This trend to smaller family units has had a profound effect on conditions and social structure. In England and Wales the average rate of increase per annum of total population from 1831 to 1851 was 1.48% but from 1931 to 1951 this had dropped to 0.47%. Throughout the 19th century the average rate had persisted at well over 1%.

At the end of World War II and in the immediate postwar years there was a considerable rise in birth rates reaching a maximum in 1947 when the crude live birth rate per 1,000 was 21.1 (civilian population). This dropped to 15.7 (home population in 1956) and thereafter rose as follows: 16.1 in 1957, 16.4 in 1958 and 16.5 in 1959. Live births in 1960–61 were 18% above the level of 1951–52. After the war there were high marriage rates in Eng-

TABLE I.—*Counties of England*

County	Administrative headquarters	*Administrative area (sq.mi.)	Population		Population density (persons per square mile)	
			1951	1961	1951	1961
Bedfordshire	Bedford	473.3	311,937	380,704	660	805
Berkshire	Reading	710.5	288,945	383,487	407	540
Buckinghamshire	Aylesbury	749.1	386,291	486,183	516	650
Cambridgeshire	Cambridge	492.4	166,887	189,913	339	385
Cheshire	Chester	973.1	815,884	921,212	840	945
Cornwall (excluding the Isles of Scilly)	Truro	1,350.2	345,442	341,746	256	253
Cumberland	Carlisle	1,511.0	217,540	223,050	144	148
Derbyshire	Matlock	992.9	685,170	745,223	690	750
Devon	Exeter	2,577.2	514,213	538,412	200	209
Dorset	Dorchester	973.1	291,123	309,176	300	318
Durham	Durham	970.1	903,080	953,410	930	980
Essex	Chelmsford	1,499.2	1,601,329	1,859,449	1,067	1,240
Gloucestershire	Gloucester	1,208.3	429,149	494,366	354	407
Hampshire	Winchester	1,453.0	628,959	762,214	431	524
Herefordshire	Hereford	842.1	127,159	130,919	151	155
Hertfordshire	Hertford	632.1	609,775	832,088	965	1,312
Huntingdonshire	Huntingdon	365.6	69,302	79,879	189	218
Isle of Ely	March	374.9	89,049	89,112	236	236
Isle of Wight	Newport	147.1	95,625	95,479	649	648
Kent	Maidstone	1,517.6	1,536,507	1,670,707	1,012	1,102
Lancashire	Preston	1,619.4	2,038,875	2,201,815	1,281	1,360
Leicestershire	Leicester	805.3	345,896	408,898	430	507
Lincolnshire						
Parts of Holland	Boston	418.5	101,555	103,388	242	246
Parts of Kesteven	Sleaford	722.0	129,785	135,317	179	187
Parts of Lindsey	Lincoln	1,501.6	309,592	330,948	206	220
London	London	117.0	3,347,956	3,195,114	28,600	27,300
Middlesex	Middlesex Guildhall, London	232.3	2,269,315	2,230,093	9,800	9,650
Norfolk	Norwich	2,035.2	375,721	389,216	185	191
Northamptonshire	Northampton	904.6	255,258	292,771	282	324
Northumberland	Newcastle upon Tyne	1,994.0	440,136	479,487	220	240
Nottinghamshire	Nottingham	818.5	533,361	591,321	652	724
Oxfordshire	Oxford	735.6	177,061	203,334	241	278
Peterborough, Soke of	Peterborough	83.5	63,791	74,442	768	895
Rutland	Oakham	152.0	20,537	23,956	136	158
Shropshire	Shrewsbury	1,346.6	289,802	297,313	215	220
Somerset	Taunton	1,603.0	472,159	517,000	295	323
Staffordshire	Stafford	1,071.0	853,220	985,113	797	922
Suffolk, East	Ipswich	855.3	321,909	342,696	376	400
Suffolk, West	Bury St. Edmunds	610.8	120,652	129,969	196	213
Surrey	Kingston-upon-Thames	701.8	1,352,639	1,480,649	1,928	2,110
Sussex, East	Lewes	772.8	333,136	374,669	431	485
Sussex, West	Chichester	633.4	322,792	411,224	508	650
Warwickshire	Warwick	873.0	490,743	612,578	562	701
Westmorland	Kendal	788.9	67,392	67,222	85	85
Wiltshire	Trowbridge	1,344.7	386,692	422,753	288	314
Worcestershire	Worcester	684.7	398,251	441,029	582	616
Yorkshire						
East Riding	Beverley	1,149.9	211,799	223,783	184	195
North Riding	Northallerton	2,116.6	378,209	397,074	179	188
West Riding	Wakefield	2,510.8	1,588,413	1,649,135	636	657

*This figure does not include the area of any county borough that is within the geographical county because a county borough is an independent unit.

TABLE II.—*Changes in Population by Births, Deaths and Migration, 1871–1961 (England and Wales)*

(in 000s)

Period	Increase in population	Births	Deaths	Natural increase	Net migration
1871–81	3,262	8,604	5,178	3,426	−164
1881–91	3,028	8,894	5,265	3,629	−601
1891–1901	3,525	9,157	5,563	3,594	− 69
1901–11	3,343	9,290	5,246	4,044	−501
1911–21	1,816	8,281	5,845*	2,436	−620
1921–31	2,066	6,928	4,692	2,236	−170
1931–39 (June 30)	1,508	4,969	3,985	984	524
1939–51	2,298	8,330	6,263*	2,067	231
1951–61	2,314	7,121	5,159	1,962	352

*Including losses of 577,000 and 240,000 of noncivilians and merchant seamen belonging to England and Wales who died outside the country during 1914–18 and 1939–48 respectively.

TABLE III.—*Towns With Populations Over 100,000 in 1961 (England)**

Area	Acreage	1961 census (persons)	Intercensal increase, 1951–61	
			Number	Per cent
Birkenhead C.B.	8,618	141,683	− 818	− 0.6
Birmingham C.B.	51,147	1,105,651	− 7,034	− 0.6
Blackburn C.B.	8,088	106,114	− 5,104	− 4.6
Blackpool C.B.	8,609	152,133	4,801	3.3
Bolton C.B.	15,280	160,887	− 6,280	− 3.8
Bournemouth C.B.	11,448	153,965	9,120	6.3
Bradford C.B.	25,525	295,768	3,365	1.2
Brighton C.B.	14,347	162,757	4,689	3.0
Bristol C.B.	26,350	436,440	− 6,554	− 1.5
Coventry C.B.	19,140	305,060	48,818	18.1
Croydon C.B.	12,672	252,387	2,517	1.0
Dagenham M.B.	6,728	108,363	− 6,205	− 5.4
Derby C.B.	8,116	132,325	− 8,942	− 6.3
Ealing M.B.	8,781	183,151	− 4,535	− 7.6
East Ham C.B.	3,324	105,359	−15,477	−12.8
Enfield M.B.	12,401	109,524	− 941	− 0.9
Gateshead C.B.	4,560	103,232	−11,875	−10.3
Harrow M.B.	12,555	208,963	−10,531	− 4.8
Hendon M.B.	10,371	151,500	− 4,357	− 2.8
Heston and Isleworth M.B.	7,261	102,897	− 3,950	− 3.7
Hornchurch U.D.	20,308	128,127	24,035	23.1
Huddersfield C.B.	14,147	130,302	1,276	1.0
Ilford M.B.	8,411	178,210	− 6,496	− 3.5
Ipswich C.B.	9,957	117,325	9,907	9.2
Kingston upon Hull C.B.	14,421	303,268	4,163	1.4
Leeds C.B.	40,615	510,597	4,717	0.9
Leicester C.B.	16,985	273,298	−11,883	− 4.2
Liverpool C.B.	27,810	747,490	−43,348	− 5.5
London (administrative county)	74,898	3,195,114	−152,842	− 4.6
Luton C.B.	8,773	131,505	21,124	19.1
Manchester C.B.	27,255	661,041	−42,041	− 6.0
Middlesbrough C.B.	7,131	157,308	10,036	6.8
Newcastle upon Tyne C.B.	11,094	269,389	−22,353	− 7.7
Northampton C.B.	6,201	105,361	929	0.9
Norwich C.B.	8,141	119,904	− 1,332	− 1.1
Nottingham C.B.	18,370	311,645	3,795	1.2
Oldham C.B.	6,392	115,426	− 7,792	− 6.3
Oxford C.B.	8,785	106,124	7,377	7.5
Plymouth C.B.	13,140	204,279	− 3,733	− 1.8
Portsmouth C.B.	9,249	215,198	−18,347	− 7.9
Preston C.B.	6,357	113,208	− 8,159	− 6.7
Reading C.B.	9,105	119,870	5,674	5.0
Romford M.B.	9,342	114,579	26,577	30.2
St. Helens C.B.	8,865	108,348	− 4,175	− 3.7
Salford C.B.	5,203	154,963	−23,231	−13.0
Sheffield C.B.	39,586	493,954	−18,896	− 3.7
Southampton C.B.	11,543	204,707	14,886	7.8
Southend-on-Sea C.B.	10,286	164,976	13,170	8.7
South Shields C.B.	4,876	109,533	2,935	2.8
Stockport C.B.	8,440	142,469	668	0.5
Stoke-on-Trent C.B.	21,209	265,506	− 9,609	− 3.5
Sunderland C.B.	8,575	189,629	8,105	4.5
Thurrock U.D.	44,700	114,302	32,194	39.2
Tottenham M.B.	3,013	113,126	−13,803	−10.9
Twickenham M.B.	7,014	100,822	− 4,841	− 4.6
Wallasey C.B.	5,913	103,213	1,844	1.8
Walsall C.B.	8,780	117,836	3,301	2.9
Walthamstow M.B.	4,343	108,788	−12,347	− 1.1
Wembley M.B.	6,300	124,843	− 6,541	− 5.0
West Ham C.B.	4,689	157,186	−13,807	− 8.1
Willesden M.B.	4,634	170,835	− 8,862	− 4.9
Wolverhampton C.B.	9,126	150,385	−12,287	− 7.6
York C.B.	6,933	104,468	− 947	− 0.9

*C.B.=county borough; M.B.=municipal borough; U.D.=urban district.

land and Wales and increasingly early marriage ages, which caused a temporary rise in births. The effect of this, however, may be largely offset by the growing tendency for families to be completed in the early years of marriage and not extended by the increase of married life falling within the reproductive period.

The marked decline in the mortality rates since the 19th century slowed in tempo during the 1920s and there has been relatively little change since. As is to be expected in an aging population, deaths in the 1951–61 intercensal period were higher than in any corresponding period since 1911–21. The annual average death rate per thousand of the population in England and Wales from 1871 to 1875 was 22; from 1881 to 1885 it was 19.4; from 1901 to 1905 it was 16.1; in 1921 it was 12.4; and in 1959 it was 11.6. In general the death rates for males are significantly higher than for females and since the beginning of the 19th century there has always been an excess of females within the population. The reduction in death rates has resulted in a progressively aging population with the cohorts at higher age ranges having an increasing proportion of females within them. The shift of age distribution within the population is creating new social problems and pressures which are likely to intensify as the century continues. The proportion of persons under 15 in England and Wales in 1841 was no less than 36%, whereas 100 years later (1939) it was only 20%. In 1891 the corresponding figure was still as high as 35%. Only 7% of the population of 16,000,000 in 1841 was over 70 years of age. However, by the 1960s the aged were relatively twice as numerous and this notable increase has occurred wholly since 1891.

VI. ADMINISTRATION AND SOCIAL CONDITIONS

Central government administration is discussed under GREAT BRITAIN. As England and Wales are frequently considered together for administrative and statistical purposes, the following subsections sometimes embrace both countries.

1. Local Government.—In England and Wales local government assumed its present basic shape in the 19th century. A system of locally elected councils was formed, to control various services. Since then local authority responsibility has greatly increased. Two local government commissions were established under the Local Government act, 1958, to review the general county and county borough pattern in England and Wales (excluding Greater London) and in special review areas. The act also made changes in the local government financial system. The ministry of housing and local government is the main link between central and local government in England and Wales. For local government, England (excluding London after April 1, 1965; *see* below) and Wales are divided into county boroughs and administrative counties. The latter comprise three types of county district: noncounty boroughs, urban districts and rural districts (the administrative county of London being divided simply into 28 metropolitan boroughs and the City), all of which are subdivided into parishes. Each group is administered by a different type of local authority and responsibilities differ with groupings.

A royal commission on local government in the Greater London area reported in Oct. 1960. On July 31, 1963, the London Government Act 1963 became law. Under its provisions, effective April 1, 1965, a new administrative area called Greater London would absorb the administrative county of London and nearly all Middlesex (the rest going to Surrey and Hertfordshire) and would include parts of Kent, Surrey, Hertfordshire and Essex. The 28 metropolitan boroughs and other boroughs or districts taken into the new Greater London area would be reorganized into 32 units called London boroughs. The City would be unaffected. (*See* LONDON.)

Councils are formed from elected councilors (and in some cases aldermen elected by the councilors) presided over by a chairman. Councilors usually hold office for three years and aldermen for six. In most boroughs the chairman is the mayor; in the City of London and certain boroughs he is the lord mayor.

Functions and Services.—Every local authority must provide such social and environmental services as parliament requires and may also act under permissive legislation or use powers provided by adoptive acts and by local legislation or private bills. County borough councils are all-purpose authorities. County councils and county district councils have certain functions allocated to them, and although there is considerable independence, there is also a certain degree of delegation from county councils to county district councils. Parish councils or meetings have a few functions they may exercise independently. In London those services which must be uniform throughout the area are provided by the London County council (after April 1, 1965, the Greater London council) and those which can be locally administered fall to the metropolitan borough councils (after April 1, 1965, the London borough councils) and in the City, to the Common council.

Council services fall into three main groups: (1) environmental, which aim at preserving and improving amenities in various fields such as public health and sanitation, the control of nuisances, drainage, sewage, street cleaning, refuse collection, street lighting, the provision of parks, public safety, town and country planning, etc.; (2) protective services, which include fire services, civil defense and police; and (3) personal services, which aim at promoting individual welfare and range from health services, housing, education, to the provision of entertainment and leisure time activities.

Local Authority Elections.—Procedure is regulated by rules laid down in the Representation of the People act, 1949. Any person aged 21 years or over who is British or a citizen of the Irish republic and registered as a local government elector may vote at

a local government election. The voting procedure closely follows that for parliamentary elections. Most local authority areas are divided into electoral districts which may be divisions, wards or parish wards. Outside London, county divisions each return one member: the London electoral divisions return three members each. All other local authority areas return three or a multiple of three members for each ward.

Internal Organization.—This remains to a considerable extent the choice of local authorities. Usually questions of policy and principle are decided by the whole council, which establishes committees to carry out detailed administration of the various services. More important committees of larger authorities may appoint subcommittees to aid them in their work. The work of the council and committees is carried out by salaried executives and employees, who fall into three main categories: (1) heads of departments or chief officers; (2) subordinate officers employed in a professional, technical or clerical capacity; and (3) a large number of manual workers. Rates of pay and conditions of service for local authority staff, although in the power of the employing council (except in rare cases where proposed salaries must have ministerial approval) are based on Whitley council recommendations. (*See* INDUSTRIAL RELATIONS.)

Local Government Finance.—Income is derived from government grants payable to county and county borough councils by the ministry of housing and local government, from local rates, loans, trading receipts, rents, fees and other miscellaneous sources. The largest single government grant is the general grant, payable under the Local Government act, 1958. Other government grants are percentage grants, unit grants and assigned revenues.

Local rates are levied at so much per pound sterling on the ratable value of property and are levied and collected by the councils of the county and non-county boroughs and of urban and rural districts. In London they are collected by the borough councils and City corporation. Internal control of finance lies with a finance committee, while an external audit is carried out by district auditors appointed by the ministry of housing and local government, except for certain general accounts of a majority of borough councils outside London. (*See* also LOCAL GOVERNMENT.)

2. Housing and Planning.—Housing problems in England and Wales stem largely from the enormous population increase since the beginning of the 18th century, together with rapid industrial expansion which caused great internal population mobility. Despite a high rate of building in the 1930s, there was a desperate need for new housing programs after World War II. Over-all responsibility for housing lies with the ministry of housing and local government, while local authorities are responsible for housing in their areas.

The central office of information publication *Britain: an Official Handbook, 1964* states that there are more than 15,000,000 houses in England and Wales of which about 3,750,000 are owned by local authorities. Between 1945 and the end of 1962 more than 4,500,000 new flats and houses were built in England and Wales. While local authorities provided most of these, private provision greatly increased after the removal of building restrictions in 1954. The chief agencies which lend money for house purchase are building societies, insurance companies, industrial and providential societies and local authorities.

In returns made by local housing authorities in 1955, about 850,000 or 6.5% of all houses in England and Wales were estimated as unfit for habitation and local authorities demolished or closed thousands of unfit houses. Many local authorities had completed their clearance programs by the early 1960s. The legislation under which improvement grants are available are the Housing (Financial Provisions) act, 1958, and the House Purchase and Housing act, 1959. The former act also consolidated legislation on subsidies for new housing in England and Wales.

Rent control, introduced in 1915, was maintained by successive legislation until 1954 when a policy of progressive abolition began, culminating in the Rent act, 1957. This removed rent control from dwellings with rateable values of more than £40 in London's metropolitan police district, or of more than £30 elsewhere. It also removed control from all new lettings after 1957 and provided the possibility of decontrol of further categories of property by order in council with parliamentary consent. Rent tribunals established under the Furnished Houses (Rent Control) act, 1946, still operate, but the scope of that act was later limited. (*See* also HOUSING: *Great Britain.*)

Town and Country Planning.—The Town and Country Planning act, 1947, provides for a general planning system throughout England and Wales. Financial amendments to this legislation can be found in the Town and Country Planning acts, 1954 and 1959.

Twelve New Towns (*q.v.*) have been established in England and Wales under the provisions of the New Towns act, 1946. Eight of these aim to absorb excess population in the Greater London area; the other four, in Durham, Northamptonshire and Monmouthshire, are to serve the special needs of those areas. There was subsequent legislation for New Towns in 1959. The National Parks and Access to the Countryside act of 1949 created a commission for England and Wales which by June 1960 had created 11 national parks. The Forestry commission also set up national forest parks, and the Nature Conservancy established a number of reserves.

3. Welfare Services.—Religious orders, manor houses and merchant and craft guilds provided early social services. However, the decline of the feudal system and dissolution of the monasteries meant that traditional sources of relief were removed making it imperative to find an alternative system. Various acts were passed in an attempt to deal with the situation, culminating in the Elizabethan Poor Relief act, 1601. This measure introduced the notion of state responsibility for those in need and enforced on local authorities in England and Wales the duty of providing for the sick, destitute and homeless from local rates. Administration of the Poor law continued under this legislation until 1834 when the "new Poor law" was introduced. In the interval between the Elizabethan act and that of the 19th century, there was a great expansion in voluntary charitable activities, especially in the 19th century. However, at the time of the Industrial Revolution, poor relief was the only statutory welfare service in operation, and in a repressive atmosphere the government had no desire to intervene in what was considered as the private sphere of life. Parliament and the public were slowly goaded into action by humanitarians, administrators and writers. There was no overall design in the growth of the early social services, which developed different rationales according to different needs.

During the greater part of the 19th century reformers aimed at establishing control over environmental conditions and securing minimal health requirements, giving the state the role of watchdog. Later the state took a more active role, directly initiating welfare legislation, filling in the gaps left by voluntary effort.

The foundations of the modern system of welfare are the first Old Age Pensions act, 1908 and the National Insurance act, 1911. Between World Wars I and II, the state began to engage in many new activities. For example, this period saw a large scale implementation of the Maternity and Child Welfare act, 1918. Other measures between the wars established state responsibility through local authorities for the blind and physically handicapped and dealt with unemployment and unemployment insurance, factory conditions, widows, orphans, old age pensions and tuberculosis.

During World War II the reconstruction and rationalization of the social services was planned. Temporary measures were passed to deal with pressing needs and some of these were permanently adopted, like the supply of free or cheap milk and vitamin supplements to mothers and young children. The Beveridge report on social insurance and allied services appeared in 1942, forming the basis of postwar legislation. It recommended: a system of children's allowances; comprehensive health and rehabilitation services; and the avoidance of mass unemployment.

In 1943 and 1944 two White Papers discussed the educational system and the proposed national health service.

After the end of World War II a series of acts increased the scope of the social services, forming a unified system. Certain of these, such as the Family Allowances act, 1945, the National Insurance act, 1946, and the National Insurance (Industrial In-

juries) act, 1946, were based on the Beveridge report.

Since the 19th century there has been a revolution in attitudes toward social security. Whereas social service used mainly to consist of charity from rich to poor, it now involves the satisfaction of the basic needs of individual members of society, provision for this being made by the whole community.

Professional social work has rapidly grown in status and many official bodies now employ social workers whose job is to relate statutory provision to individual need. Many universities offer social science courses and various types of social work require an additional short specialized training course. The report of the working party on social workers in local authority health and welfare services, 1959, made a number of suggestions as to reorganization of training.

Voluntary organizations pioneered almost all the social services and in the mid-20th century complement the statutory provision, undertaking work outside its scope. Such bodies operate in almost all fields of social work and are often publicly aided. Their central co-ordinating agency in England and Wales is the national council of social service, which sponsors citizens' advice bureaus.

Youth Work.—This was originally voluntary but from as early as 1918 local education authorities have been indirectly connected with it. In 1939, the board of education issued a circular, which accepted "a direct responsibility for youth welfare." The board encouraged local education authorities to provide facilities for the leisure activities of youth and to co-operate with voluntary bodies. The White Paper on educational reconstruction also considered the youth service and the Education act, 1944, made compulsory the local authorities' duties in connection with youth work, which had until then been permissive.

The Albemarle report on the youth services in England and Wales (1960) points out that with the McNair report (published 1944) on youth leadership and with the Youth Advisory council's two reports (1943 and 1945) there was a sense of optimism with regard to the service just after the war and youth work expanded. Economic crises, however, forced the ministry to hold back the development of youth services in favour of educational expansion. The Jackson committee on recruitment, training and conditions of service of youth leaders and community centre wardens of 1949, and the Fletcher committee on the recruitment and training of youth leaders and community centre wardens of 1951, produced reports that were not implemented. Since then, although voluntary and professional workers have done very valuable service to youth work, on the whole it has probably not received either the statutory or public support it needs. Many authorities appoint full-time youth workers and often have a specialized youth committee to deal with the work. If this is not the case, supervision is undertaken by some other responsible committee, answerable to the local education authority. In addition to local authority assistance and ministry of education direct grants, the youth service obtains funds from voluntary sources—from subscriptions, gifts, trusts and special appeals, etc.

In 1936 the Standing Conference of National Juvenile Organizations was formed, later renamed the Standing Conference of National Voluntary Youth Organizations. There are 26 constituent members to the conference, 14 observers and 3 associated organizations. Questions of common policy may be discussed between the ministry of education, the standing conference or individual youth organizations. The Albemarle report estimated that the youth service attracts only one in three young people. Youth organizations fall into the following main categories: adventure and character training organizations; local youth groups; junior political groups; preservice organizations; and practical organizations such as the National Federation of Young Farmers' clubs.

The National Association of Boys Clubs and the National Association of Mixed Clubs and Girls Clubs unite local youth groups. Other organizations concerned with youth include the National Playing Fields association, the Central Council of Physical Recreation, Outward Bound trust, Youth Hostels association and the British National Committee of the World Assembly of Youth.

National Health Service.—The first Public Health act was passed in 1848. Central control of public health grew in the second half of the 19th century when such services as pure water supply, sewage and refuse disposal and street cleaning were provided. Hospitals were in some cases set up by local authorities, although most medical and hospital services remained private and voluntary. The National Insurance act, 1911, provided a measure of state-aided medical care. The early years of the 20th century saw increased provision of local authority hospitals, improved control of tuberculosis and the establishment of maternity and child welfare services. There was, however, no over-all co-ordinating medical and hospital system until the 1946 National Health act, which established a comprehensive health service, universal in scope and to a great extent (except for national insurance contributions) free of charge. Subsequent legislation is the National Health Service (Amendment) act, 1949, the National Health acts, 1951 and 1952, and the National Health Service Contributions acts, 1957–61.

Ultimate control for the service lies with the minister of health, advised professionally and technically by the central health services council. The services provided under the act are in three separate groups: hospital and specialist services, personal practitioner services and local authority services. The greater part of the cost is met by general taxation, a small part from local rates, and some from payment for services where this is required.

Hospital and specialist services cover all types of care including the treatment of maternity cases, infectious diseases, tuberculosis, the chronic sick, convalescents and the mentally disturbed or defective. England and Wales are divided into 15 hospital regions, each of which is administered by a regional hospital board.

Doctors and dentists are free to contract in or out of the service and may have private patients while within the scheme. Doctors' prescriptions, for which there is a small charge, may be dispensed by any health service chemist (druggist). Ophthalmic treatment is also provided.

The local health authority services cover maternity and child welfare, midwifery, health visiting, home nursing, aftercare, temporary domestic help during sickness or confinement, vaccination, immunization, local health centres and ambulance services. Maternity and child welfare services include free dental care, the provision of antenatal and postnatal clinics, health visiting and often the provision of special toddlers' clinics. Approximately three out of four babies in England and Wales attend the child welfare centres for regular supervision. Expectant mothers may be booked for priority hospital beds at confinement because of medical or obstetric reasons or because of home conditions. If there is a surplus of hospital beds in an area, mothers may choose whether or not to be confined in hospital or at home. Hospitals usually provide ante- and postnatal care for their patients. A general practitioner obstetrician may attend a home confinement or the family doctor if willing, as well as the trained local authority midwife. Antenatal and postnatal examinations are given and the midwife gives routine attention. There is also an emergency mobile obstetric squad.

Many local health authorities make special arrangements for premature infants born at home and they may provide day nurseries for under fives. Private and factory nurseries must be registered as must privately employed persons caring for more than two children not of the same family.

Unmarried mothers obtain special help in addition to the normal services. Much of the work carried out for these mothers is voluntary. Various bodies (often religious) provide homes for mothers and babies. Local authorities may contribute to the cost of the work and some run their own homes and appoint social workers to deal with this problem.

Local authorities must be notified of infectious diseases and cases are isolated. Other principles applied against these types of illness include the supervision of contacts, vaccination and immunization, and special hospital accommodation is provided. There are hospitals, clinics and sanatoriums for illnesses such as tuberculosis and venereal diseases where both specialized treatment and aftercare are given. Mass miniature radiography was introduced in 1943 to help combat tuberculosis and preventive

measures against disease are receiving attention.

Mental health services conform closely to the other health services. Mental hospitals and mental deficiency hospitals are administered by regional hospital boards. Local health authorities are responsible for admissions, which in the case of mental illness are usually voluntary. The local authorities and hospitals both carry out preventive measures and aftercare. Local health authorities also arrange home visiting for the mentally subnormal and occupation and training, usually in special centres. Short-term care in emergencies is also their responsibility.

The general hospital services stem from two roots—from voluntary hospitals and from the municipal hospitals which in turn arose from the old Poor law infirmaries. Local authorities also once provided the majority of hospitals for the mentally sick or defective.

In 1939 the ministry of health under the emergency hospital scheme became responsible for organizing existing facilities on a national level while hospitals remained responsible for actual treatment. The National Health act meant that almost all hospitals (with a few exceptions) were organized in a single service. Routine hospital administration devolves upon management committees, usually responsible for a group of hospitals. Teaching hospitals are separately administered. Usually the patient's own doctor makes arrangements for hospital admission, treatment and specialist consultations. Treatment is free and a certain amount of aftercare during convalescence is provided. Certain hospitals have diagnostic ear clinics which may supply national health hearing aids to their patients. Specialist rehabilitation facilities for those in need are available at main hospitals and in a few specialist centres. The ministry of labour and national service also operates a disablement resettlement service. The blood transfusion service is administered by the regional hospital boards.

Medical and psychiatric social work is becoming increasingly important. In addition to psychiatric social workers many hospitals have almoners' departments which apply social casework principles to patients in need of help through family and other difficulties. Some local authorities employ almoners in their health and welfare services.

Voluntary organizations provide many types of aid to the sick and handicapped. More than half the hospitals in England and Wales have "leagues of friends" or similar bodies. Hospitals may also be helped by the British Red Cross society, the Order of St. John and the Women's Voluntary services.

National Insurance and Related Services.—Before World War II social insurance and assistance schemes in England and Wales were far reaching but piecemeal and unco-ordinated. In 1944 the ministry of pensions and national insurance was created; it administers family allowances and national insurance under the National Insurance acts, 1946–60, the Family Allowances and National Insurance acts, 1952–61, and National Insurance (Industrial Injuries) acts, 1946–61. Single weekly insurance contributions are obtainable from three classes of persons: self-employed, employed and nonemployed. The only persons not liable to contribute are children, students, certain married women and widows, certain people over pensionable age and those with very small incomes. Employers contribute as does the exchequer. A small sum to cover industrial injuries is included in the class 1 contributions met by employed people and their employers.

The national assistance board is responsible for grants under the National Assistance act, 1948. For these there is no necessary qualification but need. The Children's act, 1948, created an independent scheme to care for children without normal home care. Married women may claim exemption from insurance contributions in which case they may only claim benefits through their husbands. Benefits may be paid for maternity and home confinements, for unemployment, death, industrial injury and disability, to widows and to guardians. Family allowances which require no insurance qualification are also paid to families with two or more children as are children's special allowances in certain circumstances. The National Insurance act, 1959, introduced a scheme of graduated contributions and pensions.

The national assistance board has many functions including provision for those unable financially to maintain themselves; the supplementation of insurance benefits where necessary; and administration of noncontributory pensions under the Old Age Pensions act, 1936. The national assistance board also administers hostels under the Polish Resettlement act, 1947, and provides accommodation for those "without a settled way of living," although most reception centres are run by local authorities on the board's behalf. The board has two residential centres for those demoralized by long unemployment.

All local welfare authorities either provide services for the blind or delegate this work to voluntary organizations. Most authorities also either provide welfare services for the deaf and handicapped. Voluntary organizations are very active in the whole of this field. National associations include the Royal National Institute for the Blind, the National Library for the Blind, St. Dunstan's (for those blinded in war) and the National Institute for the Deaf. Local authorities must maintain a register of the blind in their areas and of other physically handicapped persons wishing to register.

After the end of World War II there was a great expansion of services for the old. Local old people's welfare committees are aided by the national corporation for the care of old people, established in 1947.

Care of Children Deprived of Normal Home Life.—Under the Children's act, 1948, local authorities must receive into care any child under 17 years of age who is without parent or guardian, or who is abandoned or whose parents cannot provide for him or her. Under the Children and Young Persons acts 1933–52, local authorities must also care until the age of 18 for those regarded by juvenile courts as needing care or protection. Children in care are fostered or, if this is not possible, maintained in a children's home run by a local authority or voluntary body. Voluntary homes must be registered and all homes and boarding out arrangements may be visited by home office inspectors. The home office is the central government department concerned with child care. Local authorities appoint a children's committee and a children's officer responsible for the area. The officer is assisted by child-care officers. Cost of the services is shared by the local authorities and the exchequer. Children's departments may also investigate cases where children living with parents or guardians may need care or protection. The National Society for the Prevention of Cruelty to Children maintains inspectors and visitors to investigate cases of cruelty and neglect.

Adoption, which was first legally regulated in 1926, is now governed by the Adoption act, 1958.

Various local authorities have appointed committees to co-ordinate statutory and voluntary services dealing with the family. (*See* also SOCIAL SERVICE; SOCIAL SECURITY; CHILDREN, LAWS CONCERNING.)

4. Police and Fire Protection.—There are 125 separate independent police forces which may be county forces, borough forces, combined forces (where the area of responsibility extends over neighbouring counties or boroughs), the metropolitan police force and the City of London force. Forces vary in size according to their area and population coverage. There are also part-time special constables and police forces for certain public bodies and the services.

In the counties the police authority is the standing joint committee. In the boroughs the police committee is known as the "watch committee" and in the combined areas the authority is composed of representatives of the constituent areas. The authority for the metropolitan police is the home secretary; that in the City of London is the Common council, which usually appoints a standing committee. Co-ordination of the police service is exercised through the home secretary's police regulations. The home secretary is assisted by her majesty's inspectors of constabulary (five men and one woman in England and Wales). Police officers may be serving regulars for 25 years or more, members of the first police reserve and special constables. Police cadets are recruited by most forces, which may also employ civilian staff. A small number of women also are recruited. Police duties vary greatly, from work such as traffic control to preparing and pre-

senting straightforward cases in magistrates' courts.

All forces have their own criminal investigation departments and may have other specialized departments such as river police, mounted police and police dogs. There are a number of common services supplementing those provided by the individual authorities. For example, there is a common training service. The metropolitan police maintain a criminal records and fingerprinting service, publish the *Police Gazette*, maintain a special branch of the Scotland Yard criminal investigation department for protection duties, expedite extradition orders and act as agents of the International Police commission. For certain of these services the metropolitan police receives an exchequer grant. The Police Council for Great Britain is the negotiating body for the service. It was established in 1953 and is composed of representatives of both the police service and local police authorities.

The fire services are under the central control of the home secretary, assisted by inspectors of fire brigades and organized on a local basis under the Fire Services act, 1947. By the early 1960s there were more than 130 local authority fire brigades in England and Wales. The local county council directly controls the London fire brigade, which serves all London including the City. Cost is borne by the local authorities, the exchequer contribution being included in the general grant. There is a Joint Fire Research association dealing with all aspects of fire prevention and fire fighting. (*See* also POLICE; FIRE PROTECTION.)

5. The Judiciary.—In England and Wales the judiciary is independent of the legislature, central responsibility for it lying partly with the home secretary and partly with the lord chancellor. The prime minister is responsible for recommendation to the crown for certain appointments.

Common law based on custom has been gradually built up by the courts. A supplementary system of law, introduced during the middle ages, was known as "equity." Equity and common-law courts were amalgamated in 1873. Statute law includes acts of parliament and delegated or subordinate legislation made under powers conferred by parliament. Generally speaking, although there are certain exceptions, all indictable or more serious crimes require trial by jury in a superior court.

Courts may be broadly divided into civil and criminal, although the demarcation is not rigid. The more important civil courts in England and Wales are: (1) County courts, of which there are nearly 400, most of which do not sit continuously; as well as ordinary county courts, there are still a few local courts of similar jurisdiction. (2) The high court of justice, comprising the high court and the court of appeal; the high court sits in three divisions—the queen's bench division, the chancery division and the probate, divorce and admiralty division. (3) The house of lords, which in addition to the court of appeal may be used as an appellate court in civil-law actions.

Criminal courts include: (1) Petty sessional or magistrates' courts (juvenile courts are specially constituted magistrates' courts). (2) Courts of quarter sessions, of which there are 65, which may be county or borough sessions. (3) Courts of assize, which are branches of the high court of justice; the assize judges work on circuit. (4) Crown courts at Liverpool and Manchester which act as quarter sessions for those two cities and also as assize courts for the West Derby division and the Salford division of Lancashire. (5) The central criminal court of the Old Bailey which acts as court of assize for London, Middlesex and parts of the home counties.

Other courts in England and Wales are coroners' courts, the ecclesiastical courts of the established church and military courts or courts-martial. There are also a number of administrative tribunals to enforce professional discipline and to deal with disputes where specialized knowledge is required or where a government department or public authority is interested. These tribunals receive general supervision from an advisory council on tribunals, under the Tribunals and Inquiries act, 1958.

The legal profession is divided into two categories of lawyers: barristers and solicitors. In general it may be said that solicitors pursue legal business for lay clients, while barristers advise on legal problems submitted through solicitors and conduct proceedings in the higher court. (*See* also ENGLISH LAW.)

Legal Aid.—The Legal Aid and Advice act, 1949, which applies to England and Wales, based upon the Rushcliffe committee report (Cmd. 6641, 1945) as amended by the Legal Aid act, 1960, aims to provide free or low-cost legal aid for those whose "disposable income" does not exceed £700 a year. Where, however, the litigant has disposable capital of more than £500, or such larger figure as may be prescribed, the grant of legal aid may be refused if it appears that he could afford to proceed without it. England and Wales are divided into 12 areas and within each area a committee of solicitors and barristers administers the legal aid scheme and sets up legal aid centres to which those seeking legal aid apply. General responsibility for the scheme lies with the Law society guided by the lord chancellor; costs are met from the legal aid fund and supplied by contributions and fees from aided persons, costs recovered in litigation or an exchequer grant.

Before any applicant's case goes forward his means are investigated by the national assistance board. The question of whether or not a litigant should make a contribution to cost is determined, after certain deductions and allowances, by his annual "disposable income" (prescribed at £250 or less for free legal aid and not in excess of £700 for contributory aid) and his annual "disposable capital" (likewise £125 or less and not in excess of £500). Oral legal advice is also available at low cost through the legal aid fund to those who could not normally afford it. No charge is made to those receiving national assistance. For persons ineligible for this, the Law society runs a voluntary scheme. Free legal aid in criminal courts may be granted in certain cases. The legislation governing this is the Criminal Appeal act, 1907; the Poor Prisoners' Defence act, 1930; the Summary Jurisdiction (Appeals) act, 1933; the Legal Aid and Advice act, 1949; the Cost in Criminal Cases act, 1952; and the Legal Aid act, 1960 (*see* LEGAL AID).

Treatment of Offenders.—Courts may impose fines, sentence to imprisonment or other forms of detention, discharge absolutely, discharge conditionally, or place on probation. Two special forms of treatment for persistent offenders are corrective training and preventive detention. The Homicide act, 1957, restricts the death penalty to "capital murders" or those forms which most directly assail law, order and public peace. The death penalty is subject to reprieve on the exercise of the crown prerogative of mercy.

The home secretary is responsible for all penal institutions, and he is advised by the advisory council on the treatment of offenders. The prison commission is responsible to the home secretary for the administration of prisons, borstal and detention centres, while general supervision of individual institutions lies with visiting committees or boards of visitors. There are 48 prison establishments in England and Wales which fall into the following categories: (1) local prisons, receiving all classes of prisoner direct from the courts; (2) special local prisons, either open or walled, for short-term prisoners, often first offenders; (3) six young prisoners' centres for those under 21, serving sentences of three months or more; (4) regional training prisons, which may be open, for first offenders and other types of trainable offender; (5) central prisons for recidivists, preventive detention prisoners and long-term prisoners of "star" class (one of the last type of prison is open).

Classification aims at separating the sexes, the young from adults, the untried from convicted and civil from criminal prisoners. It also aims at providing training suitable for those who may benefit and at preventing contamination. Prison training endeavours to provide work of a type which may help released prisoners to earn their living, educational facilities which may include vocational courses, and opportunities for developing personal responsibility, which may include residence in open prisons. There is a remission (parole) system dependent on good conduct. Prerelease courses operate at all prisons and home leave may be granted to prisoners. However, gross overcrowding of local prisons greatly hampers remedial measures. A White Paper (1959) initiated a formidable building program.

Treatment of Young Offenders.—No child under the age of 8 may be charged at law and between 8 and 14 children must be dealt with in juvenile courts unless charged jointly with an adult

or charged with homicide. In England and Wales a person between 14 and 17 years will usually be dealt with summarily for an indictable offense or may be tried by jury. In certain cases young persons may claim to be tried by jury. In addition to the methods of treatment listed for adult offenders, courts may commit young offenders to the care of a fit person (usually the local authority). The following institutions exist: (1) remand homes, which are provided by local authorities for the custody of those under 17 before appearance in court, between court appearances, while awaiting approved school vacancies or for up to a month of punitive detention; (2) approved schools, which are residential and provide education and training for young offenders or for those in need of care or protection.

Approved schools may be provided by local authorities, voluntary bodies or local committees and must be approved by the home secretary; in the early 1960s there were about 120 in England and Wales. Approved schools are classified by age of admission. The period during which a child may be detained in an approved school is determined by law. Statutory responsibility for aftercare rests on the school managers. Children may be released on licence as soon as the situation warrants this, and there is a further period of supervision.

After 1959 remand centres were established for those between 17 and 21 who might otherwise be committed to prison and those under 17 unsuitable for remand homes. About 40 attendance centres have been established in England and Wales for those aged between 12 and 17, guilty of offenses for which adults can be imprisoned. Boys attending do so in their spare time. An experimental attendance centre operating on similar lines for those aged 17 to 21 was established in Manchester. There are four detention centres in England and Wales (two junior centres and two senior centres) which aim to provide a short, strict lesson. Length of detention is usually three months and provides a brisk, disciplined regime.

Borstal institutions vary in type but are for offenders aged between 16 and 21. In England and Wales there are 2 allocation centres, 19 training borstals, 1 correction centre and 1 recall centre. For girls there are three training borstals and one recall centre. Sentences range from nine months to three years with a total period of control extending to four years. (*See* BORSTAL SYSTEM; CHILDREN, LAWS CONCERNING; JUVENILE DELINQUENCY: *Great Britain;* CHILDREN'S COURT.)

Probation Service.—The probation service is governed by the provisions of the Criminal Justice acts, 1948, and relies on the offender's co-operation. The only offenses for which probation may not be used are those punishable by death or life imprisonment. Supervision must be for not less than one year and not more than three years. Other requirements may be asked of the offender and if he fails to comply with the probation order or commits a further offense, he may be dealt with for the original offense. The probation officer's primary functions are to supervise juvenile and adult offenders and children and young persons in need of care or protection or beyond control. He may also undertake matrimonial conciliation, advise the courts on custody of children and assist in the aftercare of certain discharged prisoners and those released from approved schools and borstals. A probation training scheme (instituted in 1936) is managed by the home office, with the advice of the probation advisory and training board.

6. Education.—In 1833 the government began making annual grants to voluntary societies providing schools, and in 1870 the Elementary Education act legally introduced the idea of compulsory primary education. The statutory provision of secondary education in England dates from the 1902 Education act. Wales had a lead in this field and had provided many grammar schools as a result of the Welsh Intermediate Education act of 1889. The 1918 Education act (or Fisher act) raised the age of compulsory school attendance to 14 years. The 1921 Education act embodied the 1918 act and consolidated earlier acts. An Education act passed in 1936 was shelved because of World War II. Early in the war the government began to prepare for new legislation and presented a White Paper to parliament in 1943 containing proposals for educational reconstruction. The proposals were generally accepted and a new act received royal assent in 1944.

The cornerstone of education in England and Wales is the Education act, 1944. This remodeled the public system of education. Further amending acts were passed in 1946, 1948, 1953 and 1959, aimed at improving the working of the 1944 legislation while preserving its intentions. The 1944 act stated that

> the statutory system of public education shall be organized in three progressive stages, to be known as primary education, secondary education and further education, and it shall be the duty of the local education authority for every area, so far as their powers extend, to contribute towards the spiritual, moral, mental and physical development of the community by securing that efficient education throughout those stages shall be available to meet the needs of the population of their area.

Every parent is required to cause his child "to receive efficient full-time education suitable to his age, ability and aptitude" between the ages of 5 and 15. The act also made special provision for nursery education, but there has been comparatively little development of nursery facilities.

Administration.—The minister of education is assisted in his duties by the permanent officials of the ministry and by her majesty's inspectorate of schools. He supervises the work of the local education authorities, which comprise 62 county councils, 83 county borough councils and 1 joint board representing the areas of a county and a borough in England and Wales. They have, however, a considerable measure of autonomy and the educational system is highly decentralized. The ministry does not own or control educational institutions or appoint teachers. These duties lie with the local education authorities or occasionally with the individual school authorities. The minister's control of the local authorities, apart from the provisions of the 1944 act, operates through statutory regulations, circulars, inspection and conditional grants.

In addition to providing and maintaining what are known as "county schools," the local education authorities maintain a considerable number of schools which have been provided by voluntary organizations (often religious bodies) and which are known as "voluntary schools." A third category of school is the "direct grant school," completely independent of the local education authority, managed by its own governing body and in receipt of a grant-in-aid from the ministry.

Teachers are free from official direction over teaching methods, curricula, etc., and such matters are usually settled within the individual school in conjunction with the heads of departments or principal and with relevance to examination needs. Universities are completely free from central or local governmental control.

Primary Education.—In England and Wales this lasts from age 5 to about age 11. A primary school may be an infant school for children aged 5 to 7, a junior school for children aged 7 to 11, or, if a smaller school, may contain both infant and junior children. When the local authorities allocate children to different types of secondary school, this is often done on a basis of the children's general work standard at primary school coupled with objective tests of various kinds undertaken at about age 11. There is a large amount of experiment with selection techniques.

Secondary Education.—Secondary schools in England and Wales are normally of three types.

1. Grammar schools provide education for those children with an academic bias up to the age of 16, when the ordinary level general certificate of education is taken, or beyond this to the ages of 17, 18 or even 19. General certificate of education examinations are also set at advanced and scholarship levels. University entrance requirements and those of most professional bodies are expressed in terms of this examination, which is set by nine examining bodies.

2. The small number of technical schools provide a general education relating to industry and commerce.

3. Secondary modern schools, the largest group, supply an education with a practical bias for those with no special technical or commercial aptitudes, usually up to the age of 15, although an increasing number of secondary modern children are remaining at school until 16. It is intended to raise the leaving age to 16 as soon as practicable.

Some authorities have established comprehensive schools providing all types of secondary education.

There is a large number of independent schools, which are visited by her majesty's inspectors. The best-known independent schools are those known as "public schools"; these usually cater for the 13–18-years age group and lay great emphasis on character training.

Teachers.—By the early 1960s the average ratio was one teacher to 26 pupils in primary and secondary schools publicly maintained in England and Wales. Nevertheless, large classes are found, and to help reduce their size the government has undertaken a place program of teacher training college expansion which involves an extensive building program. The normal two-year training course was also extended by a further year in Sept. 1960. Teacher training may be undertaken in training colleges by students of 18 and over, or by graduates in one-year courses at the various university departments of education.

Further Education.—Numerous opportunities for further education exist for those who have finished full-time schooling. These may be full-time or part-time, vocational and nonvocational.

Technical or vocational education is for the most part provided by the local education authorities in technical, commercial or art colleges and evening institutes. Higher technical education is also provided in technical colleges. In contrast with the compulsory schooling, period fees are charged, though these are usually low and often nominal for students under 18. Technical colleges fall into several groups: (1) local colleges that provide mainly part-time courses of two to five years, orientated to such technical qualifications as ordinary national certificates and the city and guilds of London institute certificates; (2) area colleges that also provide certain advanced courses of two to four years for those of 18 or 19, aiming at higher national certificates and diplomas; (3) regional colleges that cover larger areas than the first two institutions and provide a substantial amount of advanced work; (4) colleges of advanced technology, of which there are eight, which provide no low-grade courses, concentrating on advanced work, postgraduate studies and research.

There are seven national colleges providing advanced courses for specialized work relating to particular industries. The diploma in technology, first awarded in 1958, has been introduced as a qualification equivalent to a university honours degree; and a higher distinction, membership of the college of technologists, is being instituted and will be awarded after a three-year period of advanced work of research.

There are a number of specialist institutions offering courses in art, music and dramatic art.

Universities.—There are 16 universities in England and 1 in Wales. These are independent and self-governing, although the greater part of their income derives from public funds. Each university awards its own degrees and all of them provide for postgraduate work and research. The universities of Oxford and Cambridge (*qq.v.*) date from the 12th and 13th centuries and each consists of a group of largely self-governing residential colleges. London university (*q.v.*) and all other universities in England and Wales were founded in the 19th and 20th centuries. *See* also the appropriate sections of EDUCATION, HISTORY OF; ELEMENTARY EDUCATION; ADULT EDUCATION; UNIVERSITY; TEACHERS, TRAINING OF; TECHNICAL EDUCATION; SCHOOL ADMINISTRATION. (JO. DA.)

VII. THE ECONOMY

The economy of England is an integral part of that of the United Kingdom and while for some purposes the interests of Scotland and Northern Ireland, and to a lesser extent Wales, are separately considered and administered, industry, trade and commerce are economically interwoven.

Supplying but one-half of its population's food requirements, poor in natural resources (with the exception of coal), the United Kingdom, which was the first country in the world to become highly industrialized (*see* INDUSTRIAL REVOLUTION), is primarily concerned with manufacturing industry and commerce and produces almost 20% of the world's exports of manufactured goods. Agriculture, forestry and fishing contribute only 4% and mining and quarrying 3.5% to the gross national product, while manufacturing industry accounts for about 36%. England makes the most important contribution to the nation's economy because of its size (53% of the total land area), population (82%) and greater natural resources. The chief of these last are coal, iron ore, crops and grass.

1. Agriculture.—The physical environment and natural resources of England are much more favourable to agricultural development than those of other parts of the U.K.; a greater proportion of the land is rich-soiled lowland where the lower rainfall and higher temperatures are more favourable to crop growth. The poorest land, the moors and rough grazings, make up two-thirds of Scotland and more than one-third of Wales, whereas in England they represent only one-tenth of the country.

Prior to 1939 the farms of Britain produced approximately 30% by value of the country's food requirements. The overwhelming dependence on imported food was the product of long-term trends beginning in the 19th century. The growing population of Britain could be fed more cheaply by foods produced under more favourable conditions overseas. This foreign competition affected cereals to the greatest extent but also, with the introduction of refrigeration, came to affect livestock products until only liquid milk, by nature of its perishability, remained completely immune. Legislation to protect the home producer of foodstuffs was introduced in the 1930s and brought to an end the free trade economy which had prevailed for almost a century. Between 1871 and 1939, however, the arable area of England and Wales had declined by almost 6,000,000 ac. and the emphasis of farming shifted from crops to livestock. Changes in English agriculture during the 20th century reflect the attempts of farmers to avoid foreign competition, the degree of protection afforded them by government legislation and the increase in living standards with consequent changes in diet and taste. The majority of farmers became stock farmers and by 1939 about 70% of the value of the annual agricultural output was made up of livestock products and only 15% of farm crops. Milk was by far the most important item, followed by beef, poultry and eggs, pig and sheep meat. Potatoes were the most valuable cash crop, followed by wheat, barley and sugar beet.

The heavy dependence on imports was a severe disadvantage during the blockade of World War II. The war economy required more food production per acre and more land under the plow, and by 1944 agricultural production was 25% up on the prewar level with a much-reduced fodder import and a smaller farmed area because of the land requirements of the armed forces (*see* Table V).

The Agricultural act of 1947 laid the foundation of postwar policy. The act was designed to assist the growth of agricultural production to 50% above the prewar level. As the world's food supplies improved and as British production rose, government control was relaxed, though the principles of agricultural support were maintained in all postwar legislation. Stress was laid upon guiding production to meet home demand at the lowest possible cost. The 50% increase in output was achieved by 1953–54 and rose to 68% by 1960, an increase achieved with a smaller labour force (about 500,000) but with greater mechanization.

The majority of English farms are small, the average size being between 50 and 100 ac. There are more than 300,000 holdings, including many under 20 ac. in size which are but part-time enterprises. Approximately half the farms are owner-occupied.

TABLE IV.—*Size of Holdings* (*England, 1959*)
(in acres)

Under 20	20–49¾	50–99¾	100–149¾	150–299¾	300–999¾	Over 1,000	Total
136,108	50,123	47,410	25,838	30,597	12,736	658	303,467

Source: Ministry of Agriculture, Fisheries and Food.

Major Crops.—Wheat growing is mainly confined to the drier, sunnier counties of eastern and southern England. New strong varieties have become increasingly widespread in these areas and yields with the aid of chemical fertilizers and weed killers are at 24 to 25 cwt. per acre, well above the prewar level of 18.5 cwt. The use of combined harvesters has concentrated wheat sales into

TABLE V.—*Crop Acreages*
(in 000 ac.)

Crop	1939	1944	1960
Wheat	1,670	2,950	1,825
Barley	888	1,641	2,768
Oats	1,198	1,876	990
Potatoes	438	913	546
Sugar beet	336	415	417
Turnips and swedes	370	431	213
Mangolds	199	284	120
Fruit, vegetables and flowers	549	718	684
Temporary grass	1,858	2,588	3,995
Total arable	8,365	13,333	12,661
Permaneot grass	13,549	8,439	9,199
Total acreage, crops and grass	21,946	21,733	21,860

Source: Ministry of Agriculture, Fisheries and Food.

the early autumn and led to a reduction in price at that period. Combines have also necessitated special drying and storage facilities if harvests are to be millable. The standard price guaranteed for wheat is varied seasonally to offset the low price periods and help in storage costs.

Barley is the chief grain crop in England. It is grown in approximately equal proportions for livestock feed and for the malting and industrial market, the latter commanding the higher price. Yields have increased by some 3 to 4 cwt. above the 1939 average of 17.5 cwt. per acre.

The acreage under oats has declined greatly since World War II. In arable areas it is at a disadvantage compared with barley, which gives higher yields and is a dual-purpose crop. The horse population which once it fed has been much reduced. The virtues of the crop lie in its tolerance of humid, cloudy conditions and acid soils. Under these conditions, which pertain in northern and western England, it is the most profitable cereal, but during the 1950s these same areas reduced their arable acreage and concentrated on the production of better and cheaper grass for livestock.

The sugar-beet crop has to compete with imported cane sugar and but for government protective legislation would be uneconomic to grow. The industry was created by the Beet Sugar act of 1925, the first legislation to protect an agricultural enterprise in Britain since the Corn laws of the 19th century. Within ten years of the passing of the act 18 beet factories were erected in suitable arable districts; transport costs have limited beet growing to these districts. Sixteen of the factories are in the English arable belt between Yorkshire and Essex, two in the dry border counties of Shropshire and Worcestershire and only one, in Fife, is outside England. The British Sugar corporation operates the beet factories and the whole crop is grown under contract to it. Average yield is usually in excess of 13 tons per acre. The Sugar act of 1956 which established the Sugar board and the Agriculture act of 1957 maintain the protection of the industry.

The potato yields heavily but requires high fertility and intensive cultivation. The early crop which commands the highest price is grown in many widely scattered areas where climate and soil favour early growth. Main crop potatoes are most important in areas of intrinsic soil fertility, particularly on silts and reclaimed peats. Accordingly the principal potato areas are the peats and silts of the Fenland, the silts of the upper Humber estuary and the peats of southwest Lancashire. The average yield in England in 7.8 tons per acre and the bulk of seed stock is obtained from Scotland.

Horticulture.—The growing of vegetables, fruit and flowers in market gardens is one of the most intensive forms of cultivation practised in England. Though accounting for only 2.5% of farmed land, horticulture contributes almost 9% to the value of agricultural production. The growing importance of vegetable and fruit crops reflects the changing diet in England consequent upon a rising standard of living. Market gardening is found round all large towns, the proximity of a market, so important for fresh and perishable produce, being of more consequence than climate and soil. The intensive nature of the economy allows the producer to manipulate both soil and climate (by greenhouses and irrigation) to suit the crops grown. Other horticultural areas are, however, to be found where soil and climatic conditions are so favourable as to stimulate production even at some distance from the main urban areas. This is particularly true of regions where climate and soils allow early growth.

Greater London has peripheral horticultural areas in Essex and Hertfordshire; the west Midlands conurbation buys produce grown in Staffordshire, Warwickshire and Worcestershire; while Merseyside and Manchester are supplied from light-soiled areas in southwest Lancashire and Cheshire. Farther afield are important areas depending on transport by rail and truck to serve a much wider market. The southern coast of Cornwall with its mild climate produces early crops marketed throughout the country. Other specialist areas, generally with light, easily worked soils, include the Cottenham and Wisbech districts in the Fens, the Botley area of Hampshire, the Swanley and Dartford districts of Kent, the Vale of Evesham, famous for both its vegetables and tree fruit, and the Sandy and Biggleswade area in the Bedfordshire. The main centres of greenhouse production are in the Lea valley in north London, and round Worthing in Sussex.

Kent is the most important county for the production of tree fruit. Cultivation was established on a commercial scale in the 16th century and includes apples, pears, plums and cherries. Worcester, another famous orchard county, is most noted for its plums, while Hereford, Somerset and Devon produce England's cider apples. Of the soft fruits strawberries are a specialty in Worcester, Hampshire and Cambridgeshire and black currants in Norfolk.

Livestock.—The agriculture of England, though to a lesser extent than in Wales and Scotland, is essentially concerned with livestock husbandry and, in particular, with milk production. Two thirds of the agricultural output is made up of livestock products, milk and dairy products alone accounting for one fifth.

TABLE VI.—*Livestock Production*
(in 000 head)

Livestock	1939	1944	1960
Cows and heifers in milk	1,977	2,006	2,199
Total cattle	5,911	6,260	7,309
Sheep and lambs	13,337	8,654	13,029
Pigs	3,307	1,353	4,481
Poultry	52,019	28,559	80,804
Horses	476	418	63*

*1958.
Source: Ministry of Agriculture, Fisheries and Food.

About 85% of the cows and heifers kept on farms in England are milk producers in dairy herds. Dairying is practised in every county though the main concentrations of milch cows occur in the northwest in Shropshire, Cheshire, Staffordshire, Derbyshire and Lancashire, and in the southwest in Somerset, Dorset, Cornwall and Wiltshire. Throughout the 20th century dairying has come increasingly to dominate English farming but its expansion has been particularly rapid since the establishment of the Milk Marketing board in 1933. More than 80% of the milk sold is consumed in the liquid market where it can command a much higher price than milk bought for manufacture into dairy products. The immunity of the liquid milk market from foreign competition and the attractiveness of a regular monthly income derived from sales to the board, has doubled milk production since the board's establishment. The quality of dairy cattle was considerably improved after the war by the artificial insemination service, operated largely by the board, which makes available the very best dairy strains to all farmers. These developments have resulted in the higher yielding dairy breeds, the Friesian and Ayrshire becoming more numerous than the once dominant Shorthorn. Average yields per cow have increased from 550 to 745 gal. per year. A scheme for the eradication of bovine tuberculosis in British cattle introduced in 1934 was completed in 1960 when the whole country became an attested area.

Home production supplies approximately three-quarters of the country's beef needs and, since most English cattle are dairy cattle, a large proportion is cow beef derived from dairy herds. Approximately one-quarter of the dairy herd is killed each year for meat; in addition calves born to dairy cattle but sired by beef breeds, frequently by artificial insemination, are increasingly important sources of veal and beef. Special beef breeds, for which

Britain is famous, are bred throughout the country, but long-established specialist areas still persist: the Hereford is of great importance in its home area along the Welsh borderland, North Devons are bred in Devon and Cornwall, and the Sussex survives as a minor beef breed in the county of its origin. The rearing, store production and final fattening of beef cattle does not always take place in the same area. Movement of cattle is a feature of the beef trade and while rearing is more characteristic of upland Britain the final fattening takes place on lowland arable farms in eastern England and the Northampton-Leicestershire fattening pastures.

Hill sheep are bred in the Pennines, the Lake district and in Devon. In these areas sheep are occasionally the farmer's main concern; more usually they are of subsidiary importance to cattle. The draft ewes and lambs from these hill areas move down to the lowlands to be fattened or to be crossed with rams of larger, quicker maturing lowland breeds. Grass-fed breeds, giving the leaner joints the public now demands, are much more important than the large arable breeds, characteristic of the 19th century. The production of fat lambs rather than wool is the main concern of English sheep farmers.

Pig keeping is rarely a specialist form of farming though pig farms do exist supplying the large sausage and bacon companies. More commonly pigs are minor enterprises found on most farms, 70% being kept on farms of under 100 ac. Though less pig meat is eaten than before World War II a greater proportion is produced on English farms.

Poultry, like pigs, are kept in small numbers on most farms and though specialist poultry farms are increasing in number, more than 80% of laying birds are in flocks of fewer than 1,000 birds. Lancashire and Yorkshire have by far the greatest head of poultry. On poultry farms rearing, egg production and the feeding of table birds are specialist enterprises. A later development has been the production of broiler fowls which are intensively fed in cages as table birds on what are essentially nonagricultural establishments. The annual production of broilers in the early 1960s was estimated to exceed the total number of fowl kept on agricultural holdings. (*See* also AGRICULTURE: *British Commonwealth: The United Kingdom.*)

2. Forestry.—Britain, once heavily forested, has now less than 8% of its area in woodland. Of the total of more than 4,000,000 ac. about half are in England. Prior to 1914, with the exception of a few crown forests, the woodland of England was privately owned and was composed, in the main, of broad-leaved deciduous trees of which oak was the most common species together with beech, ash, birch, sycamore and elm. The bulk of timber, particularly of softwoods, was imported and the blockade of World War I caused about 450,000 ac. of woodland in Britain to be felled. The seriousness of this situation led to the establishment of the Forestry commission in 1919. Under the general direction of the ministry of agriculture and fisheries, the commission was entrusted with the development of British timber resources by the introduction of large-scale afforestation and reafforestation and by giving advice and assistance to owners of private woodland.

The commission made considerable progress between 1918 and 1939 but the fellings of World War II consumed a larger acreage of woodland than the commission had been able to establish since its foundation. This felling was chiefly of private woodland as the commission's forests were for the most part too young to be felled by 1939. After 1945 the Forestry commission planted at a faster rate than before the war and at almost twice the rate of private owners.

As a result of these changes the Forestry commission has come to operate a much larger proportion of the nation's woodland. State forests amount to one-third of the total forest area of Great Britain and 42% of the productive woodland. The bulk of Forestry commission planting has taken place in upland Britain, particularly in Scotland and Wales, contrasted to England, which has more extensive areas of private woodland. The chief state forests in England are the Kielder forest in Northumberland, Thetford chase in the Breckland of East Anglia, the New forest in Hampshire, the Forest of Dean in Gloucestershire, Sherwood forest in Nottinghamshire, Cannock chase in Staffordshire and the forests of the north Yorkshire moors. The commission, which maintains four forester training schools, has plantations in every county except London and Middlesex. The expansion of forestry in England has been limited by a number of factors. The greater part of Britain's most productive land is in England, hence agricultural interests inhibit woodland expansion. Not only has England less poor hill land but a large proportion of it is common land which cannot be acquired for planting. In the Lake District National park planting, by agreement, has been restricted to preserve the park's character. Tree growth on a commercial scale is difficult because of wind force in some extreme westerly and upland areas, while in the southern Pennines industrial fumes from the peripheral manufacturing areas are an added hazard to the planting of conifers.

The predominant species of tree differs in England from the rest of Britain, partly because the private woodlands which are more extensive in England contain a preponderance of broad-leaved deciduous trees, and partly because the Forestry commission plants different species in England. Since World War II, in England south of a line from the Severn to the Humber, where planting has been largely the reafforestation of felled woodland, the commission has planted more than a third of its woods with broad-leaved species, chiefly beech, while Corsican pine is the main plantation conifer. North of this line conifers are more important and represent about 90% of plantings, with Sitka spruce the principal species, as in Wales.

3. Fishing.—About 14,000 men are employed in fishing in England and Wales, the vast majority in England. The fish brought into English ports are caught over a wide area in distant waters, off Iceland, northern Norway, Spitsbergen, in the Barents sea and, increasingly, off Labrador and Newfoundland; in the middle grounds around the Faeroe Islands; in near waters in the North and Irish seas and, to a minor extent, inshore around the British coast.

The fish caught are of two main kinds: white fish, which are demersal or sea-bottom fish; and near-surface or pelagic fish, of which the herring is most important. Pelagic fish are caught in middle and near grounds often within 60 mi. of the British coast. Herring fishing begins in summer off the coast of Scotland and reaches the East Anglian coast by autumn. Lowestoft and Great Yarmouth are the principal herring ports of England. White fish are landed in much greater quantities. About 70% of the white-fish catch is brought in from distant waters by trawlers based on Hull, Grimsby and Fleetwood; cod, plaice, haddock, hake, halibut and sole are among the important species. The main branches of the industry are organized by the Herring Industry board and the White Fish authority. About 10% of the catch landed is brought in by foreign vessels. (*See* also FISHERIES: *British Fisheries.*)

4. Mining.—England is poor in mineral wealth and only its coal production (between 10% and 15% of the world output) is of international stature. Gold, copper, tin, zinc and lead have all been mined, and tin and copper production were once important. The ore bodies are now largely exhausted and the discovery of richer, more easily worked deposits abroad has closed almost all English mines since the mid-19th century. The annual production of tin from the mines of Cornwall and Devon averaged 1,000 tons metal content in the mid-20th century.

The output of iron ore is sufficient to meet less than half the demand of British industry and represents about 3% of world production. Salt deposits in Cheshire and on Teeside, and kaolin (china clay) deposits in Cornwall are substantial and serve the chemical and pottery industries respectively, while clays and shales for brickmaking and limestone for cement, agricultural use and ironmaking are mined in large quantities throughout the country. Minor minerals include barytes and fluorspar, which are mined chiefly in Derbyshire.

Despite widespread exploration no substantial mineral oil resources have been discovered in England. Small oil fields in Nottinghamshire and Lancashire yield 80,000 tons of crude oil per year, less than 1% of British consumption.

Coal.—This is the most important mineral and the chief indigenous source of power in England. Workable reserves are large

(estimated sufficient for about 400–500 years at the current rate of consumption) and cover the full range of bituminous coals, anthracite being found only in Wales and Scotland.

Coal production in Britain reached a maximum of 287,000,000 tons in 1913. Subsequently the rising costs of coal mining due to the senescence of some coal fields, and the loss of export markets as new sources of power were developed abroad and as foreign coal fields came to compete more vigorously in the export trade, led to a decline in output. In addition the conversion of merchant ships to oil firing, accelerated by the sinkings of two world wars, reduced the demand for bunker coal, while at home the increasing efficiency of coal-burning engines served to offset the rise in the internal demand for power.

In the period immediately after World War II Britain and the whole of Europe were faced by an acute coal shortage. The need was to reverse the trend of declining coal output and to produce more coal for home consumption and for export. The coal industry was faced with the problems of labour shortage, the physical and economic disadvantages of age and exhaustion in the coal fields and the urgent need for capital investment and re-equipment on a massive scale. In 1946 the coal industry was nationalized and passed into public ownership under the management of the National Coal board. In 1950 a long-term development program was begun which aimed at increasing output by 20% to 240,000,000 tons per annum by 1965; of this total, coal fields in England were to contribute 177,400,000 tons. By 1957 it became apparent that this target would exceed the coal demand of the mid-1960s and that the surplus would be difficult to export as a similar situation of excess coal was developing in Europe. This deceleration in the demand for coal was due largely to the use of other sources of energy, in particular of oil. In 1959 a *Revised Plan for Coal* was published in which production targets were reassessed for each field.

TABLE VII.—*Deep-Mined Coal Production by N.C.B. Divisions and Coal Fields (England, 1959)*

Division and fields	Production (million tons)	Production as percentage of U.K. total	Cost per ton (shillings)	Output per man (shift-tons)
Northern division	12.72	6.5	85/11	1.29
Northumberland	11.78	6.1	83/-*	—
Cumberland	0.94	0.4	115/9	—
Durham division	23.32	12.1	96/4	1.07
(Durham field)				
Northeastern division	41.9	21.7	74/11	1.49
(Yorkshire field)				
Northwestern division†	14.01	7.2	91/8	1.24
Lancashire	11.94	6.2	91/-*	—
East Midlands division	44.5	23.1	62/2	2.00
North Derby and Nottingham	37.49	19.5	—	—
South Derby and Leicester	7.09	3.6	57/5	—
West Midlands division	15.37	7.7	82/-	1.34
North Staffordshire	5.6	2.8	86/5	—
Salop and South Stafford	1.21	0.6	90/-	—
Warwickshire	4.53	2.3	78/7	—
Cannock Chase	4.01	2.0	77/5	—
Southeastern division	1.60	0.8	104/-	1.07
(Kent coal field)				
Southwestern division‡	20.4	10.5	99/6	1.01
Bristol and Somerset	0.63	0.3	99/1	—
Forest of Dean	0.4	0.2	86/4	—
England	152.53	79.2	—	—
United Kingdom	192.5	100	82/4	1.33

*Approximate costs. †Includes North Wales. ‡Includes South Wales.
Source: National Coal Board.

The open-cast working of near-surface coal deposits begun on a large scale in 1942 proved to be the most economical method of mining coal. By 1958 open-cast production in England was about 11,000,000 tons (U.K. 14,300,000). In the *Revised Plan for Coal* the planned output from this lucrative method of mining was reduced to 2,000,000 tons to avoid the premature closure of deep mines and the consequential distress in mining communities. (*See* also COAL AND COAL MINING.)

Iron.—The annual production of iron ore in Britain averages between 14,000,000 and 18,000,000 tons, almost entirely produced in England. In terms of metal content this output represents 40% of iron ore consumption in Britain.

The iron ores of England are found in four main formations, the Carboniferous Limestones of Cumbria, the Carboniferous Coal Measures, the Liassic and Oolitic rocks of the Jurassic system and the Cretaceous Sandstones of the Weald (no longer worked). Historically the Coal Measure ironstones have been of great importance in influencing the location of metal working. They no longer produce ore, and the Jurassic ores have taken their place as the most important indigenous source of iron. These ironstones, though lean, can, over wide areas, be cheaply mined in open quarries. (*See* also *Iron and Steel Industries* below.)

TABLE VIII.—*Iron Ore Production (England)*

County	Ore	Metal content (1959) (% Fe)	Production (000 tons) 1938	Production (000 tons) 1959	Proportion of U.K. output (1959) (%)
Cumberland	Carboniferous limestone	48	678.7	322.5	2.1
	Hematite				
Leicestershire	Middle Lias	26	878.7	960.7	6.4
	Inferior Oolite	31			
North Lincolnshire	Lower Lias	20	3,438.8	6,120.4	41.1
South Lincolnshire	Middle Lias	26			
Northamptonshire	Middle Lias	23	3,832.2	4,830.0	32.5
	Inferior Oolite	31			
Oxfordshire	Middle Lias	23	532.0	848.9	5.6
Rutland	Middle Oolite	31	530.6	1,243.4	8.3
Staffordshire	Coal Measures	30	127.0	0	0
North Yorkshire	Middle Lias	28	1,516.5	418.2	2.8
England	—	—	11,651.8	14,744.1	99.1
United Kingdom	—	28	11,859.2	14,869.5	100

Source: British Iron and Steel Federation.

5. Power Resources and Utilization.—Despite the great increase in oil imports and the development of hydroelectric and nuclear power, coal remains by far the most important source of energy used in Britain. In England its importance is even greater as the bulk of hydroelectric power is generated in Wales and Scotland. Less coal is burned, however, to provide direct heat and power, as more is being converted into electrical energy and gas. After World War II an increasingly large amount of energy was derived from imported oil used both direct and in thermal electricity generating stations. In 1949 oil contributed 9% to the energy consumed; by the early 1960s the proportion exceeded 20%.

Nearly one-quarter the energy used in Britain is in the form of electricity and the proportion is increasing rapidly. Most of the electricity is produced in coal-fired steam generating stations, an increasing amount from oil-burning plants, while hydroelectric and nuclear power stations make a small but significant contribution. After 1930 the sales of electricity in Britain increased tenfold to more than 100,000,000,000 kw.hr. in the early 1960s. The production and distribution of electricity in England grew up piecemeal in the first decades of the 20th century. In 1926 the Central Electricity board was established to co-ordinate production and consumption. Some power stations were closed, the most efficient being retained, and a system of interconnected transmission lines, the grid, was established, operating at 62,000 and 132,000 volts. During World War II it became necessary to transfer large amounts of current over long distances for which the grid was not suited. After the war, developments in transmission technology made such movements of energy less difficult, and upon the nationalization of the industry in April 1947 it was possible to plan the supergrid, operating at 275,000 v. and connecting large generating stations with the main centres of demand. In the 1960s the new British Electricity authority was engaged in a massive program of power station construction, siting the stations where cheap coal and cooling water were available and where demand could best be met; of particular importance were the coal-fired stations along the Aire, Calder and Trent. The power stations away from coal fields have relied increasingly upon imported fuel oil and many have been designed to deal with the alternative fuels according to their relative costs.

Britain has few hydroelectric resources, almost none in England. This, with the rising cost of coal, has stimulated research into atomic energy as an alternative source. Though conventional thermal power production is becoming more efficient, nuclear power will ultimately be cheaper. The nuclear generator is a thermal station using compact fuel rods whose cost of transport is insignificant. The plant may therefore be put where most convenient: it must have supplies of cooling water; for psychological reasons, it must at present be kept away from large centres of

population. The pioneer power station at Calder Hall in Cumberland began working in 1956, with a capacity of 140 megawatts. Other nuclear projects are at Berkeley in Gloucestershire, Bradwell in Essex, Hinkley point in Somerset and Dungeness in Kent.

The increasing use of oil was a marked feature of postwar power development. Most of the oil consumed is imported as crude oil. The bulk of refinery capacity is in England, the principal plants being the Fawley refinery near Southampton, Shellhaven and Kent refineries on the Thames and the Stanlow plant on the Mersey. The bulk of crude petroleum imports is from the middle east, with Venezuela as second source. More than half the imports of refined products come from the Americas, with Europe as second supplier. The use of pipelines in England is limited and is confined to oil products. However, an oil company was engaged in the construction of lines from the refinery at Fawley, one to carry gas to a chemical plant near Bristol and the other to carry 11 different light fuel oils to a distribution depot at Staines. The government plans for reorganizing transport, published in Dec. 1960, envisaged the construction of pipelines on railway property.

The manufacture of gas for heating, illumination and power has been carried on in England since the first decade of the 19th century. Largely replaced by electricity as an illuminant, gas is still an important source of heat and power. Gas is chiefly made from coal though oil is of increasing importance as a source of combustible gases. The gas industry was nationalized in April 1948 and is administered by the Gas council and 12 area boards, 10 of which serve England.

6. Iron and Steel Industries.—Basic to the engineering industries which contribute so largely to British exports is the manufacture of iron and steels. England produces about two-thirds of British iron and about 60% of British steel. Two of the major raw materials are fully supplied from home measures, the best coking coals being mined in Durham, Yorkshire and south Wales; metallurgical limestone is of wide occurrence in England. English iron ore supplies, however, are insufficient to meet the demand and are supplemented by imported ores which average about 12,000,000,000 tons per year (60% by metal content of ore used). The fourth major raw material, scrap, is of great importance in steelmaking and represents just under half the furnace charge. Scrap largely arises in the main engineering centres of the country and in the steelworks themselves.

The most important district for both iron and steel production in England is in the northeast with Teeside as its main focus. Coming into prominence after the discovery of the Cleveland ironstones in 1850 and using Durham coking coal, this region became the leading British iron producer. With local ore almost exhausted, production is based on imports. Steel is made chiefly by the basic open-hearth process in integrated iron and steel works, and the output is largely of heavy rolled products such as rails and plates.

The second most important district is in Lincolnshire, at Scunthorpe, using local ore and Yorkshire coal. The open-hearth basic steel produced is converted into heavy standardized products such as girders. Farther south along the Jurassic escarpment other ore-field plants occur in south Lincolnshire and Northamptonshire. As at Scunthorpe the ore is mined on the spot and the blast furnaces fired with Yorkshire coal. More pig iron is produced than steel and the excess is transported to the west Midlands for use in forges and foundries. The retained pig iron is converted into basic steel by the Bessemer process which is admirably suited to a non-coal-field location using highly phosphoric ore. The remaining ore-field iron and steel area in England is the west Cumberland and Furness district which was founded on the local nonphosphoric hematite ores; much ore is now imported. The tradition of acid steel production has, however, persisted and the region is the only producer of acid Bessemer steel in Britain.

The iron and steel areas on the coal fields have somewhat different characteristics. The need for fuel saving and hence integration of iron and steel processes is not so pressing; many coal-field sites date from presteel days and were located by the presence of both coal and Coal Measure ironstones. The Coal Measure ores are no longer economic to work and both ore and pig iron are imported. Steelmaking capacity exceeds ironmaking in all coal-field districts, the most important in England being in the west Midlands and Sheffield. The former, a pioneer iron area in the Industrial Revolution, is only of minor significance and specializes in the production of irons for foundry and wrought-iron work. The latter is the leading British producer of high-quality alloy steels. In Sheffield the unit of production is small, the steels are made to exacting specifications and the electric furnace achieves an importance not found elsewhere in Britain. Other steelworks are to be found in south Lancashire and north Staffordshire.

7. Engineering Industries.—Engineering covers a multiplicity of activities and embraces many distinctive industries though their products are generally some form of machine or instrument made from metal. In total these industries are the largest employers of labour in Britain.

Shipbuilding and Marine Engineering.—In the first two decades of the 20th century Britain, the pioneer of the steamship and the iron vessel, was building more than half the world's ships. The interwar years saw depression and unemployment in the shipyards but World War II brought a revival and prosperity and full order books prevailed until the late 1950s when once again orders declined in the face of intense competition, particularly from Japan and the German Federal Republic, and led to surplus shipyard capacity. Of the major British centres three are in England, the most important in the northeast where shipyards flank the Lower Tyne, Wear and Tees, on Merseyside and at Barrow-in-Furness. The northeast coast is the principal centre of marine engineering.

Railway Engineering.—Locomotive and rolling stock are manufactured in two types of works. The railway companies in the 19th century established their own works generally at nodal points in the rail network. These works were taken over by British railways in 1947 upon the nationalization of the railways, the most important in England being at Crewe, Darlington, Derby, Doncaster and Swindon. In addition separate private companies in these localities manufacture locomotives and rolling stock for both the home and the long-established export market. The last steam locomotive to be built for British railways went into service in 1960. The electrification and dieselization programs which were introduced in the 1960s made new demands of the industry, and both British railways and private company works produced the new vehicles. The private works count South Africa, India and Pakistan among their major overseas customers.

Motor Vehicles.—Of particular importance in postwar Britain was the growth of motor vehicle production, 90% of which is in the hands of five leading companies; smaller companies were progressively absorbed by these five firms. The industry is concerned with the assembly of components to form the finished vehicles. These assembly plants are entirely in England, concentrated in a belt from Birmingham to London with Birmingham, Coventry, Cowley, Luton and Dagenham as the main centres. Components and accessories are produced by a large number of manufacturers scattered throughout the country. The output of passenger vehicles doubled after 1945 and that of commercial vehicles increased by 50%. Further expansion during the 1960s changed the locational pattern of the industry. Merseyside, with three new plants scheduled, became a new and major centre of vehicle production due in part to the government encouragement and financial help given to firms establishing plants in areas of high or chronic unemployment. The west Midlands is the main centre for the manufacture of motorcycles, motor-scooters and bicycles.

Aircraft Industry.—Like the motor vehicle industry, aircraft manufacture uses components drawn from a wide variety of sources. Its location tends to be influenced by the availability of skilled labour such as that recruited by motor vehicle manufacturers; indeed several aircraft companies either grew out of or are associated with automobile firms. During World War II the demand for military aircraft led to the establishment of large factories where there was access to skilled labour and sufficient space for assembly hangars and testing airfields. Since the war the resources of the contracted industry have been concentrated into five large companies, two concentrating on engine production, two on airframe assembly and one specializing in the manufacture of helicopters. Originally focused upon Greater London and the

southeast, the wartime dispersion from this more vulnerable area has persisted. In addition to London and the home counties other important aircraft factories are to be found in Coventry, Wolverhampton and Derby in the Midlands, Bristol and Yeovil in the west country and in Manchester, Preston, Leeds and Brough (near Hull) in the north.

Textile Machinery.—The textile industries were the first to become mechanized during the Industrial Revolution with the result that the textile machinery industry is one of the oldest forms of engineering in Britain and the world. Growing up in close contact with the textile mills on the coal fields, the textile machinery factories are located in the yarn- and clothmaking districts. The greatest concentration is in Lancashire, where the cotton industry was the first to adopt machinery; others occur in the Yorkshire wool textile area, the lace and hosiery districts of Northampton and Leicester and, to a lesser extent, in the silk and cotton areas of Cheshire and Derbyshire. The industry produces a wide range of machines to work both natural and synthetic fibres, supplies many parts of the world and has earned a high reputation for its products.

Agricultural Machinery.—Agricultural machinery is manufactured in a large number of works throughout the countryside. Motorized machinery, such as tractors and combine harvesters, is produced mainly by subsidiaries of the motor vehicle companies.

Electrical Engineering and Electronics.—These industries are growing rapidly; the British economy is coming to rely more on highly developed skills and complex processing to produce goods that can compete successfully in the world market.

The electrical engineering industries include the manufacture of massive complex machines, such as electricity generating plant, as well as small domestic appliances. The heavier side of the industry requiring large plant and skilled labour is located in or near mechanical engineering centres from which it derives some components and to which it sends a considerable part of its output. Major centres include Manchester, Tyneside, Stafford and Rugby. The lighter side, manufacturing a wide range of domestic appliances, is more widely distributed, and works occur in most urban areas. The electronics industry, producing radio, television and radar equipment and data-processing machines, expanded rapidly after the war. It supplies the home market and exports components and equipment to many areas, particularly to North America and Australia. The industry is dispersed with a certain emphasis on metropolitan England, together with light engineering, has been attracted to the trading estates and towns of the development areas where old staple industries such as cotton textiles are in decline and where labour and buildings are in supply.

8. Textile Industries.—Though no longer occupying the commanding position they once held, the textile mills and clothing factories of Britain together employ the third largest labour force and contribute about 10% of the value of British exports. These industries are located almost entirely in England.

Wool Textiles.—The West Riding of Yorkshire dominates this industry, accounting for 95% of the worsted and 85% of the woolen cloth produced. This area, favoured by its abundant water supply, its social and economic structure, and, in the 19th century, by its coal, brought about the eclipse of the once famous medieval textile areas. East Anglia is no longer significant, while the Cotswolds woolen district has survived only as a specialist in high-quality woolen cloths in which traditional craftsmanship compensates for the small scale of production.

In Yorkshire, woolen cloths are of particular importance in the Huddersfield and Wakefield districts; worsteds, though widespread, are concentrated in the northwest, with Bradford, Keighley and Bingley as major centres, and farther south in Halifax and Huddersfield. Woolen manufacture is vertically integrated with all processes taking place in one mill; the unit of production is small, and many are still family concerns. In contrast, the worsted firms are larger and the industry is horizontally organized with combing, spinning and weaving processes carried out by separate firms in distinct premises. There is, however, increasing economic integration as companies are formed with interests in all processes, though their physical separation has remained. An important contributor to exports and still the world's largest single consumer of raw wool, the West Riding industry is steadily increasing its consumption of man-made fibres for blending with the natural wool yarns; synthetic fibres account for one-tenth of the fibres used.

Cotton Textiles.—Lancashire dominates cotton to an even greater extent than Yorkshire does wool. More than 90% of English cotton mills are in Lancashire and most of the others are in the adjacent counties of Cheshire, Yorkshire and Derbyshire. A commanding position was assumed as early as 1840 and Lancashire, the first textile area to adopt steam-driven machines, continued to expand until it came to be one of the classic examples of 19th-century industrial specialization. By 1840 cotton goods accounted for almost one-half of Britain's exports. Production increased until World War I; thereafter a decline set in which has persisted. The annual consumption of raw cotton in the 1960s was less than one-third of that in the peak year of 1913. The loss of overseas trade resulting from the expansion of manufacture abroad left the industry with a productive capacity far in excess of the requirements of a home market already invaded by foreign goods and textiles made from artificial fibres. The recession continued after World War II; the labour force declined to less than half the inter-war level and the industry, burdened with surplus capacity, old buildings and old machinery, was, in the 1960s, undergoing a drastic reorganization in an attempt to concentrate capital investment into a smaller number of modernized production units. The geographical specialization in processes and cloths persists. While weaving and spinning are found in most cotton towns, weaving to the exclusion of spinning is much more characteristic of the towns north of Rossendale, such as Nelson, Colne and Darwen, while the bulk of spinning capacity is found in southeast Lancashire. The availability of empty cotton mills and the labour force, largely female, once employed in them has encouraged a wide variety of light industries to be established in Lancashire.

Synthetic Textiles.—Man-made fibres are of great importance in England. Rayon was the first to be introduced. Essentially a chemical process, rayon manufacture uses a cellulose raw material such as wood pulp. It can be produced as a staple fibre and as such can be blended with natural fibres or as continuous filament yarn. Though more widely diffused than the traditional textile industries, the rayon industry became located in the same general area as the established textile districts, with Lancashire as the leading county. The other synthetic fibres have been produced largely since 1948, and both rayon and chemical companies have become involved in their production. Nylon, the first of the truly synthetic fibres, is the most important. In addition to the plant at Pontypool in south Wales where large-scale production first began in 1948, two other factories, at Doncaster and near Gloucester, are in production. Terylene, the British trade-name for a polyester fibre (called Dacron in the United States) developed in Manchester, is produced at Wilton near Middlesbrough. A third, Courtelle, an acrylic fibre, is in quantity production at Grimsby. Ten times more cotton and wool than nylon continues to be used in British cloth manufacture.

9. Textile-Using Industries.—The clothing, carpet, hosiery and knitwear industries are the main users of yarns and cloth. Hosiery and knitwear manufacture is widely distributed, though the long-established district in Nottinghamshire and Leicestershire is still important. Woolen yarns account for half the consumption. The principal carpet manufacturing towns in England are Kidderminster, Halifax, Dewsbury and Durham. Wool makes up 90% of the yarn used, though the tufted carpet industry, with Lancashire as one of its important centres, uses a considerable quantity of synthetic fibres.

The clothing industry is particularly important in England. The unit of production is characteristically small and its locational requirements are such that a wide degree of dispersion is possible; in the years after World War II clothing factories were established in many development areas. Three major foci can be distinguished: Greater London with its prestige value, transport facilities, labour force and the largest single market for consumer goods in Europe, is the main centre; the West Riding, and in particular Leeds, is second in importance, and its specialization in factory-

tailored suits reflects the availability of wool textiles; the third region, centred on Manchester, produces shirts, blouses and dresses, indicating the textile characteristics of Lancashire.

10. Chemical Industries.—The chemical industries are major employers of labour and account for about 9% of British exports. The description, as in engineering, covers a wide range of activities, though production is concentrated into fewer companies than in the engineering trade. The heavy chemical industries produce materials such as acids, alkalis and their salts (sulfuric acid and soda being particularly important) which become the raw materials for a whole range of subsequent manufactures. Large deposits of salt, a major requirement of heavy chemical manufacture, occur in Cheshire and on Teeside and have given rise to the heavy chemical industries near Northwich and Runcorn in Cheshire, Widnes in Lancashire and Billingham on Teeside. The Lancashire-Cheshire area benefits from the facilities of the Ship canal and access to imported raw materials, and from the proximity of major consumers in the textile towns of Lancashire, soap manufacturers on Merseyside and glassworks in St. Helens. The heavy chemical plants on the rivers Severn, Thames and Humber use imported raw materials.

After World War II the output of the chemical industries expanded threefold. This was due in part to the rapid growth of two relatively new branches of the industry, petrochemicals and plastics. England is the largest producer of petrochemicals in Europe. Major plants were built after the war near the oil refineries at Stanlow on the Mersey and Fawley on Southampton water. A third plant at Wilton is near other chemical installations on Teeside, while a fourth centre constructed near Bristol is linked with Fawley by pipeline. Petrochemicals become the raw materials of the detergent, synthetic rubber, plastics and synthetic fibre industries.

The production of plastics quadrupled in England after 1948. Increasingly, plastics are replacing conventional materials for many uses, both domestic and industrial. Of particular importance has been the use of plastic sheet for wrapping and packaging. There has been a rapid expansion in the output of thermosetting plastics, which include polyesters, resins and amino plastics, and of the thermoplastic materials (particularly the polyvinyl chlorides and polystyrene plastics). Medical progress in England leading to the development of many new drugs has been a factor in the post-war expansion of the pharmaceutical chemical industry. Factories are widespread, though Nottingham is a long-established centre. More than a quarter of the output is exported.

11. Glass and Pottery Manufacture.—Drawing upon the chemical industry for part of its raw materials, the glass industry requires in addition silica and fuel for furnaces. The largest glassworks in England, at St. Helens in Lancashire, has access to the soda produced in the mid-Mersey chemical area from salt, to the Shirdley hill sands (an almost pure silica sand deposit covering thousands of acres in southwest Lancashire) and to local coal.

Pottery manufacture is perhaps the most highly localized of all English industries. Almost the entire production comes from the vicinity of Stoke-on-Trent on the north Staffordshire coal field known as "the Potteries." Originally based on local clays, the raw materials, apart from coal, are brought into the area, the kaolin from Cornwall being the chief. Traditional skills are of great importance in maintaining the prosperity of the industry, which enjoys a thriving export trade, in particular of fine chinaware to North America. Derby and Worcester are also pottery centres.

12. Rubber Manufacture.—The world's first rubber factory was established in London in the mid-19th century; 100 years later Britain was second only to the U.S. in rubber production. The growth of the industry has been closely connected with that of motor vehicle manufacture, so that, although the raw rubber is an imported commodity, manufacture has developed in consuming areas, particularly in the Midlands and in Manchester. Tires and tubes compose more than half the output and footwear accounts for one-tenth. Synthetic rubber represents one-third of the total output and is likely to increase. A synthetic rubber plant was set up near Fawley in 1958.

13. Paper Manufacture.—Paper manufacture in England is almost entirely dependent on imports for its major raw material, wood pulp. Only wastepaper and textile materials can be regarded as arising in England. The industry, while widespread, demonstrates two major concentrations, one in Lancashire and Cheshire where import facilities, water supplies, waste textile materials and a large market have favoured growth, the other along the Thames where similar facilities, together with the London market, exist. Output increased by 150% after World War II, and exports to the commonwealth countries are important. (E. S. Sl.)

14. Transport and Communications.—England has a highly developed system of transport by road and rail. Traffic also moves by inland waterway, coastal shipping, pipelines and the airways. Services are provided by a mixture of state, municipal and private enterprise. Public transport facilities are good because of the tradition of public service in an industry which has always subsidized its unremunerative services from its profitable activities.

Road Transport.—The public road network is maintained communally, and tolls are virtually nonexistent. New roads and improvements depend to a varying degree on grants from the central government, but maintenance is the responsibility of local authorities except for about 8,160 mi. which were classified as trunk routes by the Trunk Roads act of 1936 and were transferred to the ministry of transport. Before this the Roads act of 1920 had divided the highways into three categories—class I (which originally included the trunk roads) were main routes with a large proportion of through traffic, class II were the intermediate roads, while the remainder were unclassified. All the class I and class II roads were numbered, with an A prefix for class I and a B for class II, and the numbers appeared on maps and signposts. Between World Wars I and II, main roads in many towns carried tracks for electric tramcars (streetcars), but these survive only in Blackpool. In 1960 M1, the first motorway for high-speed through traffic, was opened between London and the Midlands.

The English road system carries goods and passenger vehicles, the majority being operated privately, but there is a considerable number of public service vehicles. On the passenger side these are provided by nationalized bus undertakings, by independent bus companies and by municipal corporations. Their operations are regulated by traffic commissioners appointed in accordance with the Road Traffic act of 1930. Public freight is conveyed by numerous independent hauliers and by the nationalized undertaking, British Road services, which conducts its business through five companies whose shares are owned by the British Transport commission. Competition is regulated by licensing authorities in accordance with the Road and Rail Traffic act of 1933. All road users pay fuel tax and a vehicle tax, but this is not related to the amount spent on roads. The capacity of parts of the road system has lagged behind the rapid increase in vehicles, and in the early 1960s congestion was severe at peak periods on the trunk roads and in the cities.

Railways.—The railway system of England was developed earlier than any other in the world, and was based on 19th-century conditions. The "high summer" of the British railways was the period immediately before World War I, when large sums of money were invested to complete a system which, within about 20 years, began to undergo a painful process of "pruning." The permanence of railway installations made adjustment to changed conditions particularly difficult, but adaptation and modernization were begun on a large scale in the mid-1950s. It was necessitated not only by the redistribution of population and industry but also by the mechanization of road transport.

With these problems in mind, a Labour government passed the Transport act of 1947 which envisaged an integrated, state-owned transport system under the British Transport commission, with road undertakings being absorbed on a piecemeal basis. When the Conservative government came into power it abandoned the plans for integration in favour of increased commercial freedom. The Transport act of 1953 partially denationalized road freight transport, and, in Dec. 1960, plans were announced to replace the commission by separate boards for different parts of the transport system, including a British Railways board.

In general, railway policy is to electrify the more important

E. A. COURSE

TRANSPORTATION AND COMMUNICATIONS IN ENGLAND AND WALES

main lines and the suburban services and to replace steam by diesels on the minor lines. Country branch lines and some cross-country routes are being abandoned in favour of road transport. Important electrification schemes were completed in the suburbs of London and of Glasgow in 1960; the main-line electrification between London, Birmingham, Manchester and Liverpool was scheduled for completion in 1966. Because of their comparative advantage in handling large-scale movements of freight and passengers, the future of main-line railways of England and Wales would seem to be assured.

Water Transport.—Water transport takes three forms—coastal shipping, river navigation and canals. The basis of the canal system was a diagonal cross with its centre in Birmingham and the Black Country and its arms reaching the rivers Mersey, Trent, Severn and Thames. Partly because of the hilly nature of the country, and partly because the early construction of railways checked canal development, water navigation is less significant in England than in certain other European countries. After 1948 most of the inland waterways (about 2,150 mi.) came under the British Transport commission. Postwar development was concentrated on the more prosperous river navigations, including the river Lee in the London area and the river Trent in the Midlands.

Coastal shipping receives some protection from road and rail competition for strategic reasons. In the past, small sailing vessels linked numerous small ports, but this traffic has passed to the roads. The main lines, such as Newcastle to London, still carry general cargo. However, it is for the transport of bulk cargoes, such as coal and oil products, that coastwise shipping is most significant. Terminal facilities are of great importance, and apart from special installations such as oil jetties, there is a

tendency for traffic to be concentrated at the large ports. Some are controlled by an *ad hoc* authority, such as London and Liverpool; by municipal bodies, as at Bristol; or by the British Transport commission, as at Hull.

Air Lines.—In a country as small as England, the use of air transport for internal traffic is limited. On the busiest routes, such as London to Manchester, a significant proportion of the passengers are completing journeys from other parts of the world. The state corporation, British European airways, has a monopoly of scheduled air services, but licences are granted to a number of independent air lines. Private flying, for business or pleasure, is increasing. As with shipping, the terminal facilities are very significant and airports are costly and, inevitably, remote from city centres. The principal terminals, such as the London airports at Heathrow and Gatwick, are controlled by the ministry of aviation; others, such as Southend, are municipal, and a few are privately owned.

See GREAT BRITAIN; *see* also references under "England" in the Index volume. (E. A. Co.)

BIBLIOGRAPHY.—*General Works:* Current developments and statistics are summarized annually in *Britannica Book of the Year; Britain: an Official Handbook* published annually by the Central Office of Information in co-operation with government departments and many national organizations is one of the best comprehensive reference books on the United Kingdom and contains a select bibliography.

Physical Geography: A. E. Truman, *Geology and Scenery in England and Wales* (1938); L. D. Stamp, *Britain's Structure and Scenery* (1946); J. A. Steers, *The Coastline of England and Wales* (1953); "Regions of the British Isles Series," ed. by W. G. East (1960–); G. Manley, *Climate and the British Scene* (1952); H.M.S.O., *British Regional Geologies, Geological Survey Memoirs;* P. Lake and R. H. Rastall, *Textbook of Geology,* rev. ed. (1941); E. G. Bilham, *The Climate of the British Isles* (1938); R. Kay Gresswell, *The Weather and Climate of the British Isles* (1961); Sir A. G. Tansley, *The British Islands and Their Vegetation,* 2nd. ed. (1946); W. B. Turrill, *British Plant Life* (1948); J. Gilmour and M. Walters, *Wild Flowers* (1954); A. R. Clapham, T. G. Tutin and E. F. Warburg, *Flora of the British Isles,* 2nd ed. (1961); E. Harrison Matthews, *British Mammals* (1952); R. Peterson, G. Mountfort and P. A. D. Hollom, *A Field Guide to the Birds of Britain and Europe,* 3rd rev. ed. (1958); M. Smith, *The British Amphibians and Reptiles* (1951).

All branches of British natural history are covered in the "New Naturalist Series," which contain bibliographies, and in the "Wayside and Woodland Series."

The People: H. J. Fleure, *A Natural History of Man in Britain,* 2nd ed. (1959); Sir Cyril Fox, *The Personality of Britain,* 4th ed. (1943); H. M. Chadwick, *The Origin of the English Nation* (1907); V. Gordon Childe, *Pre-Historic Communities of the British Isles,* 3rd ed. (1949); Stuart Piggott, *The Neolithic Cultures of the British Isles* (1954); J. L. and Barbara Hammond, *The Village Labourer,* 2nd ed. (1920); E. T. Leeds, *The Archaeology of the Anglo-Saxon Settlements* (1913); Sir Halford John Mackinder, *Britain and the British Seas,* 2nd ed. (1930); J. G. D. Clark, *Prehistoric England,* 4th ed. (1948); R. H. Hodgkin, *A History of the Anglo-Saxons,* 3rd ed., 2 vol. (1952); Sir Frank Stenton, *Anglo-Saxon England,* 2nd ed. (1947).

Archaeology: D. A. E. Garrod, *The Upper Palaeolithic Age in Britain* (1926) and "The Upper Palaeolithic in the Light of Recent Discovery," *Proceedings of the Prehistoric Society* (1938); J. G. D. Clark, *The Mesolithic Age in Britain* (1932), and "A Microlithic Industry From the Cambridgeshire Fenland and Other Industries of Sauveterrian Affinities From Britain," *Proceedings of the Prehistoric Society* (1955); S. Piggott, *The Neolithic Cultures of the British Isles* (1954); V. G. Childe, *Prehistoric Communities of the British Isles,* 4th ed. (1956); S. S. Frere (ed.), *Problems of the Iron Age in Southern Britain* (1961).

Population: B. Benjamin, *Elements of Vital Statistics* (1959); A. Carr Saunders *et al., Survey of Social Conditions in England and Wales as Illustrated by Statistics* (1958); V. G. Childe, *Prehistoric Communities of the British Isles,* 4th ed. (1956); H. C. Darby, *An Historical Geography Before 1800* (1948); T. W. Freeman, *The Conurbations of Great Britain* (1959); L. G. Johnson, *The Social Evolution of Industrial Britain* (1959); D. C. Marsh, *The Changing Social Structure of England and Wales* (1958); J. Saville, *Rural Depopulation in England and Wales. See* also publications issued by the General Register office.

Administration and Social Conditions: J. J. Clarke, *Outlines of Local Government of the United Kingdom,* 18th ed. (1957); W. E. Jackson, *The Structure of Local Government in England and Wales,* 3rd ed. (1957); Central Office of Information Reference pamphlet, *Local Government in Britain* (1959).

P. Gregg, *A Social and Economic History of Britain 1760–1950* (1950); *The Census Reports for England and Wales* (1951); A. Flanders and H. A. Clegg (eds.), *The System of Industrial Relations in Britain* (1954); D. V. Glass (ed.), *Social Mobility in Britain* (1954); Central Office of Information Reference pamphlet, *Housing in Britain* (1954); G. D. H. Cole, *The Post-War Conditions of Britain* (1956); A. Ashworth, *Housing in Great Britain* (1957); Central Office of Information Reference pamphlet, *Labour Relations and Working Conditions in Britain* (1958); A. M. Carr-Saunders *et al., A Survey of Social Conditions in England and Wales* (1958).

Central Office of Information Reference pamphlet, *Fire Services in Britain* (1955).

M. P. Hall, *The Social Services of Modern England,* 4th ed. (1959); Central Office of Information Reference pamphlet, *Social Services in Britain* (1959, 1960) and *Youth Services in Britain* (1959); H. Eckstein, *The English Health Service* (1959); Central Office of Information Reference pamphlet, *Health Services in Britain* (1959); Ministry of Pensions and National Insurance, *Everybody's Guide* (1958) and Cmd. 782, *Improvements in National Insurance* (1959); Central Office of Information Reference pamphlet, *Children in Britain* (1959).

C. C. H. Moriarty, *Police Procedure and Administration,* 6th ed. (1955); Central Office of Information Reference pamphlet, *The Police Service in Britain* (1959).

Sir C. K. Allen, *Administrative Jurisdiction* (1956); W. A. Elkin, *The English Penal System* (1957); R. Rubinstein, *John Citizen and the Law,* 4th ed. (1958); Home Office Parliamentary Paper, Cmd. 645, *Penal Practice in a Changing Society,* 4th ed. (1959); Home Office publication, *Prisons and Borstals* (1960); J. F. S. King (ed.), *The Probation Service* (1952).

S. J. Curtis, *History of Education in Great Britain,* 4th ed. (1957); W. O. L. Smith, *Education in Great Britain,* 3rd ed. (1958); Central Office of Information Reference pamphlet, *Education in Britain* (1958).

The Economy: W. Smith, *An Economic Geography of Great Britain,* 2nd ed. (1953); L. D. Stamp and S. H. Beaver, *The British Isles,* 4th ed. (1954); L. D. Stamp, *The Land of Britain,* 2nd ed. (1950); H. I. Moore, *Background to Farming* (1947); K. A. H. Murray, *Official Histories of the Second World War: United Kingdom: Civil Series. Agriculture* (1956); E. L. Crossley, *The United Kingdom Dairy Industry* (1959); H. L. Edlin, *England's Forests* (1958); H.M.S.O., Natural Resources Technical Committee, *Report on Forestry, Agriculture and Marginal Land* (1957); G. P. Wibberley, *Agriculture and Urban Growth* (1959); H.M.S.O., *Forest Commissioners Annual Reports; Fisheries Year Book and Directory* (1960); E. C. Allen, *British Industries and Their Organization,* 4th ed. (1959); W. A. Robson, *Nationalised Industry and Public Ownership* (1960); National Coal Board, *Plan for Coal* (1950), *Investing in Coal* (1956), *Revised Plan for Coal* (1959); R. Robson, *The Cotton Industry* (1957), *The Man-Made Fibres Industry* (1958); G. Maxcy and A. Silberston, *The Motor Industry* (1959); Sir Henry Self and E. M. Watson, *Electricity Supply in Great Britain* (1952); Ministry of Power, *Statistical Digest* (annual); H.M.S.O., *Annual Abstract of Statistics.*

ENGLAND, CHURCH OF. Members of the Church of England at the Reformation and afterward have commonly seen themselves as the heirs and successors of the Anglo-Saxon and medieval English church. They have looked back as much to Bede and Pope Gregory the Great as to Henry VIII and Archbishop Thomas Cranmer. They have not regarded themselves only as a conservative branch of the European Reformation but as the national church, the church of the English people since the time that England received the Christian faith. They have thus possessed a sense of continuity with the medieval church more lively than that found in other Protestant churches. Though they have usually held the fundamentals of Reformation doctrine, such as the proposition that "all things necessary for salvation" are in the Bible, they have valued membership of a church which preserved more of the traditional framework of medieval Catholicism than any other reformed church. Their services, although in the vernacular, are liturgical in form and convey a traditional atmosphere, their bishops are still consecrated in the medieval manner and the administration preserves much of medieval custom and practice.

The historical accidents of the 16th century compelled the Church of England to tolerate a wide variety of opinion. Subsequently England, especially during the Civil War and its aftermath, was deeply divided, and the Church of England became the church of a majority of Englishmen. Nevertheless, most Anglican churchmen have continued to think of the Church of England as the national church, and have valued its position as the established church to which the sovereign must belong. They have valued their position as a "reformed" church, based on the principles of the Reformation, and as a "catholic" church in the sense of being the authentic representative of ancient Christianity in the country. In the 20th century this consciousness of an inheritance at once reformed and Catholic has led some Anglicans to claim a unique place for their church in healing the divisions of Christendom, by providing a possible bridge over the chasm between Protestant and Roman Catholic.

This article is divided as follows:

I. Pre-Reformation History
1. Origins
2. Anglo-Saxon Church
3. Medieval Church
II. The Reformation
1. Abolition of Papal Jurisdiction
2. Dissolution of the Monasteries
3. Coming of Protestantism
III. From Elizabeth I to the Oxford Movement
1. Elizabethan Church Settlement and the 17th-Century Conservatives
2. 18th and 19th Centuries
3. Quest for Autonomy
4. The Church and the New Scholarship
IV. 20th-Century History and Development
1. Anglo-Catholicism
2. Liberalism or Modernism
3. Constitutional Changes
4. Prayer Book Revision
5. Relations With Other Churches
6. Foreign Missions
7. Establishment
8. Church and Education
9. Clergy
10. Church Commissioners
11. War Damage
12. Canon Law Revision
13. Membership
14. Mid-20th Century

I. PRE-REFORMATION HISTORY

1. Origins.—Christianity reached Britain under the Romans, and a very few archaeological remains of Roman-British Christianity have survived. British bishops from London, York and Lincoln attended the Council of Arles in 314, and three British bishops who attended the Council of Rimini in 359 were so poor that they needed to claim expenses from the imperial treasury.

The Anglo-Saxon invasions divided British Christians from their European source and drove them westward to Glastonbury, St. Davids, Ireland, Whithorn in Galloway, Brittany and the Scilly Isles. The Celtic church (*q.v.*) became in time strong enough to undertake missionary and monastic expansion eastward again, to Scotland, where St. Columba (*q.v.*) was at Iona from *c.* 563, and to France. But no impression was made upon the heathen Saxons of England until after the arrival of St. Augustine. In 597 Pope Gregory I the Great sent Augustine to Kent, where the king, Aethelberht, was married to Berhta, a Christian princess from Gaul, and during the 90 following years all the Saxon kingdoms of England gradually accepted the faith. The missionary endeavour was led by Celts from Iona, such as St. Aidan (*q.v.*) who worked in Northumbria, but it was Roman discipline, particularly inspired by Theodore, archbishop of Canterbury from 668 to 690, that organized the different kingdoms into a single church order and thus helped to create the unity of England. It is an old epigram that in England the state is, in a sense, the child of the church.

Pope Gregory had designed an elaborate plan for his new church. There were to be two independent provinces, each with 12 dioceses. The two archbishops were to hold their sees at the two chief Roman towns, London and York, and precedence should belong to the senior archbishop. This arrangement could not be carried out exactly, but it profoundly influenced the subsequent history of church order in England. St. Augustine's successor St. Laurentius obediently set up his see at London, but was ejected by a heathen reaction and retired to Canterbury; and thus the provincial sees have remained at Canterbury and at York, which was made an archbishopric in 735. King Offa of Mercia created (788) an archbishopric at Lichfield but it collapsed in 802. The claim to precedence made by the archbishops of Canterbury was long contested by York, as is shown in the unusual titles still borne by the two archbishops, Canterbury being "primate of all England" while York is "primate of England." The province of York always sought to guard the independence of Canterbury which Gregory had planned. The plan for 24 dioceses, making 26 bishops in all, was long in fulfillment. Theodore did important work in creating more dioceses, and the kings after the Norman conquest created more. There were 22 dioceses before the Reformation, Gregory's original number 26 was first achieved by Henry VIII in 1541 and the number in 1961 (exclusive of Wales) was 43.

2. Anglo-Saxon Church.—In the first century of Saxon Christianity there was a conflict between Celtic customs and Roman customs, resolved in favour of the Roman at the synod of Whitby (663–664). Then and thereafter Saxon Christians looked to Rome as the European centre of their faith, a place of pilgrimage and a court to which disputes might be carried. But the relation between the pope and the English church was less close than it subsequently became under the later Normans, partly because the nature of papal authority was less well established until the 12th century, and partly because the Danish invasions of the 9th to the 11th centuries again severed communications between England and the continent.

The stimulus derived from the communications and even the disputes between Roman and Celt in England contributed to a remarkable flowering of life and scholarship. St. Chad (d. 672), who had been a pupil of St. Aidan at the great monastery of Lindisfarne on Holy Island, founded the see of Lichfield and spread the faith in Mercia. His brother Cedd worked in Essex. St. Wilfrid (d. 709) devoted his life to establishing the Roman order over the Celtic and fought a stormy series of controversies over discipline and precedent. St. Cuthbert, after living as a hermit on the Farne Islands, became bishop of Lindisfarne (685). The most famous churchman of the age was perhaps the Venerable Bede, a monk at first of the monastery of Wearmouth and later of Jarrow after its foundation (681–682). He, like Wilfrid, was devoted to the establishment of the Roman discipline, and even his great history was in part written with this in mind. He was soon an outstanding figure in the tradition of European scholarship and divinity (*see* BEDE).

During the age of Bede and the remaining years of the 8th century relations between England and the continent were close, English churchmen working abroad as scholars, reformers and missionaries. St. Willibrord began to evangelize the Frisians in 690. St. Boniface was working in Bavaria and Thuringia from 719 onward and, with the support of the pope and of the Frankish kings, laid the foundations of the church in Germany and reformed the Frankish church. Alcuin (d. 804), educated at York, became director of Charlemagne's palace school at Aachen and represented the strong impetus given by English learning and churchmanship to the Carolingian renascence.

The Danish invasions destroyed monasteries and weakened learning, and it was not until the successful campaigns of King Alfred (*q.v.*) (d. 899) and of his son Edward the Elder (d. 924) began to pave the way for the establishment of political unity under the Wessex kings that a reform of the church could be undertaken. Alfred was concerned for the conversion of the Danish settlers. Administrative reorganization of the church was begun in his time, and although his contemporaries seem to have shown scant enthusiasm for monastic revival he founded a monastery at Athelney and a nunnery at Shaftesbury. He also caused important works of divinity to be translated into the vernacular. Under his successors, particularly under Edgar (d. 975), considerable reforms were effected. Influenced by continental reform at Cluny and in Lorraine, St. Dunstan (*q.v.*), archbishop of Canterbury, St. Oswald, archbishop of York, and St. Aethelwold, bishop of Winchester, did much to re-establish order in the existing monasteries and founded various new houses. Many English cathedrals were gradually brought under the control of monastic chapters and from this time until the Norman conquest most of the bishops were chosen from among monks. Religious houses founded or refounded at this time included Glastonbury (where Dunstan had been abbot), Bath, Exeter, Ely and Peterborough. Westminster abbey was built by Edward the Confessor (d. 1066).

3. Medieval Church.—The Norman conquest united England more closely with the culture and organization of Latin Europe. Under William the Conqueror and his successors, archbishops like Lanfranc (*q.v.*) undertook the reform of the English church according to Roman ideas, reviving local synods, insisting upon the celibacy of the clergy, and introducing the canon law of western Europe. The canons ensured the separation between civil courts (already effected by an ordinance of William I) and the lawfulness

of appeals to Rome. Everywhere in western Europe after the mid-11th century the papal jurisdiction was growing in effectiveness, and it was exercised no less fully in England. The archbishops of Canterbury, to ensure their own authority, found it convenient or necessary to claim that they were papal legates by virtue of their office; a claim the popes were careful to control, by insisting that the legateship was their gift and was not *ex officio*, by bestowing the legateship also upon archbishops of York, or by sending their own envoys as legates to England with special powers. (Certain of the duties of the legateship remained to the archbishops of Canterbury after the abolition of papal authority, by acts of Henry VIII; *e.g.*, the right to issue special licences for marriage and the right to confer university degrees; *see* DISPENSATION.)

To the kings of England it was always important, if not to control, at least to limit this connection between Rome and the high officers of the church. If the bishops, who were also tenants in chief, became too powerful and independent, government of the state would be impossible. The struggles of Archbishop Anselm against William Rufus and Henry I (*see* ANSELM, SAINT; HENRY I), and of Thomas Becket against Henry II (*qq.v.*), were in part struggles for a minimum justice and freedom for the church and in part struggles for an ecclesiastical independence that no king could allow if he wished to retain an effective government. The terms imposed in reparation for the murder of Becket in Canterbury cathedral weakened the crown in face of the growing administrative power of the papal curia; and although John's surrender (1213) of his kingdom as a fief of the papacy brought him temporary advantages, the feudal relationship laid England particularly open to the rapacity of papal tax collectors during the 13th century. The crown thereafter established a *modus vivendi* by keeping a measure of control over the appointments to bishoprics, which became a matter for agreement between king and pope. Further papal claims to appoint or "provide" to future sees or benefices encountered a popular belief that these claims were intended simply as a means of surreptitious gain on the part of the papacy, and were legally forbidden, in 1351, 1365 and 1390, by the statutes of Provisors (*see* PROVISION). But the system was useful to the king as well as the pope, and the statutes were often evaded. By the later 15th century, with papal complaisance, the crown had secured an effective control over the appointments to all bishoprics. When in 1534 Henry VIII abolished the papal right in the appointment of bishoprics and retained the crown's authority therein, he was giving legal effect to what was in substance current practice.

At the same period as the statutes of Provisors, when the authority of the crown was being strengthened and the public reputation of the papacy was declining, the statutes of Praemunire (*q.v.*) were passed in 1353, 1365 and 1393. These statutes limited appeals from English courts to Rome and forbade the introduction from Rome of bulls or excommunications, under penalty of forfeiture of goods and imprisonment at the king's pleasure. Praemunire was to be invoked by Henry VIII against Thomas Cardinal Wolsey and later against the clergy of the Church of England, as the legal instrument for securing a general control by the crown and a preliminary to the abolition of papal authority.

During the middle ages English clergy and laity made important contributions to the life and activities of the western church. One Englishman became pope, as Adrian IV (1154–59); English contingents went on the crusades; the Cistercians, Cluniacs, Carthusians and Premonstratensians went to England, and the whole country was profoundly influenced by the arrival of the friars early in the 13th century. Among English saints were Gilbert (*q.v.*) of Sempringham, founder of the English order of Gilbertines, and Hugh (*q.v.*), bishop of Lincoln. The foundation of Oxford and Cambridge brought England into the stream of university development, and English scholars such as Robert Grosseteste and philosophers such as Thomas Bradwardine and William Ockham achieved a more than national repute. In the 14th and 15th centuries the English church produced a rare group of mystics; among the most famous were Richard Rolle of Hampole, Walter Hilton, Julian of Norwich and the author of *The Cloud of Unknowing*.

The English church had its full share of the abuses and unrest characteristic of the later middle ages. John Wycliffe (*q.v.*) became a revolutionary critic of the papacy and of some accepted doctrines such as transubstantiation. His followers, the Lollards (*q.v.*), were driven underground, but some at least persisted into the 16th century, and in Bohemia the Hussite movement drew much upon Wycliffe's ideas.

The local organization of the English church, as developed during the later middle ages, survived the Reformation in some respects. The cathedrals and their chapters remained, although the monastic chapters were given new statutes when Henry VIII dissolved the monasteries; these became the cathedrals of the "new foundation." Under the medieval system the dean and chapter elected the bishop, though except in rare cases this election had become a mere form (*see* CONGÉ D'ELIRE). The church had its provincial councils, the convocations of Canterbury and York which met partly for the purpose of providing representation for the clergy and of giving recognition of their right to tax themselves and partly for local registration and application of the general canon law of the church. The parochial system was effectively established in Norman times and soon fell so much under the control of local lords that parish clergy were often appointed by a lay landowner. The assimilation of the benefice to secular property, by reason of its origin, placed the resulting legal disputes in the royal or secular courts, not in the ecclesiastical. The advowson or right to appoint to the benefice became a piece of property that could be sold or exchanged or bequeathed; and the incumbent possessed a freehold in his benefice ("the parson's freehold")—that is, it was his property provided that he performed the duty.

This system of parochial benefices was left largely untouched by the Reformation, and despite many subsequent modifications is still responsible for the fact that the parochial life of the Church of England is in some respects unique in Christendom.

II. THE REFORMATION

1. Abolition of Papal Jurisdiction.—The papal jurisdiction was abolished by Henry VIII and his parliament in the Act in Restraint of Appeals (1533) and the Supremacy act (1534), and by further acts that prohibited the payment of papal taxation. Some of the necessary functions that medieval law had ascribed to the court of Rome (*e.g.*, dispensations) were allotted to the archbishop of Canterbury; but the effect was to make the crown "supreme head in earth" of the church, as Henry VIII wished, or "supreme governor" as Elizabeth I and her successors, in deference to strong feelings against the word "head," have been proclaimed.

The chief constitutional effect of the royal supremacy was to end the medieval conflict of jurisdiction between two systems of law, the law of the land and the canon law of the church; and at the same time, if not to end at least to diminish the conflict between ecclesiastical courts and secular courts by making the ultimate court of appeal in both cases the sovereign or judges deputed by him. The royal supremacy made less difference, except to the lawyers, than might at first sight appear; for the right to nominate to bishoprics or the possession of much crown patronage in the appointment to benefices or other ecclesiastical offices had already been achieved effectively in the later middle ages, an achievement which was one reason why papal authority was so easily abolished in England. Convocation, it was true, could no longer enact canons without leave of the crown; but previously the canons of convocation had been subject to the ultimate authority of Rome and an unrestricted right to make canons would rapidly have created anarchy in the legal system and a much more revolutionary Reformation.

It was never claimed, even by Henry VIII, that the royal supremacy gave authority "in the ministration of spiritual things" to a lay sovereign. But the supremacy was a grief to some consciences and helped to produce a number of recusants from the Church of England, both under Henry VIII (Sir Thomas More, Bishop John Fisher and others) and throughout the reign of Elizabeth I. Thereafter this supremacy appeared to be an essential element in the establishment of the Church of England as the national church, and was valued as part of that total relation be-

tween crown and church which included the right of the bishops to sit in the house of lords or the right of the archbishop of Canterbury (acquired in Anglo-Saxon times) to crown the sovereign.

2. Dissolution of the Monasteries.—Between 1535 and 1540 all the monasteries of England dissolved themselves or were suppressed, under the agency of Henry VIII's minister Thomas Cromwell. Their buildings were bought by, or granted to, landowners, their churches sometimes converted into parish churches (or, as with Peterborough, into the cathedral of a newly created diocese); and in the deep country the unwanted buildings, such as those of Fountains, Rievaulx and Tintern abbeys, moldered into ruins. Henry VIII had suppressed them because he thought them useless and wanted their lands and money. The Reformers taught that monks were not only useless but wrong; and so, after a brief interlude of attempted revival under Mary I, the ideals the monks represented were not renewed within the Church of England until the time of the Oxford Movement of the 19th century, and then in a very different context. (*See* OXFORD MOVEMENT.)

3. Coming of Protestantism.—Henry VIII had no intention of countenancing Protestant doctrine during his lifetime. But to throw off the pope's jurisdiction was to encourage all those Englishmen who sympathized with Martin Luther and the Reformers of Europe who were then engaged in seriously questioning the accepted doctrines of the church. In 1539, by the Act of Six Articles, Henry attempted to stamp upon those who advocated marriage of the clergy or denied transubstantiation. But on his death in 1547 a regency for the young Edward VI under his uncle, the duke of Somerset, came into power with a strongly Protestant policy. Two new vernacular liturgies, largely compiled by Archbishop Thomas Cranmer, were provided successively in the Prayer Books of 1549 and 1552, many churches were stripped of statues and their altars converted into holy tables, and the universities, which had been in some peril at the time of the dissolution of the monasteries, were purged of traditionalists and given reforming professors of divinity. The boy king's death in 1553 prevented these changes from establishing themselves deeply in the affections of the people. But during the two succeeding reigns the Protestant ideas became ever more deeply embedded in the minds of the majority of English people, and slowly stamped the English character and the church with a mark it has never lost. Mary I (1553–58) was fanatical in her attempt to repress the Protestants; and the cruelty of the burnings, together with the heroism of those who died (including Cranmer himself, bishops John Hooper, Nicholas Ridley, Hugh Latimer and Robert Ferrar) evoked a wave of popular sympathy and respect.

When Elizabeth I became queen in 1558 many English Protestants who had sought refuge upon the continent came flooding back; and the fight against Catholic Spain which culminated in the Spanish Armada of 1588, the knowledge that Catholic plotters were conspiring against the queen's life, and the excommunication of the queen by Pope Pius V in 1570 finally identified Protestantism with patriotism in the minds of many Englishmen. John Foxe's *Book of Martyrs*, first published in English in 1563, was chained to the desks of parish churches and kept alive the memory of Protestant courage and Catholic persecution. It is impossible to understand the Church of England without recognizing the importance of this epoch in which England's survival seemed to depend upon resistance to Roman Catholicism.

After some temporary Lutheran influence under Henry VIII, the English Reformers looked to Switzerland, and especially to Zürich and the successors of Huldreich Zwingli (*q.v.*), as their theological and practical guides. The Reformation stood for a raising of the educational standard of clergy and people by laying a far greater emphasis on preaching and teaching; for a Bible and for church services in the language of the people; for simplicity in ceremonial and ornament; and above all for a scriptural faith, with the emphasis upon Pauline theology and especially the doctrine of justification by faith alone. A succession of English translations of the Bible, from William Tyndale's (begun 1525) onward, culminated in the majestic simplicity of the Authorized or King James Version of 1611 (*see* BIBLE, TRANSLATIONS OF: *English*). Edward VI's second Prayer Book of 1552, of which Cranmer had been the architect, was restored in 1559 with a few variations intended to render it more acceptable to conservatives, and this was the book which, with a few more alterations mainly in a conservative direction, was re-established in 1662 and remains the only lawful Prayer Book of the Church of England (*see* COMMON PRAYER, BOOK OF). Just as all the reformed churches established formularies for the control and guidance of doctrinal teaching, there were formally approved, by both parliament and convocation, the Thirty-Nine Articles of 1571. These are a revision of earlier articles first published in 1553, and, with the addition of a preface by Charles I in 1628, remain the only formal dogmatic statement approved by the Church of England. All ordained clergy are required to subscribe to them. The most important of the doctrines they seek to impose is the doctrine of justification by faith, to which several articles are directly or indirectly related. They also declare that all things necessary for salvation are to be found in the Scriptures.

The official formularies of the Church of England are thus to be found (1) in the Thirty-Nine Articles of 1571, and (2) in the Book of Common Prayer, 1662.

III. FROM ELIZABETH I TO THE OXFORD MOVEMENT

1. Elizabethan Church Settlement and the 17th-Century Conservatives.—The English Reformers began as disciples of radical Switzerland. By the time of the Restoration of Charles II (1660) the Church of England had become the most conservative and traditional of all the churches that sprang out of the Protestant movement.

Elizabeth I herself was no radical. She wished for due decorum and would perhaps have preferred less change than in fact occurred. Moreover, the seesaw of authorized religion from the Protestantism of Edward VI to the Catholicism of Mary I and back had left the country uncertain and divided, and it seemed necessary for the government to pursue a moderate course in order to unify the country in one church. The Thirty-Nine Articles were not framed rigorously and exclusively, but allowed room for different interpretations. The Prayer Book of 1559 contained far more of the medieval rites than could be found elsewhere in Protestantism, and the government ensured that the archbishop of Canterbury, Matthew Parker, was duly consecrated by four other bishops. At first English churchmen were too intent upon reform to value these traditions. But the more they were criticized by the Protestants who followed John Calvin, the more justifiable and valuable they seemed to be. From 1593 onward Richard Hooker (*q.v.*) was publishing the classical defense of this moderate conservatism in liturgy and church practice, *Of the Laws of Ecclesiastical Polity;* and in the reigns of James I and Charles I this right wing of the Reformation flowered in devotional prose and poetry of the highest order in the work of John Donne and George Herbert. It provided some English churchmen like Lancelot Andrewes and John Cosin with a strong sense of Christian history and a lively apprehension of the treasury of Christian devotion. It represented the best side of Archbishop William Laud's quest for proper order and "the beauty of holiness" in worship, a quest never forgotten in subsequent Anglican tradition.

The Civil War and Commonwealth (1642–60) pushed this conservatism a stage further. The attempt of the Elizabethan settlement to allow room for men of differing opinions meant that by the accession of Charles I in 1625 the church was divided between Puritans and Calvinists on the one side, moderate conservatives and anti-Calvinists (Arminians) on the other. Unfortunately for the church, the argument between Calvinist and Arminian came in the popular mind to overlap with the struggle between the king and his parliament. The victory of parliament and then of Oliver Cromwell involved not only the executions of Archbishop Laud and Charles I but the legal destruction of all the conservative elements of church organization such as bishops, liturgy and ceremonial and their partial replacement by Calvinist formularies and order. Thus the restoration of Charles II naturally identified anti-Calvinism with loyalty to the crown, and imparted a new division, a new intransigence, a new rigidity, by making the former right wing seem the only authentic members of the Church of England. In 1662 ministers who refused to be reordained by the restored

bishops were evicted from their benefices, and modern Nonconformity (as opposed to small separatist congregations) dates from this epoch.

During the 17th century the Church of England produced many thinkers, such as Henry Hammond, Jeremy Taylor and George Bull, who after Hooker came to be regarded as the classical Anglican theologians, often referred to as the Caroline divines. Their work represented a rare flowering of divinity: metaphysical, dogmatic and pastoral. At the same time another school, reacting against Calvinism and against strife over religion, began to appeal to the place of reason in theology and to turn it away from controversy toward metaphysical speculation. They were known by their opponents as Latitudinarians, but their finest representatives were the Cambridge Platonists (*q.v.*), especially Benjamin Whichcote, John Smith, Ralph Cudworth and Nathanael Culverwel. Thereafter, under the stimulus of John Locke, English religious thought left Platonism behind and turned toward the rational philosophy of religion which became characteristic of the 18th century.

The accession of the Roman Catholic James II in 1685 soon alarmed and aroused English Protestant sentiment. The memory of the Civil War had identified loyal churchmanship with loyalty to the crown, and the widely accepted doctrine of passive obedience caused much heartsearching as resistance began to grow. William Sancroft, archbishop of Canterbury, and six other bishops, who protested against a royal order of 1688 that a declaration of liberty of conscience (the second Declaration of Indulgence) should be read in all churches, were put on trial and acquitted amid scenes of popular triumph. At the expulsion of James II and the accession of William III and Mary the Protestant succession was established by law in the Bill of Rights (1689), and in the Act of Settlement (1701) it was enacted that the sovereign must be a member of the Church of England. But the church suffered severely in the loss of those clergy, among them Archbishop Sancroft and some other bishops, who held the doctrine of passive obedience so rigidly and who believed that their oaths of allegiance to James precluded them from holding office under William III. A few of these nonjuring bishops consecrated others and the schism was perpetuated until early in the 19th century (*see* NONJURORS). More important than schism was the loss to the Church of England of good and learned leaders and the slur of worldliness, which by comparison seemed to be cast upon those who had accepted the new king. Meanwhile the Toleration act of 1689 at last allowed freedom of worship to Protestant dissent.

2. 18th and 19th Centuries.—The Church of England during the 18th century was strong in intelligence and learning, and Joseph Butler was only the most eminent of a line of religious philosophers. But the church was tinged with a measure of unashamed worldliness, and the evangelical movement led by John Wesley and George Whitefield (*see* METHODISM) was a reaction both against its concentration upon the head to the exclusion of the heart and against its cool attitude toward religious enthusiasm. The parochial system inherited from the middle ages was ill-fitted to meet the Industrial Revolution and the rapid growth of the new cities. Though Methodism had begun within the Church of England, it ultimately increased the number of Englishmen who did not own membership of the national church. The worst civil disabilities of Protestant dissenters (apart from the inability to be the sovereign) were ended in 1828. Meanwhile the Evangelicals who, like Charles Simeon at Cambridge and the Clapham sect (*q.v.*), remained within the church of England were beginning to exercise a powerful influence in the raising of the pastoral standards of the clergy and in recalling the church to the doctrines of the Reformation and to the biblical faith. They never ceased thereafter to form an important and influential body among the members of the Church of England, their most notable work in the 19th century being performed in foreign evangelism through the agency of the Church Missionary society, founded in 1799.

In 1801 England was united politically with Ireland, and the consequence for the Church of England was drastic, since the mainly Anglican parliament was made directly responsible for several million Roman Catholics. In 1829 the Roman Catholics were freed from most of their civil disabilities, and after the Reform act of 1832 parliament was soon forced into an impartial attitude toward the denominations of the country; thus the relations between the Church of England and the state were changed from that unity of which Hooker had been the theorist. The English need for cheap labour and the Irish famine and poverty created a wave of Irish immigration into England and largely increased the number of English citizens who were Roman Catholic. The last attempt by parliament to discriminate against the Roman Catholic (or any other) church was in the Ecclesiastical Titles bill of 1851. The admission of an avowed freethinker, Charles Bradlaugh, to the house of commons in 1886 completed this process of change.

3. Quest for Autonomy.—The emancipation of the Dissenters and Roman Catholics in 1828–29 and the later growth in the non-Anglican population profoundly affected the Church of England, apart from the change in the relations of church and state. Between 1836 and 1840 parliament reformed some parts of the ancient constitution which had existed since the middle ages—sinecures, pluralities, nonresidence—and above all created the Ecclesiastical commission (1836) to administer the consequent funds for the benefit of impoverished clergy. Queen Anne in 1703 had relinquished the crown right to the first fruits and annates (*q.v.*) and Queen Anne's bounty had been founded as a fund to help poor clergy. There were thus from 1836 two organizations, created by government, to administer these central funds, until they were amalgamated in 1947. (*See* CHURCH COMMISSIONERS.)

Realization that the church and state were now far from coterminous also made for a change of attitude. The royal supremacy, now effectively exercised through the agency of a parliament that was mainly but not necessarily Anglican, became rapidly a matter of resentment among some members of the church. The chief criticism was leveled at the court of final appeal in ecclesiastical causes, which since 1833 has been the judicial committee of the privy council. A growing body of churchmen contended, especially after the case of George Cornelius Gorham had drawn attention to this function of the privy council in 1850, that a committee of theologically untrained lawyers was unsuitable as a final court in cases involving doctrine or ritual. But every suggestion for reform encountered the difficulty that, whatever the disadvantages, legal experience is even more important than theological equipment.

Further, it was clear after 1829 that parliament was no longer fully suitable as the hub of church government. The convocations had for many years been effectively suppressed. After a period of flowering under the first two Stuarts, they had lost their original *raison d'être* in 1664 when Gilbert Sheldon, archbishop of Canterbury, conceded to parliament the right of taxing the clergy. After the revolution of 1688 the lower house of convocation vociferously opposed the Whig regime, and in 1717 convocation was prorogued and had not been thereafter permitted to meet for effective discussion. The legislation of 1828–29 aroused a demand for the revival of convocation. Amid many suspicions the Convocation of Canterbury began to meet again for discussion in 1852, the Convocation of York in 1861. From time to time they have received the licence of the crown to make canons, but hitherto always with the object of bringing the canons of the church into line with acts of parliament. Convocation was hampered in its revival first by the historical accident which made two convocations at York and Canterbury and secondly by the purely clerical nature of the assembly which failed to command the confidence of parliament or lay churchmen. In 1886 an (unofficial) house of laymen was created for the province of Canterbury, and in 1892 a similar one for the province of York. But this lay representation was not given legal status until the Enabling act (Church of England Assembly [Powers] act) of 1919. The revival of convocation, and afterward the creation of the Church assembly, were of much relief and service in meeting the needs created by the change in church-state relations of 1828–29.

The Oxford Movement of 1833 onward was in one of its aspects a movement of "high" churchmanship in the sense of stressing the authority of the Church of England, of dwelling upon those sides of church life that were Catholic and traditional and thus above the merely national. To the leaders of the Oxford movement

among them John Henry Newman (in its early stages), Edward Bouverie Pusey and John Keble (*qq.v.*), it was much more important that the church should be teaching the truth than that it was the national church. Suspecting the royal supremacy, they suspected the Reformation which gave it birth. They took hold of all the most traditional elements in church life and sought to restore to the Church of England the teaching and practice of the ancient and undivided church. Their successors little by little transformed the worship of the Church of England, attempting to turn the churches from houses of preaching and teaching into temples of awe and sacramental grace. And since some of their adherents would no longer recognize the final court of appeal in ecclesiastical causes the resulting ritual suits in the courts were painful. But perhaps the most remarkable result of the movement was the founding, from the 1840s, of communities of women and then of men under religious rule. Despite some eccentricities at first, these communities soon came to feel, and were felt to be, at home in the Church of England, and this was then unique in a church so deeply influenced by the Reformation (*see* MONASTICISM; WOMEN'S RELIGIOUS ORDERS).

Much relief was brought when the Church of Ireland was disestablished in 1869 and the Church of Wales in 1920, since these were churches of a small minority of the population. But a motion in 1871 to disestablish the Church of England failed, and most Anglicans and some non-Anglicans have continued to value an establishment, stripped of the features galling to dissent, as an important recognition of Christian standards in the national life.

One other significant stage in the partial separation of church and state was completed in 1871, when the colleges of Oxford and Cambridge, and the University of Durham, were removed from that almost exclusive clerical control that had existed since their foundation. English clergy had normally been trained at the universities; and though they continued in part to be so trained, it was felt desirable to create seminaries where exclusively Anglican teaching could continue to be given. The first theological college in the modern sense was founded at Chichester in 1839. From 1918 the bishops insisted that almost all ordinands should pass through one of these colleges.

4. The Church and the New Scholarship.—Since the age of Hooker, and indeed from very early times, the English church had a tradition of learned clergy. Especially between 1660 and 1850, English clergy like Joseph Butler contributed eminently to the philosophy, the science and the scholarship of Europe. With the decline of leisure and through the new pastoral zeal of the 19th century, this reputation began justly to decline. But as the new science and new historical criticism became powerful, such churchmen as Joseph Barber Lightfoot, Mandell Creighton and William Stubbs remained among the leaders of European scholarship.

All churches during the 19th century experienced the difficulty of adjustment to the new science and to the proof, for example, that the earliest chapters of Genesis were not to be taken as literal history. Here the existence of the judicial committee of the privy council as the final court made for liberty of opinion. In 1864 it reversed a judgment of the ecclesiastical Court of Arches which had condemned two of the authors of the book *Essays and Reviews* (1860), a book in which essays by Benjamin Jowett and others called for a new attitude to biblical studies. (W. O. C.)

IV. 20TH-CENTURY HISTORY AND DEVELOPMENT

The beginning of the last decade of the 19th century saw two events—the one the unintended result of, the other within, what the Oxford movement had become—which indicated the lines of subsequent ceremonial and theological development in the Church of England. These were the Lincoln judgment (*q.v.*; 1890) on ritual practices and the publication (1889) of *Lux mundi*, a volume of essays by various authors. The former aided the progress of the ritualistic and doctrinal changes that became known as Anglo-Catholicism. The other, more fundamentally important, revealed the progress of the principles of biblical criticism among men influenced by the Oxford movement (*i.e.*, in circles in which these principles had hitherto been repudiated or ignored) and so gave, if unintentionally, great impetus to the growth of theological Liberalism or Modernism. *Lux mundi* may be regarded as the effective beginning of the "higher," or historical and theological, criticism of the New Testament in the Church of England. Such criticism had already won some measure of acceptance as regards the Old Testament, but was curiously delayed in its application to the New, where its inevitable consequences were feared as being much more serious for traditional or orthodox Christianity and therefore met with strong if gradually weakening opposition. In the mid-20th century no recognized eminent theologians of the Church of England or any recognized teacher of theology in any English university would deny the importance or refuse to accept the principles and methods of biblical criticism in the study of either Testament, though they might differ about its alleged results.

1. Anglo-Catholicism.—This development of the Oxford movement for long encountered the strongest opposition of a kind often fanatical rather than thoughtful, which often assisted what it sought to destroy. In 1890 John Kensit founded the Protestant Truth society, pledged to such opposition. It engaged in such activities as raiding and interrupting divine worship in ritualistic churches, and even attempted to secure the election to parliament of candidates pledged to oppose ritualism. Archbishop Edward Benson strove to promote peace, but the feud was intensified by the papal condemnation of Anglican holy orders as invalid in 1896.

This disappointment of Anglo-Catholic aspiration, so far from satisfying the militant Protestant opposition, provoked a new outburst of anti-Roman controversy in which Sir William Harcourt, the champion of the earlier Public Worship Regulation acts, took a leading part. The answer produced by the archbishops of Canterbury and York to the grounds on which the pope's condemnation was based was widely accepted beyond the membership of their church as irrefutable; but, probably like Noah and his family in the ark, the English archbishops and bishops were more troubled by the stench within than by the waters without. There was flagrant disorder within the church, though it was then far less than it later became. Various questions concerning it were argued at the "Lambeth hearings" in 1899. The archbishops condemned the ceremonial use of incense and the carrying of lights in processions, and, in general, bishops accepted these decisions as the law in the matter. In the next year, the reservation of the sacramental elements was condemned, but though individual bishops attempted to secure obedience no concerted action was taken and the prohibitions were not successfully enforced.

Agitation became so acute that in 1904 a royal commission on ecclesiastical discipline was appointed, which in 1906 produced a report filling five "blue books." It revealed departures from the established use of the Church of England such as the interpolation into the order of Holy Communion of prayers from the Roman Catholic Mass, the practice of reservation, the Mass of the Presanctified, Corpus Christi processions, the invocation of the Blessed Virgin Mary and of saints and the observance of such Roman Catholic features as the Assumption of the Blessed Virgin Mary and the Sacred Heart of Jesus. These practices were bluntly declared to be "clearly inconsistent with and subversive of the teaching of the Church of England." They implied doctrines repudiated by that church and "should promptly be made to cease." There was also some neglect at the other end of the ecclesiastical variations—*e.g.*, evangelicals often omitted to say the prescribed daily services—but the commissioners held that few if any of such omissions had any doctrinal significance.

It was stated with equal frankness, however, that "the law of public worship in the Church of England is too narrow." The ecclesiastical courts were recognized as effete and their reconstruction was recommended, together with an increase in the power of bishops to deal effectively with obstinate lawbreakers, against whom it was suggested that the final penalty should be not imprisonment but deprivation. (Strictly, those who went to prison were sent there not as punishment for illegal ecclesiastical practices but for contempt in refusing to obey lawful orders of the courts.) It was further advised that letters of business should issue to the convocations to prepare a new ornaments rubric and "to make such modifications in the law relating to divine worship as may tend to secure greater elasticity." The letters were duly

issued. Convocation attempted compromises acceptable to neither party, and began a 20-year consideration of Prayer Book revision.

Like most else in the Church of England, Anglo-Catholicism was in practice a loose collection of elements by no means wholly consistent and held with varying degrees of intensity. From the time of *Lux mundi* it has contained both a liberal and a conservative or traditionalist, or even "Romanizing," wing. The former accepted biblical and historical criticism, though generally in moderate forms. It continued the best tradition of the Oxford movement, taking its standing ground upon modern learning while making full use of art and tradition in the worship of the church; and it often adopted something like Newman's doctrine of development. One of the best expositions of its theology was the composite work *Essays Catholic and Critical* (1926). It gives an account and offers a defense of a Christian belief which values historical continuity, Catholic custom and fully sacramental religion, combined with an acceptance of modern criticism and learning. Some of such Anglo-Catholics had considerable sympathy with the form Modernism took in the Roman Catholic Church, which, in essence, was Newman's doctrine of development pushed to extremes that would have shocked its author. It is significant that the Anglo-Catholic English Church union (now the Church union) criticized the papal condemnation of Modernism, while in general Liberals and Modernists in the Church of England defended the excommunication of A. F. Loisy and G. Tyrrell on the ground that their apologetic was disastrous.

But Anglo-Catholicism as taught in an increasing number of parish churches had little in common with the theology expounded by its leaders. There was a large and vigorous propaganda, more influential in the south of England than in the north, which was traditionally Roman in spirit and outlook, in its teaching about revelation, ecclesiastical authority, the binding character of church tradition, the necessity of auricular confession and priestly absolution and in forms and modes of worship. It was often denied, in practice if not avowedly, that the Church of England and her bishops, apart from so-called "western" (*i.e.*, Roman Catholic) tradition, had any separate authority to forbid teaching or practices allowed by that tradition. The baseless hope was cherished that someday the papacy would compromise about its claim to infallibility. This was a fundamental departure from the characteristic attitude of the Oxford movement. It was claimed that there was no important difference between the Anglican Eucharist and the Roman Mass. Thus, in 1924, about 3,000 clergymen presented a declaration of faith to the Orthodox patriarch of Alexandria. It stated that

> the signatories affirm that they hold it to be the genuine doctrine of the Anglican Church that they have received Apostolic Orders through their Bishops with the purpose that we should offer the unbloody Sacrifice of the Eucharist for both the living and the departed. We hold that by consecration the bread and wine are changed and become the true Body and Blood of Christ, and we hold that Christ, thus present, is to be adored.

This, as a leading Jesuit affirmed, states the pre-Reformation and the modern belief of the Roman Church; but to claim that it represents the doctrine of the Church of England is to falsify history.

Anglo-Catholicism has done much to transform and improve the outward appearance of churches and their services and also, though to a less extent, the teaching if not the preaching given in parishes. But no impartial observer could fail to note one fundamental weakness which the results of well over a century of devoted efforts made manifest to all who do not refuse to see. This is its failure to win the laity, as distinct from a small but energetic minority, to its ideals. Earlier hostility has largely yielded to mute tolerance by the mass of lay people, but the tolerance shows no sign of moving to acceptance. The movement remained predominantly clerical, and it has probably widened the difference in outlook and mentality between clergy and laity. In so far as it had a "Romanizing" section, that influence has greatly declined. But, in a more moderate sense, what "Anglo-Catholic" signifies had probably by mid-20th century won the support of a majority of the clergy. Of the large numbers who are regarded as evangelicals or "central" churchmen, few any longer vigorously oppose. But in the life and religion of the country as a whole, the influence of the distinctive tenets of Anglo-Catholicism was never great. The insensibility of public and church lay opinion on these matters would satisfy Epicurus. The laity remain fundamentally, though rarely now aggressively, Protestant.

2. Liberalism or Modernism.—The other characteristics of the Church of England from the last decade of the 19th century was Liberalism or, as those who disliked this name because of its political associations preferred to call it, Modernism. For long it was mainly liberal Protestant in character, though not exclusively so. Its twin bases were (1) that scientific knowledge and inferences from it had rendered the whole cosmology of the New Testament incredible to any educated person, having demonstrated its wholly mythical character; (2) that modern critical historical studies rendered a thorough reinterpretation of early Christianity essential and that this involved the elimination from theology of all that was irreconcilable with modern thought. There were great differences between its adherents and sympathizers, who had no one philosophy. They were agreed that problems of religious belief could not be settled by the decisions and formulations made by ecclesiastical authority at a date when the issues were not the problems they had later become. Thus the emphasis of Modernism was upon the intellectual aspects of Christianity and was a protest against a stereotyped intellectualism; traditional formal theology was expressed in terms of outmoded knowledge and philosophical notions no longer tenable. It was an attitude of mind rather than a system of beliefs or unbeliefs, though its critics could not unreasonably complain that what modernists did not believe was much more clear than what they did believe. It aimed at a reconstruction of theology and a restatement of the doctrines of the church "in terms of modern thought." It had little organization and never had a popular appeal.

At the close of the 19th century the Churchmen's Union for the Advancement of Liberal Religious Thought (since 1928 the Modern Churchmen's union) was founded and very slowly increased in membership, which, however, was never large. Until about 1930 it could probably claim to have had the support of more intellectually distinguished clergymen and laymen than any comparable organization in the church, though probably never including in its membership one-tenth of those who in varying degrees sympathized with its ideals. Modernism suffered through being too closely identified with the basic theology associated with the name of A. von Harnack and trying to maintain too much of this after it had been shattered by Loisy and then by Albert Schweitzer. But it did much to win acceptance of its main purpose, for reasonable intellectual freedom is no longer threatened, if it ever was. In 1911 J. M. Thompson's *Miracles in the New Testament* denied that the "nature-miracles" ever happened, and he was compelled to abandon the exercise of his ministry in Magdalen college, Oxford, by the combined action of the bishop of Oxford, Charles Gore, himself a liberal theologian, and the then bishop of Winchester, E. S. Talbot, the visitor of the college. This would not have happened, say, 40 years later. In 1912 a book by seven Oxford men, entitled *Foundations*, provoked controversy, particularly the contribution of its editor, Canon B. H. Streeter, who became a New Testament scholar of European reputation. His suggestion that the Resurrection appearances were visions, like that of St. Paul on the road to Damascus, produced by the living Christ, was severely criticized, and Streeter was relieved of his examining chaplaincy by the bishop of St. Albans, but was appointed a residentiary canon of Hereford cathedral by the bishop, John Percival, of that diocese. In *Essays Catholic and Critical* in 1926 a view requiring a magnifying glass of considerable power to distinguish it from Streeter's was maintained by a liberal Anglo-Catholic scholar; it aroused no controversy. A spectacular protest by the noble if extreme Anglo-Catholic missionary bishop Frank Weston of Zanzibar, which included a formal excommunication of the bishop of Hereford for appointing Streeter to a canonry, led to a debate in the upper house of Canterbury convocation. The bishops adopted a characteristic resolution, tending in opposite directions. It affirmed that the denial of historical facts stated in the creeds went beyond the limits of legitimate interpretation and gravely imperiled the sincerity of profession plainly incumbent

on ministers of the word and sacraments; yet, the new problems raised by historical criticism demanded "considerateness in dealing with that which is tentative and provisional in the thought and work of earnest and reverent students."

The Lambeth conference of 1920 gave no guidance in the controversy created by the advance of Modernism. But a commotion caused by papers read at a conference organized by the Modern Churchmen's union in 1921 on the doctrine of Christ's person led to the appointment of an archbishops' commission on doctrine. This very representative body of theologians could not be accused of haste: their report, *Doctrine in the Church of England*, was issued in 1938, when its chairman since 1925, William Temple, had been archbishop of York for nine years. The report of the commission is not a compendium of official Anglican doctrine. By another evasion, it was not accorded official approval, though it was authoritatively commended to the study of churchmen. It admirably fulfilled its purpose, which was "to consider the nature and ground of Christian doctrine with a view to demonstrating the extent of existing agreement within the Church of England," and to suggest how existing differences could be diminished or removed. There was some adverse criticism, largely based on misconceptions of its purpose; but the report remains as perhaps the best example of the modern Anglican attitude to and method of dealing with doctrinal questions. Modernists were gratified and generally accepted it as conceding the freedom they desired. But by the delayed date of its issue the matters dealt with in it were no longer so acutely controversial as when the commission was appointed. Controversy later died down, and World War II brought new perplexities.

There was one further stirring of similar waters. In 1947 a book by the bishop of Birmingham, E. W. Barnes, entitled *The Rise of Christianity*, went much further in a negatively Modernist direction than anything previously written by an Anglican bishop. Some 25 years earlier he had been rebuked by the archbishop (Randall Davidson) for the language in which he had condemned what he regarded as the sacramental extravagances of some of the clergy in his diocese. Some desired that the bishop should be put on trial for heresy in respect of parts of his later book; but this was avoided by a personal statement from the archbishop (Geoffrey Fisher) in the Canterbury convocation, expressing his dissent and saying that if he held the opinions of Barnes he could not continue to hold episcopal office in the Church of England. No further action was taken and the bishop ignored the advice. It was probable that abatement of acrimonious controversy was the result of long familiarity with doctrinal differences and of a decline of interest rather than to acceptance of radically Modernist opinions, though Barnes's book was severely criticized by theologians known to be liberal in their views. But such toleration of doctrinal differences constitutes an absolute barrier to any more friendly relations between the Church of England and the Roman Catholic Church. (*See* LIBERALISM, THEOLOGICAL.)

3. Constitutional Changes.—A commission appointed in 1906 revealed that though the church was the largest single religious body in Wales less than a quarter of the inhabitants even nominally belonged to it. In 1912 a bill to disestablish the Welsh church was introduced and eventually passed under the Parliament act, but because of World War I the separation from the Church of England was not effected until 1920. Two new dioceses were later created and the total of six form an administratively distinct province, presided over by an archbishop of Wales. Serious loss of endowments was involved, which many feared would prove crippling, but a notable co-operation between clergy and laity led to a remarkable awakening to the duty of supporting the church financially and the predicted disaster did not occur.

In England, a desire and need for lay co-operation in church affairs had brought about a measure of it, but without legal status. In 1913 the archbishops were petitioned that to appoint a committee on church and state was advisable in order to secure "a fuller expression of the spiritual independence of the church as well as of the national recognition of religion." An unofficial body known as the Life and Liberty movement did much to awaken interest in the need for change. Three years later, the committee's report formed the basis of an address to the crown recommending legislation to provide for a national assembly of the Church of England; and in 1919 the Church of England Assembly (Powers) act (otherwise known as the Enabling act) became law.

The assembly thus constituted can prepare and pass measures for submission to parliament "relating to any matter concerning the Church of England" except doctrine, which is still reserved to convocation. It has three "houses" formed by the upper and lower houses of the convocations of Canterbury and York and an elected house of laity, though the method of election is very indirect. All baptized members of the church who signify themselves as such and declare that they are not members of any other church not in communion with the established church may have their names entered on an electoral roll of the parish church they attend. They elect a parochial church council. The councils in turn elect the lay members of the diocesan conferences, who finally elect the house of laity of the church assembly, the number of each diocese being proportional to the numbers on the electoral rolls of the dioceses. The assembly meets in London, generally three times a year. A measure, corresponding to a parliamentary bill, is prepared by a committee and, if and when finally approved by the assembly, is presented to the ecclesiastical committee of parliament, a body constituted under the Enabling act. Its duty is to examine assembly measures and to report on their nature, legal effect and expediency, whereon (the report being favourable—if not, it is referred back to the assembly) any member of parliament may move that the measure be presented for the royal assent. Parliament may debate, approve or reject any measure, but no longer amends it. If thus passed by both houses of parliament, the measure becomes law on receiving the royal assent.

The church thus has a partially representative body through which it can initiate reforms and so more easily adapt itself to changing needs. It can carry out work for which parliament has neither the time nor the qualifications, though, if a church is "by law established," parliament must have the right to say what kind of church shall be or remain so established. There is increasing awareness that the church assembly has made the convocations largely superfluous, but historical sentiment has preserved them. Many and various important measures have been enacted: schemes for the better use of the church's declined manpower in the ministry; the provision of a clergy pensions plan, through the Church commissioners, no longer requiring premiums and on a more adequate scale; the formation of new dioceses; recommendations involving more effective use of the church's property and revenues; and methods of dealing with the occasional scandals caused by clergymen guilty of gross negligence or misconduct in other respects. Some contentious measures have not resulted in legislation (*e.g.*, reconstitution of courts to deal with ecclesiastical causes, first recommended in 1904). Ecclesiastical and so-called "spiritual" matters can easily involve the reputations and even the livelihood of clergymen. Neither parliament nor the lawyers are in the least likely to approve the removal of these from the jurisdiction of the royal civil courts or to deprive any citizen of his right to appeal to the highest court of the realm if he is aggrieved by an alleged failure of justice in an inferior court. The scandal of treating the right to present to a benefice as a marketable property was abolished by a measure of 1923, but the important and large patronage of the crown remains.

There has sometimes been considerable criticism of the working of the church assembly and its occasional ill-considered projects; but the benefits of many of its achievements are beyond dispute. Many long overdue changes have been effected and some increase in the corporate consciousness of the church. The assembly's press and publications board attempts to increase knowledge of church affairs, a very necessary work if the inveterate parochial isolation, with little concern for wider interests and duties, is to be overcome. The dioceses have steadily increased the percentages of the apportionment of money asked of them by the central board of finance.

4. Prayer Book Revision.—An attempt to revise the Prayer Book of 1662 began after the issue of letters of business as recommended by the royal commission in 1906, and its later stages

aroused more interest than any activity of the church in the first half of the 20th century. It ended in a double failure to secure parliamentary authorization for the revision proposed by the church in 1927 and 1928. This revision was intended as a permissive alternative to the book of 1662 but not to displace it (*see* COMMON PRAYER, BOOK OF). The new rubrics were in the direction of greater liberty. They concerned the services of Holy Communion, baptism, confirmation, the solemnization of matrimony and the ordinal. Provision was made for additional services; numerous occasional prayers were added, with thanksgivings, collects, epistles and gospels. The controversial ornaments rubric was reprinted without interpretation, but a variety of vestures at Holy Communion was allowed. Some portions of the Psalms considered un-Christian in spirit and wording were made optional for use in public worship. A form of the late evening service known as compline was included.

Controversy centred on the revised order of Holy Communion and the permissions proposed in connection with it: (1) Reservation of consecrated bread and wine wherewith to give communion to sick persons unable to attend church "on the same day and with as little delay as may be." (2) Perpetual reservation; *i.e.*, the priest was to be allowed to set aside consecrated bread and wine for giving the sacrament in other emergencies, sicknesses or approaching death, if the bishop gave a licence for so doing. The bishops directed that such reservation should be

> only for the Communion of the Sick, shall be administered in both kinds, and shall be used for no other purpose whatever. There shall be no service or ceremony in connection with the sacrament so reserved, nor shall it be exposed or removed, except in order to be received in Communion, or otherwise reverently consumed.

The manner in which reservation might be made was to be determined by rules made by the archbishops and bishops. The manifest purpose of these directions was to prohibit the Roman Catholic services in connection with exposition of the Blessed Sacrament and benediction, where they had been introduced in the Church of England by the more extreme Anglo-Catholics under the name of "devotions." (3) The addition to the prayer of consecration of the anamnesis or memorial, and the epiclesis or prayer to the Holy Spirit "to bless and sanctify both us and these thy gifts of bread and wine, that they may be unto us the Body and Blood of thy Son our Saviour, Jesus Christ," to be followed, without a break, by the so-called prayer of oblation, the first of the postcommunion prayers in the rite of 1662.

Opposition came from those who objected on principle to any alternative order of Holy Communion; from conservative evangelicals who objected to reservation in any form; from those Anglo-Catholics who desired reservation not only for the communion of the sick but also for the purposes which the proposed rules were designed to forbid; and from those who, while not opposed to reservation for the purposes proposed, had no confidence in the rule-making power reserved to bishops or that they could or would enforce them.

Moderate Anglo-Catholics and evangelicals, most modernists and all "central" churchmen were prepared to concede much that they did not like for the sake of order and peace. At the July session of the church assembly, 1927, the revised book was passed by 517 against 133 votes. When parliamentary approval was sought, the house of lords gave it by 241 votes against 88. Then on Dec. 15 came the shock of rejection in the house of commons by 238 votes against 205. This result was popular in the country as a whole. The bishops had claimed, not unreasonably, that the book did not depart from the Reformation settlement and, with possibly more hope than grounds for it, that it would assist the restoration of order. The house of commons believed neither proposition. The home secretary of the day, Sir William Joynson Hicks (himself a rather militant evangelical), pointed out irrefutably that the primate, Randall Davidson, could not deal with "Romanizing" clergymen though on two notable occasions he had promised to do so.

But the end was not yet. The bishops met and persuaded themselves that the commons' adverse vote was due to "avoidable misunderstandings," and resolved to present the book again in substantially the same form with some explanatory modifications. The only result was to reduce support for the revision. The assembly voted for approval by only 396 to 157, and in June 1928 the house of commons again rejected it by 266 votes to 220.

In the event most bishops consulted with their clergy, assembled in synods, by means of a questionnaire, the main points being: (1) whether the bishop in his administrative discretion should be guided by the 1928 book, and (2) whether he could expect support in requiring that practices consistent neither with the book of 1662 nor with that of 1928 should cease. In general it appeared that the clergy were prepared to support the bishops on both points, though a considerable minority was not; and while a majority of the synods approved the 1928 reservation rubrics, it was by a minority vote of the total votes cast in all the synods. The bishops issued a declaration that they proposed to use the book of 1928 as the standard by which to regulate additions to or departures from that of 1662. In so far as they subsequently did so, they acted neither uniformly nor energetically, and therefore met with little opposition or acceptance. Varieties and illegalities continued much as they were before Prayer Book revision began. But, as with doctrinal issues, so with forms of worship, extreme or "Romanizing" Anglo-Catholicism declined. Controversy became much less keen, and most of the formerly novel usages had continued for so long that opposition wearied; an adult generation grew up that had been more or less familiar with them from childhood. Those who disliked them in country parishes gave up attending church, in town parishes either did likewise or went to another church. On the other hand, as few bishops tried seriously to intervene, the clergy appeared to have become more amenable to such episcopal directions as were given. After 1928 there was a considerable increase of toleration in these matters. The serious widespread indifference to religion made church people realize the futility of the older controversies and encouraged a wider and a wiser sense of true values.

5. Relations With Other Churches.—Awareness of the harm resulting from a divided Christendom was heightened after World War I. The weakness of Christian witness by churches not in communion with one another came to be felt by every reflective person. Some Protestant churches achieved federation, and the former tendency to form new sects of any size or significance disappeared. The Church of England is hampered in a sincere desire for union, or at least intercommunion, with other churches by two factors. On the one hand, very large numbers of her regular and devoted members can only with difficulty be persuaded to take more than a mild interest in church matters outside the borders of their own parishes. On the other hand, the baseless hope of so many of the clergy of eventual union with Rome has made them hostile to any move by their own church that might make this more remote hope less likely ever to be realized.

According to Roman Catholic doctrine the church of Christ cannot be divided; those who refuse obedience to the bishop of Rome are outside the church and must return to it by way of submission. For Rome, reunion in any other sense has no meaning. Nevertheless, efforts have been made, and these became on the Anglican side more official in character after the Lambeth conference of 1920. A series of "conversations" was held at Malines (Belg.) during 1921–25 between Anglican delegates led by Lord Halifax, J. Armitage Robinson and W. H. Frere and certain Belgian and French clergy under the chairmanship of Désiré Cardinal Mercier and later under his successor, Ernesto Cardinal van Roey. By Anglo-Catholics these talks were regarded with hope, by many evangelicals with fear and dislike, by others with languid interest. The later talks were held with the cognizance of the pope. But the differences remained fundamental and papal favour was withdrawn. The encyclical *De Vera religionis unitate fovenda* marked the end of hope (1928). All gatherings for discussion of reunion, such as were contemplated by Anglican bishops at Lambeth in 1920, were forbidden to Roman Catholics and their taking part in them was said to be "giving countenance to a false Christianity quite alien to the one Church of Christ." The most that has been possible is co-operation in certain social enterprises and in some matters of Christian morality, though even in the latter there are serious obstacles to agreement. (*See* also ECUMENICAL MOVEMENT.)

Connections of the Church of England with other churches of Christendom have grown considerably, though with small formal result. Between Canterbury and the Orthodox Eastern Churches there has been some *rapprochement* since 1906, when a society of members of those churches and Anglicans was formed for the purpose of promoting closer relations and ultimate organic union. The patriarch of Constantinople was officially represented in discussions of Anglo-Orthodox relations, and in 1922 his synod formally recognized the validity of Anglican orders. Some other churches have given similar recognition and closer relations have been established with them. But there was no sign that the Orthodox Eastern Churches would consider organic union except on a dogmatic basis as rigid as that of Rome and that of a kind and content out of the question for the Church of England.

Anglican bishops have assisted at the consecration of Swedish bishops and there has been occasional intercommunion between the two churches. But reunion with the free churches in England alone rouses appreciable interest, and any advances in this direction hinder those made to the Roman or Eastern Churches, and vice versa. For a time, after the Lambeth conference of 1920, hopes of positive results in "home reunion" were reasonably bright but were not fulfilled. The Church of England has not modified the basis of reunion known as the Lambeth quadrilateral: the necessity of the acceptance of Holy Scripture as the rule of faith; the Apostles' and Nicene creeds; the sacraments of baptism and Holy Communion; and episcopacy. The last requirement is the stumbling back, and no effective advance toward its removal or compromise about it has been made.

Yet the Church of England is not wholly consistent in practice in this matter. Communicants of nonepiscopalian churches are welcomed to Holy Communion in Anglican churches if isolated from their own communions; both English archbishops have given addresses at annual assemblies of free churchmen, whose ministers also sometimes preach in Anglican cathedrals and churches with the approval of bishops. But a bishop of the Church of South India, though made deacon, ordained priest and consecrated bishop by Anglican bishops, if he comes to England may preach and receive Holy Communion in the Church of England but may celebrate it only if he will promise not to do so in any church not in communion with the Church of England. It is not, however, treated as a barrier to intercommunion with the Church of Sweden that it is in communion with the nonepiscopal Church of Norway. Such procedures are, of course, maintained because of a laudable desire not to alienate members of the Church of England who hold them to be fundamental matters of principle.

The 1920 Lambeth conference acknowledged free church ministries to be real ministries within the catholic church. Such ministers will not accept any need for a fresh commission or ordination on their side, while Anglican priests require only formal recognition if they wish to become free church ministers. It is hardly possible to say what any of the separated churches would be prepared to modify or surrender in order to attain organic union.

Archbishop Geoffrey Fisher (1945–61) was untiring in his efforts to promote closer and more friendly relations between separated churches. Attempts to achieve organic union having failed, in 1946, in a sermon before the University of Cambridge, he made what to many appeared to be a more promising appeal. This was that the free churches, in order to make intercommunion possible, should consider taking episcopacy into their systems of order and thus to encourage growing more closely together, leaving organic union as the ultimate goal. The implications were studied by joint meetings of Anglican theologians with representatives of the free churches, and a report, *Church Relations in England*, appeared in 1950. This showed that the discussions, instead of being confined to the limited purpose, went once more and at large over general theological agreements and differences, ground fully explored before. It, with several later individual statements, made clear that no immediate favourable response to the archbishop's suggestion would be forthcoming. Negotiations with the established Church of Scotland, which is Presbyterian, also issued in a joint report recommending continuance of discussion, interchange of preachers, theological teachers and students and limited sanction for baptized and communicant members of the Church of Scotland to receive Holy Communion in the Church of England. Three other documents resulted from the Cambridge sermon. "Grave obstacles," "formidable barriers," "unresolved tensions" and other phrases of this familiar and weary company, reveal the long journey ahead and little hope that it will end otherwise than at its starting point.

6. Foreign Missions.—The great expansion of missionary work in almost every part of the world that was so marked a feature of the 19th century continued until World War I, but thereafter the increase in manpower and money for this work did not keep pace with the opportunities or even with the maintenance of work already established, which had to be restricted. Since that date the growth of nationalism and the desire for independence in former missionary areas has tended strongly against the acceptance of what seemed to the influential parties in these lands as the religion of foreigners and a means of maintaining their superiority. This fact, in the main, was realized and accepted by the church, which concentrated less upon planting branches of the Church of England abroad and more upon carrying the Christian gospel to other lands while encouraging the growth of indigenous churches, with a minimum of direction from the Church of England. Even so, in the far east, missionary expansion has become impossible because of the conditions created by war, the rise and rapid spread of nationalism in Asia and the rule of communism in some countries. The work of the Overseas Council of the Church of England, in facing what is regarded as the essentially missionary character of Christianity, continues its efforts in supporting relatively young churches overseas where such help is possible and acceptable. In this matter Geoffrey Fisher, too often spoken of as primarily a very able administrator mainly concerned with the machinery of the church, showed real wisdom. When he became archbishop in 1945, in the vast continent of Africa only South Africa was an independent province within the Anglican Communion. By his retirement in 1961, practically all Anglican-African dioceses were completely organized into such provinces, each under its own archbishop, independent of the archbishop of Canterbury and of the Church of England, though in full communion with it (*see* ANGLICAN COMMUNION). In each case ecclesiastical autonomy was achieved before political independence.

After World War II, international conferences of practically all the non-Roman Catholic churches were held under the auspices of the World Council of Churches (*q.v.*). Their purpose was the development of an international Christian consciousness and witness in social matters, and a closer, ultimately a united, co-operation among the participating churches. Representatives of the Church of England took a prominent part in these gatherings. This ecumenical movement (*q.v.*) was claimed by Archbishop William Temple, who contributed much to it, to be something genuinely new. It is partly the outcome of the abolition of isolation formerly imposed by distance. Some think its relative ineffectiveness to be attributable in part to seeking a degree of theological agreement among the churches that does not in fact exist within the participating churches.

7. Establishment.—The Church of England is established by the state, which thereby professes to be Christian. This means that parliament has added its sanction to the church's own formulations of doctrine, order and worship as expressing the form of Christianity officially recognized in England. Though not strictly a legal contract, the establishment is more like a contract than like anything else, and therefore cannot be varied or repudiated by one party to it. The church is not free to make changes in doctrine, order or worship without the consent of parliament, if the church is to remain established. Thus the church's freedom is limited. In practice it works as follows:

1. By the Enabling act (1919) the Church assembly, composed of the convocations of Canterbury and York (*q.v.*) and an elected house of laymen, can prepare measures on matters concerning the church (except in doctrine, which is reserved to convocation). These measures "lie on the table" of the house of commons for a period, and may or may not be challenged. If challenged, they may be debated or approved or rejected. If unchallenged they in due course receive the royal assent and become part of the law

of the land. Parliament does not amend them.

2. Ecclesiastical law is part of the law of the realm. It is administered by church courts, whose judgments (unless reversed by the civil courts on the ground that justice has not been done in the courts below) will be enforced by the civil power. Church courts deal with doctrine, worship, ritual, the ecclesiastical and moral obligations of clergymen. There is a final appeal to the judicial committee of the privy council, which may have ecclesiastical assessors, which only decides, by interpreting accepted formularies according to their legal construction, whether or not justice has been done in the church courts.

3. Appointment to all bishoprics and deaneries is with the approval of the sovereign, who acts upon the advice of the prime minister.

4. The convocations meet not as of right but are summoned to do so by royal writ. They have an ancient right of direct access to the throne.

The establishment charges the church with the duty to bring the ordinances of religion within the reach of all persons in England who desire them. It gives to the parochial clergyman a recognized place in English life. It does not mean that clergy are paid or the church subsidized by the state, which pays only such ministers of any church as are directly employed by the state (*e.g.*, prison chaplains). By virtue of accepting the obligations involved by establishment the church continues to possess endowments inherited from the past when its declared and practising membership was much larger than it has since become.

There is little opposition to the continuance of the establishment of the Church of England by the other churches in England. Many Anglicans desire a larger measure of freedom for the church, but it is doubtful if this could be attained and establishment maintained. There are admitted anomalies that do not trouble the majority of Englishmen so long as practical considerations are on the whole satisfactory.

8. Church and Education.—After 1870, when the state began to provide elementary or primary education in districts where there were no church schools, church control and influence in education steadily and sometimes rapidly declined. Education acts of parliament assumed the continuance of voluntary schools (which category includes church schools) but, especially after the act of 1944, involved the church in heavy additional expenditure and therefore critical decisions. The raising of the general education standard and of teachers' salaries considerably increased costs. Much more serious, every school site, old or new, had to conform to stringent building regulations. The state was willing to assist the managers of voluntary schools to the extent of half the cost of bringing their schools up to the prescribed standard and half the cost of subsequent external repairs. Where this obligation was acceptable, the church school became "aided," which meant that two-thirds of the managers would be church-appointed, the managers could choose the teachers, the clergy could teach in the schools and the parish have the use of the school buildings for other purposes as previously. If the financial requirements could not be met the schools became "controlled." This status meant that one-third of the managers might be church-appointed and that there could be two periods a week of church teaching for children whose parents requested it; all managers were to be consulted about the appointment of head teachers, other teachers being appointed by the local education authorities, a proportion of them being competent to give religious teaching. The result was a great decline in the number of church schools. Many of the so-called public schools (*i.e.*, independent schools, chiefly boarding) are church foundations, where religious education is provided and there is daily worship in the school chapel.

The church has a number of colleges for the training of schoolteachers, and some dioceses have a theological college running two- and three-year courses for the training of ordinands. At Oxford and Cambridge some theological professorships remain confined to clerks in holy orders and most of the colleges are Anglican though there are no religious tests for membership. The other universities of England have no ecclesiastical allegiance, but some have theological faculties and also provide training in the teaching of religion. The sole requirements for appointment to professorships and other teaching posts in such faculties are competence and good character.

The attitude of the state to the church in the matter of education reveals two advantages to the Church of England in being the church "by law established" in that country: there is no anti-Christian civil policy and no church party in politics, two evils of which it would be hard to decide which is the greater.

9. Clergy.—The number of diocesan bishops was increased in order to lighten the bishops' burdens by dividing the larger dioceses. Many more suffragan and assistants were also appointed. The result has been little gain and much cost to the church. Bishops continued to be overworked and the duplication of cathedral churches, staffs and the ramifying paraphernalia of diocesan boards and organizations involve increased expenditure which could be more profitably employed. It has also become clear that more diocesan bishops do nothing to build a stronger and more influential church.

While the number of bishops altogether, including suffragans and assistants, more than doubled in the years between 1890 and 1950, the number of clergy was nearly halved. At the same time the population increased by almost one-third. Yet the clergy in the parochial ministry form the front line of the church's work in the world. Important as the central organization may be and the debates in convocation and the church assembly with innumerable boards and committees both in London and in the dioceses, there would be little for them all to do and nothing very useful they could accomplish if the work in the parishes failed. A large number of town and country parishes were united into single charges or made tenable in plurality by one parish priest, but in 1959–60 there were 900 vacant benefices and in many parishes of over 10,000 people the parish priest worked single-handed, without the assistance of curates. The practice of ordaining older men retired from the professions did something to meet the shortage, but the proposal that men not intellectually qualified for ordination but otherwise suitable should be ordained for country parishes while continuing to do rural manual work did not meet with approval. It would not greatly ease the position in densely populated cities and large towns, and it would also require some direction of labour by bishops. Further, the ministry, like other callings, is indeed a vocation, but some of its more indispensable tasks cannot be adequately performed, or performed at all, without ability and knowledge at a professional level of competence. In 1961 the total number of clergymen with parochial charges was 10,375 with (including chaplains to the armed forces of the crown) 2,700 assistant curates; if the retired but often active clergymen are included, clergy numbered 18,969. The number ordained annually had risen from 479 in 1952 to 626 in 1961, which latter figure about balances the annual loss by death and retirement. Even so, the number of clergymen had declined by over 10,000 since the beginning of the 20th century, while the population had increased by more than one-third. However, the late 1950s and early 1960s showed a marked increase in the number ordained.

10. Church Commissioners.—This important body, whose main function is to provide stipends, houses and pensions for the clergy, gained its name when two corporations were amalgamated in the interests of economy and efficiency—Queen Anne's bounty and the Ecclesiastical commissioners. The commissioners' income doubled between 1957 and 1961, largely as the result of a new investment policy of taking up industrial shares. They have none in breweries, newspapers, entertainments or in industries known to be associated with armaments. Stipends of the poorer clergy, a first charge on the commissioners' income, doubled in the same period.

A former important source of church income ceased to exist in its old form with the Tithe act of 1936 (*see* TITHES). About 7,000 rural benefices lost about 18½% of their income, cathedrals affected lost some £15,000 a year out of a total income of £95,000 and the Ecclesiastical commissioners something like £50,000 a year. The £70,000,000 worth of stock which the church received was in no sense a lavish endowment from public funds, but was simply payment for property which the government compelled the church to

sell for about four-fifths of its value. The Tithe act got rid of a long-standing cause of hostility between farmers and the church. (*See* also CHURCH COMMISSIONERS.)

11. War Damage.—Three ancient English cathedrals were seriously damaged during World War II: Canterbury (which suffered most), St. Paul's and Exeter. The Welsh cathedral of Llandaff was almost destroyed and also the fine church serving as the cathedral of Coventry. Many churches in cities and large towns were obliterated. War damage compensation went far to finance necessary restoration, and the regulations allowed diocesan authorities to use the compensation money for the building of churches elsewhere. The damage to cathedrals also was repaired. The building of the huge new cathedral at Liverpool, which suffered only superficial damage, except for some coloured windows, during the war, continued, and the first bay of its nave came into use in 1961. A new cathedral at Guildford was consecrated in the same year and a new one at Coventry in 1962. But the upkeep and maintenance of the ancient cathedrals and their services, notwithstanding the increases in the stipends of deans and residentiary canons made possible by the Cathedrals (Grants) measure (1954), is often very difficult, as they are impoverished by increased costs and prices which their revenues are insufficient to meet. Income received by them from the Church commissioners represents compensation for property surrendered by the cathedrals in the 19th century (the present annual yield from which greatly benefits the Commissioners' fund) and is but a fraction of its present yield. The restoration of a modest part of what was taken from the cathedrals would relieve their chapters of anxiety.

12. Canon Law Revision.—In the mid-20th century the convocations undertook a revision of canon law, a task which arouses little enthusiasm. It is however important that the church should have an up-to-date set of laws to regulate its life and practice, provided they are reasonably flexible without permitting chaos.

13. Membership.—It is impossible to state precisely the membership of the Church of England. The "establishment" means that this church is charged with the duty of bringing the ministrations of religion to all in England who desire them. Every square yard of the country lies within some parish, and all baptized persons who do not profess membership of another church or have no religion have the right, should they so wish, to the ministrations of some parish priest. In the fighting services all men who do not claim membership of another church or religion are classed as "C. of E."

In 1961 the total population of England was roughly estimated at 42,663,080, of whom 27,005,000 were baptized and 9,748,000 confirmed. The number on the electoral rolls of parishes was only 2,877,671, while about 2,218,500 received Holy Communion on Easter day or within six days afterward. Slightly less than one-half the babies born in England are baptized in the Church of England.

14. Mid-20th Century.—The dominant impression of any survey of the history of the Church of England from the last years of the 19th century is that of a great increase in centralization and organization. This undoubtedly made the church much more efficient as an institution; but it led to no extension of the spiritual influence of the church by whatever evidence that may be estimated. Yet without these changes that influence would probably have become much less than it is. During the period, the church was well served by its archbishops, all in their various ways remarkable men—Randall Davidson, Cosmo Gordon Lang, while the third, William Temple, though he held the primacy for only two and a half years, bears comparison with any of his predecessors. His successor, Geoffrey Fisher, resigned at the end of May 1961 and was succeeded by Michael Ramsey, previously bishop of Durham and archbishop of York. Fisher's efficiency and wisdom in many and varied services to the church were generally recognized, though he was not altogether popular. In almost the last of his travels, which included a visit to the ecumenical patriarch in Istanbul, he proposed a courtesy call on Pope John XXIII which was readily accepted. This was the first time since the Reformation that a pope and an archbishop of Canterbury had met. Some extreme Protestants within and without the Church of England opposed the archbishop's action; others hailed it as very important. *See* also references under "England, Church of" in the Index volume. (J. S. BE.)

BIBLIOGRAPHY.—J. R. H. Moorman, *A History of the Church in England* (1953); S. C. Neill, *Anglicanism* (1958); For special studies up to and including the Reformation *see* the relevant bibliographies to ENGLISH HISTORY. *See* also M. Deanesly, *The Preconquest Church in England* (1961); S. C. Carpenter, *The Church in England 597–1688* (1954), *Church and People 1789–1889* (1933); P. E. More and F. L. Cross, *Anglicanism* (1951), on the 17th century; W. H. Hutton, *The English Church From the Accession of Charles I to the Death of Anne* (1903); J. H. Overton and F. Retton, *The English Church From the Accession of George I to the End of the 18th Century* (1906); F. W. Cornish, *The English Church in the Nineteenth Century* (1910); N. Sykes, *Church and State in England in the Nineteenth Century* (1935); J. Wickham Legg, *English Church Life From the Restoration to the Tractarian Movement* (1914); A. M. Ramsey, *From Gore to Temple* (1960), *Doctrine in the Church of England* (1928); C. Garbett, *The Claims of the Church of England* (1947), *Church and State in England* (1950); C. M. Ady, *English Church and How It Works* (1940); B. Seebohm Rowntree and G. R. Lavers, *English Life and Leisure* (1951); J. C. Wand, *Anglicanism in History and Today* (1961). For contemporary information *see* the *Official Year Book of the Church of England* (annually); official reports of the Lambeth conferences (1867 *et seq.*); proceedings and publications of the Church assembly; prefaces to *Crockford's Clerical Directory* (biennially). (J. S. BE.; W. O. C.)

ENGLEFIELD, SIR FRANCIS (*c.* 1520–1596), English Roman Catholic, who was a close adviser of Mary I and a vigorous opponent, in exile, of the Elizabethan church settlement, was the eldest son of Sir Thomas Englefield (d. 1537), justice of the common pleas. Francis took the oath of royal supremacy, served as sheriff of Berkshire and Oxfordshire in 1546–47 and accepted in 1545 a grant of the manor of Tilehurst, which had belonged to Reading abbey. He was even knighted at the coronation of Edward VI (Feb. 1547), but the progress of the Reformation alienated him, and he attached his fortunes to the cause of the princess Mary. On Mary's accession, he was sworn a member of the privy council like many others who owed their promotion to their loyalty. He sat as member of parliament for Berkshire in all Mary's parliaments except that of April 1554. He received no higher political office than the lucrative mastership of the court of wards, but he was a personal friend of the queen and as such a very influential councilor.

An ardent believer in persecution, Englefield was present at Bishop John Hooper's trial, sought Roger Ascham's ruin and naturally lost his office and his seat on the privy council at Elizabeth's succession. He retired to the continent before May 1559, and from that time until his death was an active participant in plans for the restoration of Roman Catholicism. Hopeful at first of the virtues of persuasion, he grew more extreme as the Elizabethan church settled down. Englefield became the close confidant of William Cardinal Allen, Robert Parsons and the "jesuited" Catholics, who advocated forcible intervention by Spain and the succession of the infanta. In 1585 Englefield thought that Mary Stuart's succession, peaceful or other, would not be satisfactory unless it were the result of Spanish support and she were dependent on Philip II. In exile he soon occupied the position of expert on English affairs, though he proved as misinformed and prejudiced as the rest.

Englefield lived first at Rome, then in the Low Countries and finally at Valladolid, Spain. He was blind for the last 20 years of his life, and survived on his pension from Philip II. He had been outlawed in 1564 and his estates sequestered, but they were not forfeited until 1585, when an act of attainder was passed against him. A further act was passed in 1592 confirming the forfeiture to the crown. His nephew Francis eventually recovered some of the family estates, and was created a baronet in 1612.

ENGLEHEART, GEORGE (1750–1829), English miniature painter, was born in Kew on Oct. 26, 1750, the son of a German-born plaster-modeler. He studied under G. Barret, Royal Academician, and copied paintings in Sir Joshua Reynolds' studio before setting up on his own. His career is well documented, since his fee-book from 1775 to 1813 is still in existence; between those years he painted no fewer than 4,853 miniatures. Throughout this immense output he kept a steady level of conscientious craftsmanship, yet remained responsive to the changing fashions of the times.

He died in Blackheath, March 21, 1829. His nephew, JOHN COX DILLMAN ENGLEHEART (1784–1862), also was a popular miniature painter. (G. Rs.)

ENGLER, (HEINRICH GUSTAV) ADOLF (1844–1930), German botanist famous for his system of classification and as a plant geographer, was born at Sagan on March 25, 1844. He was educated at the University of Breslau, from which he graduated in 1863, and obtained the degree of doctor of philosophy in 1866. After teaching for four years he became in 1871 custodian of botanical collections of the Botanical Institute of Munich and remained there until 1878 when he was promoted to a professorship at the University of Kiel. In 1884 he was appointed professor of botany and director of the botanical garden of the University of Breslau and in 1889 became director of the Berlin Botanical garden, Dahlem. Under his directorship, which lasted until 1921, this became one of the foremost botanical gardens in the world. He made three explorations to Africa, in 1902, 1905 and 1913. In 1905 he also visited India and Java, and in 1913, undertook a journey around the world, traveling through Russia, Japan and the United States. He died at Dahlem on Oct. 10, 1930.

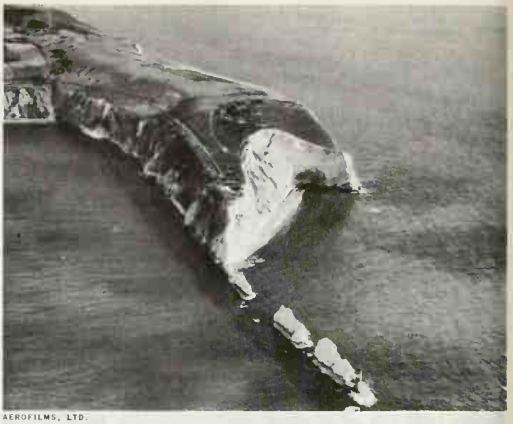
Engler was a great administrator, and the dictator of German taxonomic and geographical botany, although his contributions to other branches of botanical science are also important. He collaborated with Karl Friedrich Philipp von Martius in his *Flora Brasiliensis* (1840–1906) and with Alphonse de Candolle in his *Monographia Phanerogamarum* (1878–96), writing a number of monographs of various families. But his greatest contribution to taxonomy is his monumental *Die natürlichen Pflanzenfamilien*, edited with Karl von Prantl (1887–99, 2nd ed., 1925–), followed by *Das Pflanzenreich*, parts 1–102 (1900–37); his *Syllabus der Pflanzenfamilien* (1892), republished many times, is a most valuable reference book. His other works include *Die Vegetation der Erde*, with Oskar Drude, 15 vol. (1896–1923)—a series of phytogeographical monographs and many papers on the flora of Africa in which he was especially interested. He was also the founder of the *Botanische Jahrbücher*, which he edited from 1880 until his death.

ENGLEWOOD, a city of Bergen county, N.J., U.S., about 14 mi. N. of Jersey City, slightly inland from the Hudson river, and about 3 mi. W. of the Bronx borough of New York city. Originally founded in 1647 as a part of Hackensack township and detached in 1871 as the township of Englewood, it was incorporated as a city in 1895 and reincorporated in 1899 because of a constitutional flaw in the 1895 act. It was originally populated by English, Dutch and French colonists and has continuously been a residential community of especially fine homes with a few specialized and light industries. The Actors' Fund home for disabled, sick and retired actors has been located in Englewood since 1928. In 1951 the Englewood telephone system became the first in the U.S. to offer subscribers direct-distance dialing, by which they could reach telephones in cities as distant as San Francisco by dialing a series of code numbers. For comparative population figures *see* table in NEW JERSEY: *Population*.

(D. N. A.; M. P. M.)

ENGLISH CHANNEL (commonly called "the channel"; French *La Manche*, "the sleeve"), a narrow sea separating the southern coast of England from the northern coast of France. Taking the western limit as the line joining the Scilly Isles with the Isle of Ushant, it extends for approximately 350 mi. E. to the Straits of Dover. From west to east its width gradually decreases, with a marked constriction between St. Catherine's point in the Isle of Wight and the Cotentin peninsula. Its greatest width is about 150 mi. between the Gulf of St. Malo off the north coast of Brittany to Lyme bay in Dorset, Eng.; its narrowest width is 22 mi. from Calais to Dover. The main islands are the Channel Islands (*q.v.*). The name English channel was in contrast to the German ocean (the 19th-century name for the North sea).

Submarine Morphology.—The channel lies entirely within the continental shelf of northwestern Europe. The depth of the sea bed gradually increases from 20 to 25 fathoms near Dover to 55 to 60 fathoms south of Land's End. A remarkable narrow trough, Hurd Deep (north and northwest of the Channel Islands), has



AEROFILMS, LTD.

THE NEEDLES, CHALK CLIFFS RISING 100 FT. OUT OF THE ENGLISH CHANNEL AT THE WESTERN TIP OF THE ISLE OF WIGHT

the greatest recorded depth of 94 fathoms. The trough has closed ends; for this reason, it seems unlikely that it was created as a normal river valley, when sea level was lower, and more probably it is a tectonic feature of some antiquity. Elsewhere the sea bed is flat and rather featureless, although two miles south of Plymouth breakwater a submerged cliff exists in 20 fathoms of water. Similarly farther east in Start bay, there is a marked break of slope which causes the sea bed to drop from 8 to 24 fathoms in less than half a mile. The only other area of uneven bottom topography occurs in the Straits of Dover, where elongated sandbanks like the Varne, Vergoyer and Ridge are separated by deepwater hollows.

Geology and Geological History.—There is little observable relationship between the form of the sea bed and the rocks of which it is composed. Recent studies show that the submarine geology is more complex than the simple extrapolation of land outcrops on either side of the channel would suggest. The channel is occupied by a tongue of chalk running westward toward the edge of the continental shelf. As a result of earth movements the chalk has been folded and in places younger Tertiary beds have been preserved in synclines. In areas where the folding has created anticlines, Jurassic rocks are exposed on the sea floor. The Hurd Deep has been cut in a Jurassic clay. An extensive outcrop of New Red Sandstone floors the sea bed south of the Cornubian peninsula. Since the New Red Sandstone was deposited, more than 190,000,000 years ago, the channel has been the site of a shallow epicontinental sea, in which later marine deposits accumulated in basins subject to intermittent downwarping. The uplands of Devon-Cornwall and Brittany, however, probably persisted as land areas for much of this time. Subaerial erosion of the rocks has contributed sediments which were laid down on the adjacent sea floor. The final phase in the geological history is concerned with the Pleistocene Ice Age; although the northwest European ice sheet halted at the Thames-Severn line, cold conditions persisted to the south. More important were the world-wide oscillations of sea level at this time and in certain periods the floor of the channel was dry to the present 50-fathom line. The final postglacial rise of sea level (Flandrian transgression) which gave the channel its present form was responsible for severing the land connection between Kent and France, creating the Straits of Dover (*c.* 5000 B.C.). The lower parts of valleys were submerged and in this way the fine natural harbours of Falmouth, Plymouth and Dartmouth were formed. Farther east Southampton water and Poole, Christchurch and Pagham harbours came into being as a result of drowning.

Coastal changes have modified the plan of the channel shore during the past 2,000 years. In Lyme bay the cliffs of softer clay and sandstone retreated steadily under subaerial and marine ero-

sion. Similarly chalk cliffs from Dorset to Kent have retreated, but erosion rates have been slower. Maximum erosion has occurred at Selsey Bill, where the sea cut back the low cliff a mile during historic time; Saxon Selsey was an important town with a large church which served as a cathedral until the 11th century. As a result of marine attack no trace of the church remains. Conversely the coast has been considerably extended at Dungeness where shingle has accumulated since Roman times. The former deep embayments at Poole, Christchurch and Pagham have been infilled as material brought down by rivers accumulated behind the sandspits.

Bottom Deposits.—Superficial sediments covering the sea floor are thin or absent over large areas, exposing the solid geology. Dominant deposits are shingle and coarse sand; these occur widely in the eastern two-thirds of the channel. Fine and medium-grained sands are widespread in the western approaches and the shallow coastal waters from the Solent to Dover. Mud is rare, forming small patches in Lyme bay and off the entrances to estuaries (Dart, Solent and Somme). Coarse bottom sediment tends to exist where the tidal stream is strongest. The variation in the maximum strength of the tidal stream from over four knots off the Cotentin peninsula to less than one knot near Plymouth has had a marked effect on the distribution and grading of the sediments. The paucity of bottom deposits in the channel reflects the recent geological history of the area. Whereas the North and Irish seas were covered by the Pleistocene ice sheet and received glacial debris, the English channel lay beyond the farthest extent of the ice front. Near the Straits of Dover some sediment originating in the North sea forms the banks of Varne, Vergoyer and Ridge. The other major sandbanks occur as eddy accumulations in the lee of protective headlands; *e.g.*, the Skerries bank near Start point and the Shambles off Portland Bill.

Tides and Tidal Streams.—Tides in the channel are best understood in terms of the standing oscillation hypothesis, which suggests that an oscillating movement is set in motion within the channel by the rise and fall of the oceanic water of the Atlantic ocean. The movement is balanced about a nodal line which runs approximately north–south, near the Isle of Wight, so that while the tide is rising in the western channel it is falling in the east. The shape of the channel, the bottom configuration and the gyroscopic forces acting on the water cause a degenerate amphidromic point (where the tidal rise is zero) to occur at a point inland northwest from the Isle of Wight. The cotidal lines, *i.e.*, lines where high water occurs simultaneously, converge on this point. High water becomes progressively later from Land's End to the Isle of Wight. Farther east, the time of high water along the south coast between the Solent and South Foreland is approximately the same; this is also true of the French coast between Cherbourg and the Seine estuary. The tidal range near the Isle of Wight is small but increases both west and east. The shape of the Gulf of St. Malo gives rise to tidal ranges exceeding 28 ft. at its head.

The tidal streams, associated with the tide, enter the western entrance, and flow parallel to the English and French coasts as the flood for approximately six hours. After high water the streams turn and flow in the opposite direction for the remainder of the tide cycle. In the marginal bays the effect of coastal configuration may cause the flood stream to turn after only four hours; *e.g.*, in Start bay where the ebb runs south for eight hours in every tide cycle. There is also a residual water flow upchannel because of prevailing southwesterly winds. The rate of this wind-generated surface current is variable but averages 2–3 sea miles per day. With gale force winds the rate of flow has been known to exceed 21 mi. per day (Jan. 1930). Occasionally the direction of flow upchannel and through the Straits of Dover into the southern North sea is reversed, particularly when the wind is in the northeast quarter. As a result of the continuous loss of water from the channel into the North sea, the 645 cu.mi. of water in the channel is completely replenished in about 500 days, its place being taken in an influx of water from the Atlantic.

Of the navigational aids and lighthouses the most important are at Land's End, Lizard Head, Eddystone rocks, Bill of Portland, Beachy Head and Dungeness in England; and at Ushant, Batz Island, the northwestern tip of the Cotentin peninsula (Fréhel and La Hague) and La Hève in France.

Temperature and Salinity.—Average temperature conditions of the surface waters are typical of a temperate region. Minimum temperature is about 44.6° F. (February and March) and rises to 60.8° F. (August and September). Water tends to be warmer in the western channel; it is noticeable that a warm-water wedge extends much farther east along the French coast than along the English coast. In summer the warming of the shallower coastal waters leaves a cool pocket of less than 59° F. in mid-channel compared with 61.8° F. nearer the coasts. In the shallower eastern half of the channel, tidal streams cause a mixing of the waters so that there is little change of temperature with depth. West of the Isle of Wight there is often a pronounced difference (up to 41° F.) between surface- and bottom-water temperatures.

Salinity varies between 34.8‰ and 35.3‰, compared with 35.6‰ for the Atlantic waters. This decrease is due to the freshwater influx from both coasts. Because of the greater catchment area of the French rivers (47,339 sq.mi. compared with 7,481 sq.mi. on the English side) the annual net influx of fresh water is 61,772 cusecs (cubic feet per second) from France and only 10,072 cusecs from England, allowance being made for the difference in precipitation in the two areas. As a result, salinities tend to be lower on the French side of the channel.

Marine Life.—The annual temperature range of the surface waters means that the channel lies almost on the boundary between sea areas dominated by cold-water and warm-water organisms. The fauna is therefore composed of both cold- (cod, herring, whiting) and warm- (hake, pilchard, mullet) water species. At Plymouth indigenous fish fauna consists of 44 cold- and 20 warm-water species. Comparable figures for jellyfish are 23 and 17 respectively. The channel is therefore an important fishing area (Plymouth, Dartmouth, Brixham, Mevagissey, Fowey and many other smaller fishing villages). Along the north coast of Brittany numerous small ports like St. Malo, L'Aberwrach, Roscoff, Perros-Guirec and Paimpol are engaged in both inshore and offshore fishing. Since 1930 the herring has virtually disappeared and been replaced by the pilchard. This change has coincided with alterations in phosphorus and plankton content of the waters. A further factor has been the amelioration of the northern climate, causing the boundary of warm-water species to move poleward. The number of exotic fish caught in the channel has also increased.

History.—From early times the channel has served as a route by which invaders reached Britain from continental Europe. In prehistoric time successive waves of immigrants arrived with their distinctive cultures. Some came by the Straits of Dover bringing with them their improved stone implements for felling trees and tilling the soil. Another group, of Mediterranean origin, journeyed coastwise around Spain and the Bay of Biscay to Brittany and made the longer crossing into Devon and Cornwall. They left behind the chambered burial tombs found in many parts of western Britain. As these Megalithic immigrants developed a settled mode of life based on agriculture they began to trade, chiefly in metals like copper, tin and lead, with their continental neighbours. Thus the earliest cross-channel trading traffic was via the broader western part of the channel. The western route was also used by the Bronze Age immigrants (2000 B.C.), who arrived in southern Britain. They were to be followed by iron-using invaders, men differing in physical appearance from the earlier Neolithic settlers. All were to fuse to form the Britons who were later to face the Roman invaders beginning with Caesar's incursion in 55 B.C. These crossings of the channel, which was called Mare Britannicum or Oceanus Britannicus, and those that followed were made from Portus Itius (Boulogne) to a point near Deal. The Romans began the first large-scale land reclamation, that of Romney marsh behind the shingle of Dungeness. Another surviving example of the Roman occupation is the system of coastal fortresses built about 300 A.D. in southeast England to repel the piratical attacks of Saxon invaders. Of the ten fortresses under the unified command of the count of the Saxon Shore, six bordered the channel coast: Richborough, Dover, Hythe, Pevensey, Porchester and possibly one at Carisbrooke in the Isle of Wight. The gradual weaken-

ing of the power of Rome caused the loosely knit system of coastal defense to collapse and the channel coast was again at the mercy of fresh invaders from Europe. In the centuries which followed the Saxon colonization the West Saxons gradually extended their influence along the whole of the channel shore. The last time that the channel acted as the invader's route to England was in 1066 when William the Conqueror crossed from Normandy to Hastings. This area offered natural advantages for establishing a bridgehead from which a landward advance could later be made. Hastings, with its small harbour, lay at the seaward end of a peninsula bordered on the north by the natural defensive line of the Brede estuary and southwest by the embayment of the Bulverhythe. Following the conquest, the channel became a much-used route, but with the subsequent loss of Normandy it reverted to its role as a line of defense. In this setting the Cinque Ports (*q.v.*) were established under Henry II. The establishment of a navy on a permanent basis followed in the 16th century; its first major sea battle was against the Armada (*q.v.*) in 1588. The importance attached to sea defense led to the establishment of naval dockyards at Portsmouth and Plymouth. These naval bases were factors in deterring Napoleon and were also significant to British defenses in World Wars I and II. In contrast Southampton developed as a commercial port; its deepwater harbour was superior to the silted-up Cinque Ports. Other small towns along the south coast developed rapidly as resorts during the 18th and 19th centuries. Brighton, Weymouth and Worthing were helped by royal patronage, and later Bognor Regis. Bournemouth developed much later. In the southwest peninsula fishing hamlets like Torquay, Brixham and Fowey shared in the expansion resulting from the popularity of seaside vacations. Along the French channel coast are a number of fashionable bathing resorts including Dinard, St. Malo, Deauville, Trouville and Le Touquet-Paris-Plage.

In World War II the channel was the scene of the Dunkerque evacuation and during hostilities assumed its traditional role of main line of defense to England. In 1944 the Allies forced a crossing for the invasion of Normandy.

John Jeffries and Jean Pierre Francois Blanchard first crossed the channel by balloon in 1785. In 1959 the British hovercraft, generating a cushion of air and skimming clear of the waves, crossed from Calais to Dover in two hours, three minutes. Since Capt. Matthew Webb first swam across in 1875 there have been many attempts, some successful, to repeat his performance (*see* SWIMMING; *Channel Swimming*).

Cross-Channel Routes.—There are regular sea-ferry services between Folkestone and Boulogne and from Dover to Calais, Dunkerque and Ostend. Shipping is mainly from Southampton to St. Malo, Cherbourg and Le Havre, and from Newhaven to Dieppe. There are also services from Portsmouth, Plymouth, Exmouth and Weymouth. Air-ferry services are frequent, especially during the summer.

The Channel Tunnel.—The century-old idea of a tunnel bored under the sea bed of the Straits of Dover was revived by the setting up in 1957 of a channel study group. The tunnel would run from near Dover to Sangatte on the French coast. Much original research was devoted to new techniques of sounding, sampling and micropaleontology. The observations have shown that the Lower Chalk, 230 ft. thick, would provide the best rock for boring, being nonporous and sufficiently compact not to require support. In the Middle Chalk above and the Gault Clay below tunneling would be more expensive but not impossible. It seemed likely that the greater part of the tunnel could run through the Lower Chalk; only off Dover and Sangatte were there likely to be difficulties, due to upfolds in the strata. The total cost was estimated at about £80,000,000. Other proposals studied were for two immersed tubes resting on or partially buried in the sea bed and for a bridge. A tunnel would carry a railway rather than a road, mainly because of the capacity and earning potential.

BIBLIOGRAPHY.—A. H. W. Robinson, "The Floor of the British Seas," *Scot. Geogr. Mag.*, vol. 68 (1952); W. B. R. King, "The Geological History of the English Channel," *Quart. J. Geol. Soc. Lond.*, vol. 110 (1954); O. Pratje, "Die Bodenbedeckung des Englischen Kanals," *Dtsch. hydrogr. Z.*, vol. 3 (1950); *The Admiralty Manual of Tides;* F. S. Russell, "The English Channel," *Transactions of the Devonshire Association*, vol. 85 (1953); J. A. Williamson, *The English Channel: a History* (1959). (A. H. W. R.)

ENGLISH COLLEGIATE STYLE, in architecture, a style developed in the colleges of Oxford and Cambridge during the 15th, 16th and 17th centuries. It is characterized by the persistence of Gothic tradition (*see* GOTHIC ARCHITECTURE) with only slight modifications due to the creeping in of Renaissance elements (*see* RENAISSANCE ARCHITECTURE); the grouping of buildings around quadrangles, entered by an arched gate in a tower, the combination of many window lights into one mullioned and transomed opening; parapets, sometimes battlemented; long rows of similar small gables; bay or oriel windows; and the frank exterior expression of the various types of halls, large rooms and small sleeping rooms which form the college. In general it is based on the domestic architecture of the time, in earlier examples resembling late Gothic manors and, in the later, the larger Elizabethan or Jacobean country houses; during the 17th century classic motives, often of great quaintness and crudeness, are general in the detail of doors, etc., without seriously changing the spirit of the entire style.

ENGLISH HISTORY. The following is a general account of the history of England from the second half of the 5th century A.D. Events anterior to the period of the Saxon migrations are covered in the article BRITAIN. The history of other parts of the British Isles is mentioned only to the extent necessitated by its relation to English history; for a fuller account reference should be made to the *History* sections of the articles IRELAND; IRELAND, NORTHERN; IRELAND, REPUBLIC OF; SCOTLAND; WALES. The separate articles on many of the Anglo-Saxon rulers, on all English monarchs since William I and on the country's principal statesmen supplement the information contained in the general narrative. (*See* also ENGLAND, CHURCH OF; ENGLISH LAW.)

The main divisions of this article are as follows:

I. Anglo-Saxon Period
II. The Normans (1066–1154)
III. Early Plantagenets (1154–1216)
IV. Economic and Social Development in the Early Middle Ages
V. The 13th Century
VI. The 14th Century
VII. Lancaster and York (1399–1485)
VIII. Economic and Social Development in the Later Middle Ages
IX. The Tudors (1485–1603)
X. The Early Stuarts and the Commonwealth (1603–60)
XI. The Restoration and the Revolution (1660–1714)
XII. The First Hanoverians (1714–60)
XIII. Reign of George III to 1789
XIV. Economic and Social Development, 1603–1789
XV. Revolution and Reform, 1789–1837
XVI. The Victorian Age, 1837–1901
XVII. The 20th Century

I. ANGLO-SAXON PERIOD

THE MIGRATION TO BRITAIN AND THE SETTLEMENTS

The central theme of Anglo-Saxon history is the process by which a number of diverse Germanic peoples settled in Britain during the 6th century came to form the centralized kingdom which the Normans inherited from their English predecessors. According to English traditions, the migration to Britain was preceded by the descent of small companies on different parts of the British coast in the second half of the 5th century. The one contemporary account of these events written by a Briton, the *De excidio et conquestu Britanniae* ("Destruction and Conquest of Britain") of the ecclesiastic Gildas, makes them originate from an invitation sent by an unnamed British prince to certain "Saxon" adventurers to enter his service and defend his country against Pictish and Irish raiders. Later British and English traditions give the name Vortigern to the British prince, and state that the leaders of the Saxon auxiliaries were two brothers named Hengist (*q.v.*) and Horsa. A revolt of these mercenaries, who had been joined by many others of their own race, was followed by devastations which spread as far as the western sea, by the creation of new British forces after the original mercenaries had dispersed to their own countries, and by a conflict between the Britons and fresh invaders which culminated soon after the year 500 in a British victory at a place called Mons Badonicus,

BY COURTESY OF GEORGE PHILIP & SONS LTD.

ENGLAND IN THE ANGLO-SAXON PERIOD, SHOWING PLACES MENTIONED IN THE TEXT AND THE APPROXIMATE BOUNDARIES OF THE KINGDOMS IN THE 8TH CENTURY

as yet unidentified. Thenceforward for more than a generation the invaders were confined to the southeastern coastlands which they had occupied in the earlier phases of the war. Many of them returned to the continent and obtained land for settlement from the kings of the Franks. The remainder formed the nucleus of the earliest English kingdoms.

How far these primitive kingdoms stretched into the interior of the country is an open question. The historical evidence leaves no serious doubt that the kingdoms of Kent (*q.v.*) and Sussex had come into being before the end of the 5th century and that the dynasty which afterward appeared in Wessex (*q.v.*) was descended from leaders who belong to this period. But of the other southern kingdoms—Essex, East Anglia, Mercia and Lindsey—and of the northern kingdom of Deira (*qq.v.*), corresponding roughly to central and eastern Yorkshire, no traditions of origin have been preserved. In the farthest north, the foundation of the kingdom of Bernicia (*q.v.*) was assigned by a reasonable approximation to the year 547.

For the beginnings of settlement in the eastern half of England there is some evidence to be obtained from the distribution of heathen English burial grounds, and English place names of an archaic type. It suggests that by the year 500 permanent settlements had been founded well to the rear of the entire coast line from Flamborough head to Southampton water, and that in some districts, notably the valleys which merge into the estuary of the Wash, the English had penetrated far inland.

The expansion which carried the English peoples to the mastery of southern Britain began in the middle of the 6th century. Between 552 and 584 it had brought the West Saxons, the one people whose early history can be traced in outline, from the Wiltshire river Avon to the river Severn at Gloucester, the river Lea at Luton, and the hills above the valley of the upper reaches of the Great Ouse. They were unable to retain the whole of this vast territory. Within a generation they had lost its outlying portions to the rival kingdom of the Mercians, finding their own destiny thenceforward in a gradual extension of their power toward the southwest. But the midland country which they had occupied was never reconquered by the Britons. In the north the English of Bernicia, after a generation of precarious existence between the Pennines and the North sea, suddenly became a formidable power under King Aethelfrith (*q.v.*; 593–616), who repelled attacks from both Gaels and Britons, annexed the kingdom of Deira and toward the end of his reign won a long-remembered victory over the British king of Powys at Chester. It is with these events that the history of the English peoples in Britain becomes at last a matter of consecutive record.

Origins of the Invaders.—The English tradition of the origins of their race was summarized in the 8th century by Bede (*q.v.*), the greatest of Anglo-Saxon historians. His analysis derived the English people from three Germanic races, whom he called *Iutae*, *Saxones* and *Angli* (*see* JUTES; SAXONS; ANGLES). By the *Angli* he meant the men of Angel (now Angeln) in Schleswig. He placed the *Saxones* to the south of them and the *Iutae* to the north, that is, in the country afterward known as Jutland. In this he was probably mistaken, for the culture of the Jutes in England disassociates them from the other northern peoples and suggests that they had migrated to Britain from the Rhineland. (For a discussion of this problem, *see* KENT, KINGDOM OF.) From the *Iutae*, Bede derived the inhabitants of Kent, the Isle of Wight, and the mainland districts facing the island. From the *Saxones* he derived the East, South and West Saxons of his own day; the remaining English peoples, of whom he mentioned the East and Middle Angles, the Mercians and the Northumbrians, being descended from the *Angli*. Modern criticism has shown that this analysis is too simple. Other races than the three which Bede names—the Frisians, for example—certainly took part in the migration to Britain. Intertribal movement among the English peoples within Britain itself must have caused an early fusion of different racial elements. It is remarkable that in spite of these complications there is sufficient evidence from archaeological remains and ancient place names to suggest that in its main outlines Bede's famous survey was substantially correct.

It was certainly more than a literary reconstruction by a monk writing in isolation. Bede himself stated that he had sent a draft of his history to Ceolwulf, king of Northumbria (d. 737), "to read and examine . . ." Its statements therefore carry the approval of a Northumbrian king in an age when courts were the chief repositories of national traditions. The English court tradition about the original home of the *Angli* was in fact remarkably persistent. King Alfred (d. 899), a century and a half after Bede, places them in the Jutish peninsula and the islands to the east of it. All this gives weight to Bede's incidental statement that Angel, from which the *Angli* came to Britain, "is said to have lain empty from their departure to the present day." Thus to Bede and his first readers the English possession of southern Britain was the result of the mass migration of entire peoples. That the way was opened for their settlement by small groups of adventurers, however, is clear from fragments of information preserved in other sources about their clashes with the British defenders of the land; and the number of leaders who succeeded in establishing themselves as English kings is remarkable. The essential qualification for kingship among the northern peoples of the migration age was the ability to claim descent from one or other of the ancient gods of the Germanic race. Such claims are known to have been made by the kings of the Mercians, East Angles, East Saxons, Kentishmen and West Saxons, by the kings of Lindsey, who play no part in recorded history, and beyond the Humber by the kings of the Deirans and Bernicians. The East Saxon genealogy goes back to a god of the continental Saxons named Seaxneat; the other genealogies go back to Woden (Odin). Until the 8th century, and even later, it was felt that anyone sprung from such god-descended stock was entitled to rule over at least a portion of the kingdom to which he belonged. Much disorder was caused in these early centuries by individuals trying to make good their claims.

These details give the impression that the English conquest of Britain was the result of a considerable number of independent migrations led by chiefs of approximately equal rank. The English

peoples preserved no tradition of any one outstanding leader in their earliest history, and there are facts which suggest that the occupation of Northumbria proceeded independently of the movements which led to the colonization of the south. On the other hand, during most of the period between the first settlements and the Danish wars of Alfred's time, the southern kingdoms normally formed a loosely compacted federation in which their individual rulers accepted the authority of a common overlord. Bede, the ultimate authority for this information, names seven kings who successively ruled all the southern provinces of the English people as far as the Humber, the first of whom was Aelli (*q.v.;* fl. late 5th century), king of the South Saxons. The second was Ceawlin (*q.v.;* 560–593), king of the West Saxons, under whom that kingdom reached the most northerly limits of its early expansion. The third was Aethelberht of Kent (560–616), in whose time the southern kingdoms had begun to assume the outlines of their permanent shape. A 9th-century annalist who copied Bede's list states that this overlordship brought the title *bretwalda* to its possessor. The word, which means "ruler of Britain," belongs to the complimentary language commonly addressed by poets to patrons who entertained them at court. But it indicates the range of the superiority allowed to their overlords by the lesser kings who combined to form these primitive confederacies.

The Introduction of Christianity.—Under Aethelberht of Kent (*q.v.*), the third of these overlords, new factors began to play in English history. A considerable part of the Jutish race had remained on the continent, and in the 6th century had come to accept the overlordship of the kings of the Franks. The wish to strengthen Frankish influence over the insular as well as the continental Jutes is the simplest explanation of the fact that Aethelberht was allowed to marry the daughter of Charibert (d. 567), king of Paris. The consequences of the marriage were of the first importance, for it secured an honourable reception to the mission despatched in 596 by Pope Gregory I the Great for the conversion of the English peoples. A friendly court was essential for the success of the enterprise, for popular heathenism (*see* GERMANIC PEOPLES: *Conversion to Christianity*) was strong throughout southeastern England. Aethelberht's supremacy over the other southern kings gave security to Augustine (*q.v.*), the leader of the mission, as he traveled over the country. Within his own kingdom, Aethelberht's support of Augustine led to the superficial conversion of the entire Kentish people and made possible the foundation of a cathedral and a suburban monastery at Canterbury and the creation of a second Kentish see at Rochester (604). The necessity of enforcing respect for the ministers and property of the church caused Aethelberht to issue a code of law for his kingdom, which after its first clause passed beyond its original purpose to lay down a tariff for all manner of offenses, and makes Aethelberht the author of by far the earliest piece of legislation promulgated in any Germanic language.

The tradition of Aethelberht's patronage of Augustine gave to the kingdom of Kent a distinction which always set it apart among the members of the southern English confederation.

THE HEPTARCHY

The Era of Northumbrian Ascendancy.—But politically, the Kentish kingdom was not strong enough to keep the position to which Aethelberht had raised it. Even before his death in 616, supremacy in southern England was passing from Kent to East Anglia under Raedwald (*q.v.;* d. between 616 and 628), the fourth in Bede's list of overlords. In the north, a greater Northumbria was coming into being through the annexation of Deira by Aethelfrith, king of Bernicia. The first recorded trial of strength between northern and southern England was an attack on Aethelfrith by Raedwald in the interest of his protégé Edwin, the exiled heir of the Deiran kingdom. In a battle on the river Idle, near Bawtry, Aethelfrith was killed (616). Edwin (*q.v.;* d. 632) became king of all Northumbria, and within a few years had been recognized as overlord throughout southern England, apart from Kent, with which he had a special relationship through his marriage to Aethelberga, King Aethelberht's daughter. Even so, he came much nearer than any previous king to a position of authority over all England. His reputation was great among the northern peoples beyond the English border—the northern Britons, the Picts and the Gaels of Argyll. As king of Northumbria, he annexed the British district of Elmet (in south Yorkshire) which had hitherto maintained an independent existence between the northern and southern English kingdoms. He established his people along the northwestern coastlands, occupied the Isle of Man and led a successful expedition against the Britons of north Wales. To Bede, who was not given to flights of imagination, he was lord of an empire extending over the whole of Britain.

This ascendancy produced a natural reaction. Edwin was overthrown in battle (632) on his southern border by Cadwallon, king of Gwynedd, and a member of the Mercian royal house named Penda (*q.v.; c.* 632–654), who was carried by the victory into the Mercian kingship. For a year, Cadwallon and his allies gave themselves to a deliberate harrying of all Northumbria. The kingdom was restored by Oswald (*q.v.;* 633–641), son of Aethelfrith of Bernicia. Oswald's destruction of Cadwallon and his army in 633 made him the outstanding figure among the kings of his time. His reputation among the kings of the further north was as great as that which Edwin had once possessed. Among the peoples of the southern English confederacy he is known to have had direct authority in Lindsey, Wessex and Sussex. By some of these people his power was resented. In Lindsey he was remembered as "a king sprung from another province who had taken rule over them." In Mercia, where Penda by now was king, Oswald was regarded as an enemy, and in 641 he was defeated and killed by a Mercian army.

After this battle Northumbria fell apart into its two ancient divisions of Deira and Bernicia. In Bernicia, Oswald was succeeded by his brother Oswiu (641–670) who for 13 years was confronted by rival kings in Deira, and was kept in enforced subordination to Penda by the surrender of hostages. This uneasy equilibrium was ended by one of the most obscure episodes in Old English history. In 654, for no discernible reason, Oswiu was attacked by Penda at the head of a large army to which many English and British kings and princes had brought their followers. Penda was defeated and killed in a battle fought in the neighbourhood of Leeds, and Oswiu, reuniting Deira and Bernicia, was able to annex northern Mercia to his own kingdom and to re-establish Northumbrian ascendancy in the south. Three years later, a revolt in Mercia destroyed this settlement, and with it the possibility that England might be united politically under Northumbrian leadership.

Even in the south, the movement toward unification was suspended for a considerable period. In course of time, Wulfhere (*q.v.;* 657–674), son of Penda, was able to secure recognition as overlord from the other southern kings, whose forces he led in an attack on Northumbria in 674. But it failed disastrously, and for the next 50 years there is no sign that any southern king imposed his supremacy on his fellows. In Northumbria the dominant figure in this period was Egfrith (*q.v.*), son of Oswiu, who extended his kingdom by the conquest of much territory from the Picts, but in 685 was defeated and killed far from his own country at a place called Nechtanesmere near Forfar. The Northumbrian frontier was thrown back to the Forth, and was only held with difficulty along this line. There was much fighting between Englishmen in different parts—in Mercia, Kent and Wessex—but none that made much impression on subsequent events. It is only in Wessex under King Ine (*q.v.;* 688–726) and in Kent under King Wihtred (*q.v.;* 690–725) that there are signs of any consistent attempt at the consolidation of individual kingdoms. The laws which Ine issued for this purpose in Wessex form the principal authority for the structure of the earliest English society (*see* ANGLO-SAXON LAW; ENGLISH LAW).

The Pattern of Society.—Meanwhile, behind the screen of these incessant wars, the pattern of English country life was slowly assuming what was to be its permanent form. There is little contemporary evidence to illustrate the process, which can only be traced by inference from place names and archaeological discoveries, on the one hand, and, on the other, by argument from the laws and charters, histories and biographies of a later age.

Extent of British Survival.—On the fundamental problem of the extent to which a British population survived the English conquest, the evidence is remarkably consistent. British workmanship is displayed on many articles found on English burial sites. Celtic names of rivers, woods and hills appear sporadically in all parts of the country except the eastern coastal fringe. A respectable though subordinate place was assigned to Britons of various ranks in the laws of Ine. But when due allowance has been made for these traces of British survival, they remain of minor significance when set beside the overwhelming evidence for the Germanic origin of the English people as a whole—the character of the most ancient English place names, the language of the earliest English texts, the technicalities of the oldest English laws and the social system which they reflect, and the primitive organization of English agriculture. In all its essential features, it is a Germanic society which emerges from the confusion of the English settlement.

The Ceorl.—The basis of that society was the peasant, the ceorl of the earliest English laws. He appears in early documents as the head of a family comprising both slaves and free servants, and supported by a holding which he farmed in close association with others of his kind. The holding was known in Old English as a *higid* or *hid* (*see* HIDE), an archaic word which meant a family's land and covered arable, meadow and rights of common over waste and woodland. There is no early evidence for the normal size of the primitive hide, which within any given district may have varied widely between one kind of terrain and another. In the eastern counties it seems to have been about 120 ac.; in Wessex perhaps only 40 ac. Whatever the extent of his holding, the ceorl was burdened with many public duties and charges. He owed military service in person. His holding was charged with a payment to the communal *feorm* or food rent levied for the maintenance of the king's household. Similar contributions were required by the king's chief ministers, or ealdormen. Combined with charges for the entertainment of strangers coming to court, with the supply of food for the king's household men and hunt servants, with cartage services and the provision of labour for the king's works, these requirements were a serious drain on resources which can never have been greatly in excess of the necessities of life. Nevertheless, the ceorl's position in society was honourable. His life, in Wessex at least, was valued at a wergild, or payment to his relations if he were slain, of 200 shillings—in Kent, at 10 times this sum—and penalties were exacted by public authority for violence against him and the members of his family, and for the breaking of the hedge round his farm. He was sharply distinguished in law and social custom from the men who by birth or through service to the king ranked as noble. Their lives were valued at 1,200 shillings in Wessex—the early wergild of the Kentish nobleman is not precisely known, although later it was 300 golden shillings, three times that of the Kentish ceorl—and the distinction reflected by these figures runs through the whole of the Old English social order. In course of time, through the pressure of economic forces, the alienation of royal rights over peasant holdings to noblemen and churches and the devastation of repeated wars, innumerable ceorls, once independent, were brought under the authority of private lords. Before 1066, the manor inhabited by peasants grouped in a routine of agricultural service around a lord's house had become the typical form of rural community. The beginnings of the process by which this came about can be seen in Ine's laws of the late 7th century. But the typical ceorl of whom they speak was still the independent head of a peasant household (*see* CHURL).

The Gesith.—The ceorl's round of life was spent beneath the surface of recorded history. The figures who take the centre of the historian's view are the companions of kings: the members of their households, their agents in public affairs, their guardians in peace, who formed their retinues in war. They were known as *gesithas* in English and as *comites* in Latin. It was of his own free will that the gesith had entered the service of the king his lord, and when the king died the gesith's allegiance was not necessarily given to the king's son. Nevertheless, while it was in force the bond between the gesith and his lord was at least as strong as that which united him to his own kinsmen. Its violation was the gravest of crimes and its observance, even at the cost of life or freedom, was the highest of virtues. The king, for his part, was expected to provide his gesiths, after good service, with a grant of land; and a son who had proved himself able to fill his father's place among the king's companions might hope in time to succeed his father upon his holding in the country. Except in the earliest periods it is unlikely that a gesith's endowment consisted of a stretch of unbroken land. In the oldest evidence the king seems to be giving to those of his companions who had completed their service his own rights to food rents and services over villages already in being. Whatever the nature of the gesith's primitive endowment may have been, by the end of the 8th century grants of villages to men who have earned favour by service to the king are frequently expressed in royal charters which give a hereditary estate to their recipients. One of the main foundations of the later manorial system had been firmly laid.

The Growth and Achievements of the Church.—The royal charter had been introduced into England in the third quarter of the 7th century as a means of securing religious houses in the possession of their estates. Its appearance as an instrument of social change is a good illustration of the way in which the primitive fabric of English society was modified through the conversion of the English peoples to Christianity. The course of this movement, which had reduced English heathenism to insignificance within two generations from the landing of Augustine, had been interrupted by relapses caused by war and pestilence. It had been checkered by the clash between Roman and Celtic conceptions of religious order and ritual (*see* CELTIC CHURCH) arising from King Oswald's introduction of Irish monks into Northumbria after the original Northumbrian church founded by Pope Gregory's disciple Paulinus (*q.v.*; d. 644) had been destroyed by Cadwallon of Gwynedd and his heathen allies. But there had never been any general relapse of the whole English people; the northern church was brought into conformity with Roman organization and discipline at the Synod of Whitby (663–664), accepting the Roman tonsure and its method for calculating the date of Easter. Augustine's see of Canterbury, after declining into a local Kentish bishopric, was restored to full metropolitan authority by Theodore (*q.v.*), archbishop of Canterbury from 668 to 690, a monk of Greek birth and education appointed at Rome in an emergency by Pope Vitalian. It was Theodore's great achievement to adapt the Roman ideal of a centralized church to English conditions. The councils, which illustrate this aspect of his work most impressively, gave to the English church an effective means of expressing its unity and of defining its attitude to the theological controversies of the age. But at his accession the church itself had not entirely lost the character of a missionary enterprise, and it is on his measures for the improvement of its permanent organization, such as the division of unwieldy dioceses which had represented ancient kingdoms, that Theodore's importance as an ecclesiastical statesman chiefly rests. (*See* also ENGLAND, CHURCH OF.)

The tradition of a centralized church under an archbishop of Canterbury in close alliance with secular rulers was maintained by Theodore's successors. Between his death in 690 and the creation of an archbishopric of York in 735 the organization of the church demonstrated the essential unity of the English peoples in a way that no secular authority could rival. During most of this period, southern England formed a congeries of independent kingdoms, and the unification of England under Northumbrian lordship, which had once seemed possible, was beyond the power of kings now preoccupied by the insecurity of their northern frontier. Northumbria continued as a world apart, but the situation in the south after the resignation of Ine in 726 brought supremacy in those parts within the reach of any king qualified by birth and reputation. In 731 Bede stated without reservation that all the provinces of the English people as far north as the Humber, with their kings, were subject to Aethelbald (*q.v.*), king of the Mercians.

The Rise of Mercia.—By this date, the Mercian kingdom covered the whole of the Midlands from the marches of Wales, where it had arisen, to East Anglia. It was a composite state of many provinces held together by a line of kings descended from an-

cestors who had been ruling in Angel before the migration to Britain. They were a fertile race and each of their many branches asserted, at one time or another, a successful claim to rule over all Mercia. Aethelbald, who had been compelled to live in exile by the reigning king, Ceolred, succeeded him in 716. Nothing is known of the stages by which he obtained overlordship throughout the south, but the range of his power is beyond question. Outside his own kingdom there is direct evidence of his authority in Wessex, Kent and London. A charter (749) by which he granted a general exemption to monasteries and churches from all taxes and public works except bridge and fortress building is important in the history of English ecclesiastical immunities. His own way of life gave him a bad reputation among churchmen who, however, acknowledged the good peace which he kept in his kingdom; but he signally failed to keep the loyalty of his household companions, by whom he was murdered at night early in 757.

Ascendancy of Offa (757–796).—Before the year was over, after a violent contest with a rival claimant, Offa (*q.v.*), son of Thingfrith, a remote cousin of Aethelbald, began one of the outstanding reigns of Old English history. No consecutive account of it can be written, for Offa never attracted a biographer and no Mercian annals have come down from his time. It is all the more remarkable that the disjointed materials for the period give a clear impression of a great but unsympathetic ruler with an international outlook and civilized interests, able to carry through works on the largest scale for the protection of his own country, but ruthless in his dealings with recalcitrant underkings and capable of all extremities of violence to secure his son's inheritance. In his hands overlordship of all the territories between the Humber and the English channel could become indistinguishable from direct rule. In Kent, where the process can be followed most clearly, it gave bitter offense to nobles and churchmen conscious of the peculiar distinction of their kingdom, and it was the mutual hostility thus aroused which caused Offa to enter into the negotiations with Rome which led to the creation (788) of an independent archbishopric of Lichfield, which did not, however, last beyond 802. The inevitable reaction came immediately after his death (796). No later Mercian king approached the range of Offa's power, but he had shown that the unification of southern England was a possibility.

Offa's Dike.—The effectiveness of his government is most clearly shown by the great earthwork along his western border known as Offa's dike, which has kept his name alive for more than 1,000 years. There was no longer the possibility of a British revival which could threaten the English mastery of southern Britain. But between 705 and 709 a number of devastating raids, which had caused alarm as far to the east as Crowland, Lincolnshire, had shown the weakness of the Mercian defenses. They may well have caused the building of the large-scale earthwork known as Watt's dike, which closed a dangerous gap, unprotected by any natural obstacle, between the middle reaches of the river Dee and the upper waters of the river Severn. Adequate for the defense of northwest Mercia, it was soon superseded by an entrenchment covering every stretch of the Mercian frontier that was not protected by ancient woodland, and blocking every trackway from Wales into the midlands. The attribution of this work to Offa was a matter of common knowledge already in the 9th century, and is undoubtedly correct. The skill with which his dike was planned in relation to the terrain confronting the chosen line, and the assurance with which the great design was carried out, give it a unique place among the monuments of its age and kind in England or, indeed, in northern Europe.

Relations With the Continent.—It was natural that a king who could show his power so effectively in his own country should wish to display it also on the continental stage. The link between England and Rome created by Pope Gregory I, though often attenuated, was never broken. Missionaries of English birth, notably the Northumbrian Willibrord and the West Saxon Boniface, had recently been devoting themselves under papal surveillance to the conversion of Frisia and western Germany. The large measure of success that they achieved was made possible by English support. Gregory's successors for long considered England as the province of a young church to which they owed protection, and among the English evangelists abroad there were many who regarded themselves as in some degree responsible for the moral condition of their own countrymen. But in the second half of the 8th century Anglo-continental relationships came to be determined rather by political motives and were governed by the attitude of rulers toward each other, allowing room for the play of economic interests. It is in the reign of Offa that the change first becomes apparent, and its origin lies in Offa's reaction to the new dynasty established beyond the English channel by the coronation of Pepin (Pippin) the Short as king of the Franks (751).

In antiquity and reputation, the family to which Offa belonged was outstanding among the reigning houses of northern Europe. Heroic verse remembered the exploits of his ancestors, among whom his namesake Offa, king of Angel in the 4th century, anticipated the reputation that his descendant was to win by creating a permanent boundary for his kingdom. Ancestry of such distinction was always a source of strength to an ambitious king, and Offa of Mercia was expressing political values in an important Kentish charter which enlarges the title of his Mercian kingdom with the statement that he was sprung from the royal race of the Mercians. On the other hand the new kings of the Franks, though their ancestry was ancient and noble, could not claim royal descent. Their title to rule rested on the sanction given to their authority by the religious ceremony of consecration to kingship and the coronation that was its sequel. It was certainly in imitation of this newly established Frankish practice that Offa in 787 caused Egfrith his son to be consecrated king. By this time Pepin's son Charlemagne had raised the Frankish kingdom to a height of power against which Offa's overlordship in southern England seems insignificant. But Offa himself was punctilious in insisting that his relations with Charlemagne should be maintained on equal terms. A proposal (*c.* 789) that Offa should give a daughter in marriage to a son of Charlemagne was countered by a proposal that a daughter of Charlemagne should be given to Offa's son. In anger at this reply, Charlemagne closed all Frankish ports to English traders and maintained the prohibition for at least three years. But in 796, the last year of Offa's life, Charlemagne was writing again to him as to an equal, recalling the friendly exchanges which had formerly passed between them, and agreeing to Offa's proposal that each of them should give his special protection to traders coming from the other's country.

Relations With the Papacy.—Offa's relations with the papacy show the same consciousness of mastery within his own insular world. He was prepared to allow to the pope's authority over the English church a range which it had not yet found in England. But he caused it to serve his own interests in a manner which no earlier English king had attempted. There seems no recorded precedent for the letter of Pope Adrian I granting to Offa and Cynethryth his wife possession of the monasteries which they had founded or acquired. It was with Offa's approval and probably at his invitation that the same pope dispatched a legatine mission to England such as had not been seen there since the time of Augustine. But it cannot be a coincidence that the mission was closely followed by the pope's assent to Offa's request for the creation of an archbishopric of Lichfield, which would emancipate the Mercian churches from the authority of an archbishop seated among Offa's enemies in Kent. At Rome itself Offa was regarded as an enigmatical figure, of whom the pope would be well advised to take account. An isolated letter from Pope Adrian to Charlemagne laboriously disclaims belief in a rumour which had reached the pope that Offa and Charlemagne had been planning his deposition and the substitution of a Frankish pope in his stead.

Improvements in the Coinage.—But the chief legacy of Offa's reign resulted from economic pressures with which policy had little to do. It consisted in a radical alteration of the English currency, which served the interests of traders and was in line with the whole tendency of his government. Up to the middle of the 8th century, in both France and England, the issue of a coinage was left in the hands of individual moneyers. In France the moneyer's name upon a coin was the recognized guarantee of its weight and the purity of its metal, to which the name of a king,

when it appeared, was an unessential supplement. In England the coins of this period, known to numismatists as sceattas, were normally uninscribed. In 755 Pepin assumed responsibility for the Frankish coinage, which thereafter bore the royal initial and, under Charlemagne, the royal name. In England, the corresponding change was made by certain obscure kings who were reigning in Kent soon after the year 770. Their subjection to Offa gave him control of the mint of Canterbury, where he issued under his own name a series of pieces already, it seems, called pennies, which were admirable both in design and execution. The series reaches its highest artistic level in coins bearing the portrait of Offa or of Cynethryth his queen, which offer a startling contrast to the crudely executed coins put out for the Frankish kingdom by Pepin or Charlemagne.

BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

OBVERSE AND REVERSE OF THE ONLY EXTANT EXAMPLE OF OFFA'S GOLD DINAR, THE COPY OF AN ABBASID CALIPH'S COIN, STAMPED WITH OFFA REX AND THE ARABIC SUPERSCRIPTION; SECOND HALF OF THE 8TH CENTURY. IN THE BRITISH MUSEUM

No western government of the dark ages could have maintained for long a currency of such artistic merit. The coins of Offa's successors show a rapid decline in craftsmanship, but always preserve the main features on which he had insisted—the names of king and moneyer. To this double authentication is due in great part their excellent quality and their value as sources of historical information.

The Decline of Mercia.—The 50 years after Offa's death in 796 form an anticlimax to a reign which for all its internal violence had brought England effectively onto the continental scene. Egfrith, Offa's son, died before the year was out. He was followed successively by two kings (Coenwulf, 796–821 and Ceolwulf, 821–823) with whom, apparently, the Mercian dynasty ended. In the meantime a Kentish rebel, Eadberht Praen, said to be an apostate priest, had maintained himself as king for two years (796–798). Wessex, which Offa had attached to his confederacy by the marriage of a daughter to its king, Beorhtric (d. 802), had broken away under a member of its royal family named Egbert (*q.v.*), who was regarded in his own country as the rightful heir to Kent and a number of adjoining regions. For more than 20 years there is no record of war between Wessex and Mercia, but Egbert never became the man of any Mercian king, and the loss of the provinces which he ruled seriously weakened the artificial state which Offa had founded. In Northumbria, as in Wessex, Offa had married (792) a daughter to the reigning king, Aethelred, but his murder in Offa's last year was followed by revolutions which enabled Charlemagne, for once, to intervene directly in English affairs. In 808 he brought about the restoration of a Northumbrian king whom his subjects had driven into exile. In the ecclesiastical sphere, veneration for Augustine's church defeated Offa's revolutionary plan for Mercian independence. Pope Leo III re-established (802) the authority of the see of Canterbury over all the churches which had ever been its subjects, and the archbishopric of Lichfield ceased to exist. When war at last broke out between Wessex and Mercia (825) the defeat of Beornwulf, king of Mercia, by Egbert of Wessex at the battle of Ellendun (now Wroughton, under the north Wiltshire downs) disintegrated the whole group of subject kingdoms on which Mercian ascendancy itself had rested.

The Rise of Wessex Under Egbert (802–839).—After the battle of Ellendun, Wessex, which since Ine's reign had been of small account, became the strongest of the English kingdoms. In virtue of long-dormant hereditary claims, Egbert was accepted as king in Kent, Sussex, Surrey and Essex. The king and people of East Anglia, who had come over to him for protection against the Mercians, fell back into independent obscurity when the danger was removed. But the extension of Egbert's kingdom toward the southeast gave him resources adequate for enterprises toward the north, and in 829 he conquered Mercia itself, winning thereby a reputation which caused a later West Saxon annalist to add him as the eighth *bretwalda* to Bede's list of early overlords. An expedition to the northern border of Mercia brought him to the height of his fame by imposing terms of peace on the Northumbrians without the necessity of a battle. For a short time Egbert was the direct ruler of Mercia, styling himself *rex Merciorum* on coins issued in his name at London. In 830 Wiglaf, the Mercian king who had been dispossessed in 829, recovered his kingdom, apparently against Egbert's will. There is no evidence that he ever became Egbert's man, or that, after his restoration, Egbert possessed any authority in the south outside his own kingdom of Wessex and the southeastern provinces which he had annexed in 825. But this enlarged West Saxon kingdom was to be the basis from which his descendants advanced to sovereignty over all England, and its creation makes his victory of 825 a turning point in English history.

The Period of the Scandinavian Invasions

Before Egbert's accession there were signs that England was threatened by the movements among the Scandinavian peoples (*see* VIKING). In Northumbria the monastery of Lindisfarne was sacked in 793 and that of Jarrow in 794 by raiders vaguely described as "heathen men." Between 786 and 802 three ships' companies from Hordaland in Norway landed in Dorset and killed the reeve of Dorchester, who had mistaken them for traders. But the general harrying which these descents foreshadowed was long delayed. Norwegian raiders naturally turned to the northern seas, and a strong kingdom of the Danes, closely watched by the officials of the Frankish emperor, discouraged individual adventurers. It was not until 835, when the empire was losing its grip upon its northern frontier, that heavy Danish raids began to fall on England, and it was the anarchy in Denmark after the fall of King Haarik (Horik) I in 854 which made possible the "Great Army" of 865.

The English defenses were ill-prepared to meet a succession of large-scale raids, culminating in an invasion. There was no fleet in being until King Alfred deliberately undertook the building of warships. The fyrd, or peasant militia, could not easily be brought into action outside its own shire. The striking force which every king possessed in his household of military companions was ill-adapted to warfare against shifting hordes of raiders, whose numbers were sometimes large and whose initial movements were incalculable. The walled cities which had remained from the Roman occupation were of small value as defensible centres when the raids began. Winchester, Canterbury, even London, were sacked before the great invasion of 865. The ascendancy which ultimately came to the English forces was secured by the building of fortresses at well-selected points. Work of this kind was a communal obligation of long standing. Offa had insisted with heavy emphasis that military service, bridge building and the construction of fortresses were public duties necessary for all the people, from which no one could be excused. But the demand for labour required by a scheme of national defense which might be tested at any moment by enemy action met with stolid resistance from an exhausted peasantry, and its completion belongs to the 10th century.

Between 835, when large-scale raiding began with a descent on the Isle of Sheppey, off the Kentish coast, and 865, when an army strong enough to destroy kingdoms invaded East Anglia, few parts of the country can have been secure. Descents upon Lindsey, East Anglia, Kent and Wessex; the fall of a Northumbrian king in battle; and a raid across England which brought the heathen to the Wrekin are known upon good authority. In Wessex, but, as far as the evidence goes, there only, the English resistance was effective. The last of King Egbert's achievements was to defeat a force of Danes and Cornish Britons which was about to invade Devon. Both Egbert and Aethelwulf (*q.v.*; 839–858), his son, were well served by their local officers but the outstanding English victory of the period fell to King Aethelwulf himself, who in 851 defeated a Danish army estimated at 350 ships' companies after it had driven Beorhtwulf (d. 852), king of Mercia, into flight. The victory must have given to Aethelwulf at least an informal primacy among English kings. It was as a superior, if not, indeed, as a lord, that in 853 he joined Burgred, the new king of Mercia, on an expedition into Wales, and soon afterward gave him a daughter in marriage. His position was so strong

that in 855 he could undertake a long-planned pilgrimage to Rome, leaving Aethelbald (d. 860), his eldest surviving son, as regent. On his way home a year later he married Judith, the young daughter of Charles the Bald, king of the West Franks, the most obvious ally for an English king against the enemy impending from the north. But his return to England had been too long delayed; he was compelled to leave his ancestral West Saxon kingdom in Aethelbald's hands and he died (858) as king only of Kent and the other eastern provinces which his father had acquired in 825.

The "Great Army" of 865.—The Danish raids upon England in this period were light in comparison with the devastation continuously suffered by the Frankish kingdoms. Now and then the fortune of war diverted a marauding host from France to England. In 860 one of these bands crossed the channel and sacked Winchester, being afterward defeated by the local levies of Berkshire and Hampshire. After the death of King Aethelbald in this year, the West Saxon kingdom was reunited under Aethelberht (*q.v.*; 860–865), his brother, whose reign was remembered for the occupation of Thanet by a Danish host and the devastation of eastern Kent. But in the autumn of 865 the social order of the whole country was challenged by the descent of a "Great Army," and the beginning of operations which lasted for 15 years, shattered the ancient kingdoms of Mercia, East Anglia and Northumbria, and turned a third of eastern England into a group of Danish colonies. The plan that the invaders followed year by year was to fortify each autumn a base from which they could harry the surrounding country until the inhabitants bought peace from them. The extreme rarity of English coins of this period in Scandinavian hoards suggests that most of the army intended from the first to remain in England. But their movements show no trace of any political design, and they had been campaigning for 11 years before their recorded settlements began. Nevertheless, they were not a formless horde of barbarians but a coherent army, organized with careful regard for diversities of rank and led by men—sons of the famous Viking Ragnar Lodbrok—who were skilled in all the arts of northern warfare. In action they showed a discipline that could regain a battleground after an apparent defeat. Throughout the war they pursued their devastating course with a single-minded determination which would never take an unnecessary risk, but could never be worn down.

The Attack Upon Wessex.—The movements of the "Great Army" are only recorded in a chronicle written in Wessex (the Anglo-Saxon Chronicle), which understresses the resistance offered by other kingdoms but as far as it goes is an accurate record. It was not until the autumn of 870 that the Danes attacked Wessex and the chronicle passes lightly over their original invasion of East Anglia in 865, the capture of York in 866, the occupation of Nottingham in 867, the return to York in 868 and the occupation of Thetford in 869. Neither the fighting in and around York in 866, in which two rival Northumbrian kings perished, nor the killing of St. Edmund (*q.v.*), king of East Anglia in 869, which contemporaries came to regard as a martyrdom, receives more than the briefest of possible notices. With the descent of the army on Reading in 870 the chronicle enters into more detail, recording that nine general engagements were fought with the heathen during the next 12 months, bringing out the manner in which the movement of the Danes was limited by English resistance, and enlarging upon the battle at Ashdown on the Berkshire downs, in which one Danish king and five Danish earls were killed, and the English won their first recorded victory in the war. It also shows that the West Saxons were brought dangerously near to a collapse. Their king, Aethelred (*q.v.*), who had succeeded his brother Aethelberht in 865, and whose reign of five years had been overshadowed by the war, died early in 871. Alfred (*q.v.*; 871–899), his brother, who succeeded him, was to become the greatest of Anglo-Saxon rulers but in 871 he was a young man whose only experience of the war had been under his brother's leadership. In the summer of 871 the Danes were reinforced by the arrival of a large army of their own countrymen which must have made good all their losses in the previous fighting. In the autumn, like the Mercians, East Angles, and Northumbrians before them, the West Saxons bought peace from the invaders.

From Reading the army moved to London, where it remained for a year. Late in 872, after securing its control of Northumbria, it took winter quarters at Torksey in Lindsey. A year later it struck at the heart of Mercia, entrenched itself at Repton and expelled King Burgred. From the earliest phase of the war, many Englishmen had been prepared to make terms with the Danish leaders. An Englishman, Egbert, had accepted the Northumbrian kingdom from them after the disaster of 866. In 874 they made a formal treaty with a retainer of King Burgred named Ceolwulf, who bound himself by oaths and hostages to hold the Mercian kingdom on their behalf, to surrender it on their demand and keep himself in readiness to join them at any time with all the men willing to follow him. These remarkable terms imply that the army was planning the permanent occupation of the territory now at its disposal. On leaving Repton late in 874 it fell apart into two independent forces. One of them, after a year's desultory raiding in the far north, divided the region which was to become Yorkshire among its members, who immediately turned to agriculture and fell out of the war. In the summer of 877 the other force, which had spent a year at Cambridge without recorded incident and nearly two years raiding Wessex to little profit, carried out a similar division of Lindsey and eastern Mercia, leaving the west of that kingdom to its protégé Ceolwulf. In the first days of 878 the residue of this force, which had remained in winter quarters at Gloucester, suddenly renewed the attack upon Wessex and brought the last surviving English kingdom to the verge of destruction.

King Alfred's Successes Against the Danes.—The events which followed form the most famous episode in Anglo-Saxon history before the catastrophe of 1066. The Danes under King Guthrum (*q.v.*) fell unexpectedly on Chippenham in the second week of Jan. 878, fortified the village and overran the whole of Wessex. Many Englishmen left the country, others came to terms with the enemy and King Alfred with his household men withdrew into the marshes of Somerset. For at least the next four months western Mercia was the only part of England south of the river Tees within which a settled English government was carried on. It was not until the middle of May that Alfred won the battle of Edington which reversed the course of the war. Six weeks passed in the arrangement of terms with the defeated but still formidable enemy. By the treaty then made, Guthrum gave security to Alfred that the army would leave his kingdom. But the movement was not carried out until the autumn and even then it left the Danes in force at Cirencester within five miles of the West Saxon border. Fortunately for the peace of Wessex, Guthrum, who had been baptized while Alfred's guest during the recent negotiations, kept the terms to which he had then sworn. In the autumn of 879 he and his men left Cirencester for East Anglia, which had been twice overrun, but not yet settled by a Danish army. The last of the three great Danish plantations of the war was the division of East Anglia among Guthrum's followers.

For the next five years, the Danish forces still at large were chiefly engaged upon the continent, where they were closely watched from England. In 884 one of them crossed to the Medway and besieged Rochester. The short campaign which followed had important consequences. For the first time in these wars the Roman fortifications of an English town withstood a siege. Rochester was relieved by Alfred; some of the Danes recrossed the sea, and the remainder took to casual raiding which brought them to Essex, where they dispersed. On the other hand, Guthrum's East Anglian Danes had regarded them as allies, and it had become clear that if Wessex were to be secure, London must be firmly held by English hands. The sequel to this campaign was an expedition in 886, in which Alfred captured and garrisoned the city, appointing as its governor a nobleman named Aethelred who had been ruling English Mercia for at least the previous three years. Alfred's occupation of London made him at once the leader of all Englishmen who wished to challenge the Danish control of northern and eastern England. He was immediately chosen as lord by all Englishmen who were not directly subject to the Danes, thereby rising to a position which, unlike all earlier supremacies, rested on the voluntary adherence of the entire nation. Among

the Danes it gave him a reputation which would never have been accorded to a local king of Wessex. The coins bearing the monogram of London by which he commemorated its capture were imitated in the remotest parts of Danish England. The year 886 is the most probable date for the treaty by which he and Guthrum defined the boundary between their kingdoms and the terms on which Englishmen and Danes should live beside each other. The boundary left Essex to Guthrum, but Alfred's fame secured a legal standing to Englishmen in Guthrum's kingdom such as was rarely enjoyed by the men of a defeated race.

Alfred's supremacy was made permanent by the attitude of Aethelred of Mercia and the aristocracy of which he was the head. A charter of 887 shows that he was already married to Alfred's daughter Aethelflaed, and that though lord of the Mercians in fact and title he regarded Alfred as his suzerain. The Mercian alliance thus founded was decisive in the later phases of the war and its value was soon apparent. Between 892 and 896 England was continuously harassed by a large Danish army which had been heavily defeated on the continent in 891. It was supported by the Danish colonies in Northumbria and East Anglia—where King Guthrum had died in 890—but it was kept in constant movement and at last worn down by the combined forces of Wessex and English Mercia. In 899, the last year of Alfred's life, raiders from Danish Northumbria were abroad in force on English ground. But his relations with Aethelred and the Mercian nobility had created an Anglo-Saxon state which was now beyond the risk of destruction in war.

The Revival of Learning Under Alfred.—The character of this state had been determined between 865 and 899. No country could have preserved its culture and social order unchanged through the devastation of these years, and the English capacity for resistance was dangerously weak. The intellectual and artistic energy which had once distinguished England had slackened long before the Danish wars began. The deplorable Latin of mid-9th-century charters is a symptom of a deep-seated change. Its extent is brought out in a famous sentence by King Alfred describing, in the preface to his translation of the *Cura pastoralis* ("Pastoral Care") of Gregory the Great, the churches which he remembered before the land was all harried and burned, full of books which their owners could not read. The destruction of these relics of vanished scholarship in the churches which housed them brought a tale of long decay to a dramatic end. The drift toward illiteracy which it threatened was halted by Alfred's recognition of the fact that, although the knowledge of Latin had almost disappeared, some people could read English. The translations of ancient writings through which he moved toward a revival of learning were not intended to replace the Latin culture on which religion, scholarship and the arts depended. They were simply the books which Alfred considered that every man should know. Historically they are of outstanding interest because the cultivation of English prose which stemmed from them stimulated the growth of a literature in the native language which survived the Norman Conquest itself. The Anglo-Saxon Chronicle (*q.v.*), which was begun in Alfred's reign, was continued in one version until 1154.

Social Consequences of the Wars.—The social consequences of the wars are harder to assess. Few documents remain to illustrate contemporary statements of general havoc. Nevertheless, the recorded passage of the "Great Army" year by year across England implies devastation on a scale which no society could have survived unshaken. The lesser but more erratic raids of the period repeatedly interrupted the slow process of recovery, and created everywhere a sense of insecurity peculiarly daunting to a peasantry which at the best of times was living on a narrow margin of subsistence. It is probable that these conditions go far to explain the difference between the rural society of Ine's laws (late 7th century), where communities of free peasants are expected to transact their own agrarian business without the intervention of any lord, and that of the 10th-century codes, where lordship of a sort that can fairly be called manorial is a fundamental assumption. The temptation must have been great for a group of peasants, harried beyond endurance by an enemy with whom no terms could be made, to place itself under the protection of a lord who could at least provide seed corn and stock in return for a promise of service guaranteed by the surrender of inherited holdings. The manorial system of the middle ages was the product of many factors; among them, the practice, discernible as early as the 7th century, by which an individual would bargain with a nobleman for a grant of land in return for rent or service or both. The significance of the Danish wars in this context is that they supplied a motive sufficient to explain the surrender of economic independence by entire communities in all parts of Anglo-Saxon England.

Achievement of Political Unity

The Work of Edward the Elder (899–924).—The conquest of Danish England in the generation after Alfred's death was made possible by the virtual union of Wessex and Mercia which he had brought into being. After ten years of an uneasy balance of power, the Danish armies of the east and midlands were outmaneuvered and defeated in a series of campaigns planned in close alliance by Alfred's son Edward the Elder, king of Wessex, and his sister Aethelflaed (*qq.v.*), the "Lady of the Mercians." The method adopted by King Edward in the region where he operated was to neutralize each of the Danish armies in turn by fortresses which limited its movements or could be garrisoned in sufficient strength to compel its submission. Aethelflaed's fortress building in western Mercia, planned at first for defense rather than attack, brought her in course of time within striking distance of the still unbroken Danish armies nearest to her eastern border. In 917 her capture of Derby was probably decisive in defeating a general attempt by the Midland and eastern Danes to halt the West Saxon advance. In 918 Leicester submitted to her without fighting, and the Danes of York, threatened by an invasion of Norwegians from Ireland, offered her their allegiance. But she died in the summer of 918, and her brother, then about to attack the Danes north of the river Welland, was compelled to suspend operations until he had secured recognition of his direct authority over Mercia. Before the end of the year he was able to advance again; Nottingham and probably Lincoln surrendered on demand, and the way was opened for a demonstration in the further north. In 919, at or beyond Bakewell in Derbyshire, he was acknowledged as "father and lord" by the kings of Scotland and Strathclyde; by Raegnald, a Norse Viking who had recently become king of York; by Ealdred, the English lord of Bamburgh; and by all the Northumbrian people—Englishmen, Danes, Norwegians and others.

To those who then came in to King Edward, this famous act meant little more than an undertaking to respect his territory, help him against his enemies and keep peace with one another. To King Edward himself, who had already received the submission of the Welsh princes, it meant that every other ruler in Britain had now accepted him as overlord. But it gave him no direct authority beyond the Humber. Even in Mercia there were many who resented his annexation of their ancient kingdom. The last of his wars was an expedition in which he suppressed a rising at Chester, dangerous because it was supported by Welsh allies. For the permanence of his work it was fortunate that Aethelstan (*q.v.*), the son who succeeded him, had been educated at the Mercian court of Aethelred and Aethelflaed, and knew at first hand the younger noblemen who became the leaders of the Mercian people.

The Prestige of Aethelstan (924–939).—By the end of Aethelstan's reign, Wessex and Mercia had been absorbed into a national state, with a foreign policy, and a centralized administration which set itself to emphasize the range of its master's power. English relations with the continent had never been completely broken during the Danish wars. Alfred had married a daughter to Baldwin II, count of Flanders, and had sent alms year by year to Rome. Edward the Elder had married a daughter to Charles the Simple, king of the West Franks, and their son Louis (afterward Louis IV d'Outremer) was brought up in England. But under Aethelstan these relationships were merged in a system of foreign alliances, each confirmed by a marriage, which united England to the Holy Roman empire, the kingdom of Burgundy and the new power rising to importance in France under Hugh the Great, duke of the Franks. Outside the family of western nations, Aethelstan's reputation brought him an embassy from the half-mythical Harald Haarfager, king of Norway, which presented him at York with a

BY PERMISSION OF THE MASTER AND FELLOWS OF CORPUS CHRISTI COLLEGE, CAMBRIDGE

AETHELSTAN PRESENTING BEDE'S LIFE OF ST. CUTHBERT, FROM A 10TH-CENTURY ENGLISH MANUSCRIPT

brilliantly appointed warship, a gift which makes more credible the Norse tradition that Haakon, Harald's youngest son, was fostered at Aethelstan's court. These displays of power and influence were supported by Aethelstan's determination that the princes of Wales and the Danish earls of central and eastern England should attend him on his progresses over his ancestral West Saxon kingdom. Their presence is recorded in a series of royal charters which are remarkably elaborate in style and structure. They prove that Aethelstan had a staff of clerks permanently at command, and supply a 10th-century date for the origins of the English chancery.

The Pacification of the North.—The danger to this well-ordered kingdom lay beyond the Humber. The Danes whose ancestors had colonized the north and east in Alfred's time had no wish for political independence. But a new situation had been created in the north by incursions of Norsemen from Ireland, who had settled in large numbers between the Solway and the Mersey, and in 919 had established a kingdom in York itself. In little more than three years from his accession, Aethelstan had expelled the Norwegians from York and had imposed a recognition of his overlordship upon the kings of Scotland and Strathclyde. For the rest of his life he ruled York and its neighbourhood with the good will of both Danes and Englishmen. But the northern border of England was ill-defined; Britons from Strathclyde were crossing the Solway to settle in what is now Cumberland, and their king was closely related to the king of Scots. A general attack from the Celtic north was postponed by an expedition which Aethelstan led far into Scotland in 934. The conclusive test of his power came three years later, when a host of Vikings from Dublin under Olaf Guthfrithson, head of the Norse family which had once held York, was joined by Constantine, king of Scots, and Owain, king of Strathclyde. Their defeat, on a site, still unidentified, called Brunanburh, united West Saxons and Mercians, once for all, in the memory of a great victory. It did not end the Norwegian danger. Immediately after Aethelstan's death in 939, Olaf Guthfrithson invaded England again, captured York and made himself master of all Danish Mercia as far south as the Welland. The Mercian lands were recovered in 942 by Edmund I (939–946), Aethelstan's half brother and successor, but for the next 12 years the north was kept in confusion by a succession of Norse adventurers trying to found kingdoms. It was not until 954, when Eric "Bloodaxe," a son of Harald Haarfager, was overthrown by his local enemies, that York passed permanently under English rule and a kingdom of all England became a possibility.

Its emergence was at first disguised by the policy of the West Saxon kings themselves. Aethelstan and his successors—Edmund I, Edred, Edwy and Edgar (*qq.v.*)—were ruling with traditional resources a kingdom far wider than had belonged to any of their ancestors. Naturally enough, under these conditions, they tried to reduce rather than enlarge the sphere within which they were responsible for administration. In 945 Edmund conquered Strathclyde, but immediately granted it to Malcolm I, king of Scots, "on condition that he should be his fellow worker by sea and land." A generation later, Edgar (959–975) granted Lothian, which had been English since the 7th century, to Kenneth II, king of Scots, in return for an acknowledgment of his overlordship and an undertaking to appear from time to time at his court. Within his northern and eastern lands of recent conquest, Edgar had already granted autonomy to his subjects of all races, whom he collectively describes as Danes, as a reward for the loyalty which they had always shown him. Through these grants Edgar's reign left a permanent impression on British history. In course of time the cession of Lothian transferred the balance of power from the Celtic to the English element within the Scottish kingdom. The grant of autonomy to the Danes in England preserved in full operation a body of Scandinavian law and custom which the Norman kings recognized, so that in the 12th century the whole area between the rivers Thames and Tees could be succinctly described as the Danelaw (*q.v.*). The spectacular side of kingship is stressed by the authorities for Edgar's reign, which enlarge upon the solemnities of his long-deferred coronation in 973, and the ceremony at Chester, when, according to one version, eight underkings came together and did him homage. His dealings with his Danish subjects and with the king of Scots prove the statesmanship of his government.

The 10th-Century Monastic and Literary Revival.—To modern historians, Edgar's work as a secular ruler is generally overshadowed by his patronage of the monastic revival inspired by his friend Dunstan (*q.v.*), archbishop of Canterbury. Their attitude is not unreasonable, for this renewal of monastic life coloured the whole subsequent history of the Anglo-Saxon church, and supplied the foundation for an achievement in art and letters of the highest intrinsic merit. It was made possible by the diversion of Scandinavian enterprise from England to the continent, where innumerable Danish adventurers were absorbed in the colonization of Normandy, but it was so well grounded that it survived, and indeed increased in volume, during the disasters of the next generation. In time it issued in the development of a school-taught form of English prose which carried the native language as a normal means of literate expression into the early middle ages. In the related arts of calligraphy and book production this vernacular culture produced work of far more than academic interest. Four great volumes of this period, the *Exeter Book* and the *Vercelli Book* (*qq.v.*), the *Beowulf* manuscript and the so-called *Caedmon* manuscript (*see* BEOWULF and CAEDMON), preserve most of what is known from the English poetry of an earlier age. At an exciting moment, a contemporary poet could, himself, rise to the point of creation. There is no finer expression of heroic virtue in Germanic literature than the verses on the battle of Maldon (991) which describe the companions of Earl Byrhtnoth of Essex, outnumbered by a Viking host, devoting themselves to death in a last stand around the body of their lord (*see* ENGLISH LITERATURE: *From the Earliest Times to Chaucer*).

ANGLO-DANISH STATE

The contrast between this intellectual activity and the collapse of the central government when tested by war is the outstanding

feature of the period after Edgar's death. Edward the Martyr (*q.v.*), Edgar's eldest surviving son, was murdered (978) after a short reign by the household men of his brother Aethelred. The epithet *unraed* ("no counsel") applied to Aethelred II (*q.v.*; 978–1016) by contemporaries is a severe but not unjust description of a king who could neither organize the defense of the realm nor control his greater subjects. The distresses of his reign followed naturally from the incompetence of a king confronted by danger from the Scandinavian north greater than at any period since the age of Alfred. The northern peoples were then passing through the phase of unrest which preceded the establishment of united kingdoms in Norway and Denmark. Large companies of adventurers, sometimes led by men of high rank in their own countries, repeatedly harried the English coast in Aethelred's earlier years. Their object was simply plunder, and in 991 the government began the disastrous policy of buying temporary peace by the payment of a danegeld. There was no central authority in the north with which a settlement could be negotiated, and an incalculable element had recently been introduced into northern politics by large aggregations of warriors operating from their own fortified bases in the Baltic. One of them, planted in a fortress called Jomsborg, near the mouth of the Oder river, owed allegiance to Sweyn (Swein) I, king of Denmark. In 1012, by taking Thorkel, Jomsborg's leader, into his service, with 45 ships, Aethelred became Sweyn's personal enemy. In 1013 Sweyn invaded England, secured the submission of the Danelaw without a battle, and then systematically reduced the country under Aethelred's obedience, compelling him to take refuge with Thorkel and his ships in the Thames. Aethelred's wife, Emma, who was sister of Richard, duke of Normandy, had already crossed the channel, and before long he followed her, leaving Sweyn the unquestioned king of all England.

The death of Sweyn in Feb. 1014 was quickly followed by the first recorded compact between an English king and his subjects. In a written document, afterward confirmed by the exchange of formal promises and sureties, Aethelred agreed to reform everything that his people hated and to forgive everything that had been done or said against him, if they returned to his obedience, unanimously and without treachery. Restored on these terms, he commanded a force with which the Danish host in England could not compete. Canute (Cnut), Sweyn's younger son, who had now become its leader, brought it back to Denmark without risking a battle, and spent the next 15 months in preparations for a fresh invasion. In England these months passed in an atmosphere of growing tension which paralyzed the national defenses when Canute (*q.v.*) landed again in the summer of 1015. The spirit of the ordinary fighting man was good, but there was treason and suspicion of treason in high places, and the war would have ended in an early collapse of the English resistance but for the unbreakable determination of Edmund II (*q.v.*), nicknamed "Ironside," Aethelred's eldest surviving son. Even so, when Aethelred died in April 1016, many if not most Englishmen were prepared to accept Canute as his successor. During the next six months Edmund called out the national militia five times in a war which ended in an agreed division of the kingdom, leaving him Wessex but giving the rest of England to Canute. On Edmund's death in Nov. 1016, Canute succeeded him peacefully in Wessex, and the Danish conquest of England was complete.

Canute (1016–35).—The character of Canute's reign was determined by his attitude to his new dignity and the obligations which it carried. His marriage (1017) to Emma of Normandy, Aethelred's widow, secured him against Norman intervention on behalf of Aethelred's sons by her, then living at the Norman court. He became king of Denmark when his elder brother died in 1018 or 1019. For a short time he was acknowledged as king through Norway, and to the end of his life he was closely occupied with northern interests. But his English kingdom carried him into the wider sphere of western politics, which no Viking chief before him had ever entered. Like the previous kings of his line, he was a Christian, and he set himself to rule in England as a Christian king, accepting the close interrelationship of church and state characteristic of the English polity, and making it the basis of the most elaborate code of law which had so far appeared in England. By the mass of Englishmen he was welcomed as a king strong enough to enforce internal order and protect the land against Viking adventurers. The price of a quiet life was the heavy taxation needed for the payment of the armed forces which Canute maintained on land and sea, and the replacement of Englishmen by Danes in many important spheres of local government. Even so, the English element among his officers increased later in the reign and the continuity of the English administration was never broken. The political system of which Canute was the head was too artificial to survive him for long. Norway had broken away before he died (1035). In England, secession was delayed for seven years. But in 1042, on the death of Hardicanute (*q.v.*; Harthacnut), his chosen heir, native feeling for the ancient dynasty brought back from exile Edward, son of Aethelred II and Emma of Normandy, to rule the Anglo-Danish state.

Edward the Confessor, Harold and the Norman Conquest.—The reign of Edward the Confessor (*q.v.*; 1042–66) is generally regarded as a prelude to the Norman conquest. Kinship through his mother with the Norman ducal family and gratitude for protection while England was under Danish rule seem to have aroused in this childless king the wish that the duke of Normandy might succeed him as king of England. Among other possible successors, Edward, son of Edmund Ironside, the nearest by birth, had been an exile in Hungary since 1016; Harald III Hardraade, king of Norway, who claimed to inherit England under a treaty made with Hardicanute, was a successful Viking leader with no pretensions to statecraft; and Sweyn II, king of Denmark, son of Canute's sister, Estrith, had no claim beyond this attenuated relationship to the Danish royal family. Under these conditions, the king's word was likely to be of decisive weight, and there is evidence that in 1051 Edward gave it in favour of William, duke of Normandy, who is recorded to have visited England in that year. The proposal was defeated by the powerful family founded by Godwin (*q.v.*; d. 1053), earl of Wessex, an Englishman raised to high office by Canute. Godwin's son Harold (*see* HAROLD II), outdistancing all other earls in wealth and influence, came in the next ten years to a position in which the crown itself was within his reach. Harold obtained a grant of the kingdom from King Edward on his death bed, and on Jan. 5, 1066, after election by such noblemen and higher clergy as were then in or near London, he was crowned in Edward's newly founded abbey church of Westminster.

The collapse of the Old English monarchy, which followed within the year, is often put down to weaknesses inherent in the Anglo-Saxon state, notably the loosely articulated structure of a society in which countless free men were at liberty to choose their own lords, and the accumulation of political power in the hands of a few great families. In reality it was due to a temporary combination of circumstances which only a leader of assured authority aided by exceptional good fortune could have overcome. In 1066 William of Normandy had recently become free to carry out a large-scale enterprise across the channel. The count of Flanders, his wife's father, was guardian of the infant king of France, and the county of Anjou, the old enemy of Normandy, was paralyzed by a civil war. Tostig (*q.v.*), Harold's brother, once earl of Northumbria, who had been driven from the country by a local revolt in 1065, was planning an invasion of England in alliance with Harald Hardraade. In England itself a strong party resented the dispossession of the ancient royal line, represented by Edgar the Aethling (*q.v.*), King Edward's nephew, and many Northumbrians were unwilling to accept a king from Wessex. William of Normandy raised an army and built a fleet of transports without hindrance, while the long delay before action began wore out the patience and exhausted the supplies of the forces mobilized for the defense of the south coast. The end of these uncertainties came suddenly. Early in September, Harald of Norway and Tostig, his renegade ally, entered the Humber and destroyed an English army drawn up in front of York. Their own destruction by Harold of England while awaiting the delivery of hostages at Stamfordbridge, seven miles east of the city, showed that Harold commanded the forces and could move them with the speed necessary to retrieve a desperate situation. The final

crisis of this catastrophic year was introduced by a change of wind which brought William and his army into Pevensey bay on Sept. 28, three days after the battle of Stamfordbridge. It was decided on Oct. 14 when Harold, with his army unarrayed, was surprised by William on the hill nine miles from Hastings now marked by the town of Battle. After a long day's fighting, Harold was killed among the last defenders of a position which William's horsemen in reiterated assaults had failed to carry. (*See* Hastings.) The campaign was continued by a sweeping movement of William's army which isolated London, where resistance was still possible, from all sources of reinforcement. The city surrendered without a siege and, on Dec. 25, the coronation of William in Westminster abbey ended the Anglo-Saxon phase of English history.

II. THE NORMANS (1066–1154)

Anglo-Norman State

England before the Norman conquest was a highly organized state, in which the king's writ ran everywhere. It had inherited a large body of written law, administered in local courts where the king's financial interests were always watched and proceedings were commonly directed by an executive agent of the crown (*see* Sheriff). Under the pressure of the danegeld (*q.v.*) the country had developed an elaborate system of national taxation. It had produced a currency, based on an orderly sequence of types, which was issued from mints in every borough, but was regulated in weight and design by the central government. At every point where comparison is possible there was a sharp contrast between England and Normandy, where the organization of society in its higher ranges was determined by military necessities, and it had taken 20 years for a duke of wholly exceptional power to bring a fiercely competitive aristocracy into effective control. The Normans were not a race of high creative ingenuity, but their surplus energy could give fresh life to ancient institutions and under firm direction could carry through hazardous adventures. Everything in their history turned on the personal quality of their ruler. Under the first three Norman kings, England became the most formidable as well as the best-administered state in western Europe. Under the fourth, it collapsed into anarchy.

The Work of William I (1066–87).—William I claimed to reign in England as King Edward's heir. He had no desire to remodel the framework of the English government and as far as possible he wished to use Englishmen in its administration. Here his policy was defeated by an irreconcilable element among Englishmen of rank and local influence, which was particularly strong in the north, where it was only suppressed by a savage devastation of the countryside (1069–70). By William's later years the government of England at all levels had come to be controlled by Normans who with some notable exceptions showed remarkable skill in the adaptation of English institutions to the new regime. The subjection of large areas to a forest law imposed in the interest of the king's sport, and the exaction of a heavy fine (the *murdrum*) from districts within which a Frenchman had been murdered by an unknown or unarrested slayer were the innovations which touched the common man most closely. But the law administered in the local courts was in all essentials English, and even in the routines which governed life within the manor there is little discernible variation from Old English practice. The manorial courts of the early middle ages were the direct descendants of the assemblies which had met for similar business in the halls of Anglo-Saxon lords.

Introduction of Norman Feudalism.—This continuity was remarkable, for in the higher ranks of society the Norman conquest was the prelude to a revolution. England under the Norman kings was a state organized for war, in which the distribution of land among the Conqueror's followers had been intended before all else to provide the king with a sufficiency of men practised in the art of fighting on horseback. In rewarding his followers William required each of them to be responsible for a definite number of these knights, whom he would bring, properly mounted and equipped, to the king's host, at his summons (*see* Knight Service; Feudalism). Each of these men was normally provided with a "knight's fee," a holding in land sufficient for his maintenance, so that a new class of landowners, with technical qualifications for military service, quickly replaced the Old English country gentlemen or thegns, who had been good fighting men but had never learned the cavalry tactics in use abroad. By imposing this military burden on the ancient estates of the church, William expanded a force already large into a formidable army. The French innovation of the castle was used by William to forge into coherent groups the knights of which this army was composed. In military administration, as throughout his government, William's success turned on the effectiveness with which he enforced the principle that the interests of crown and baronage were interdependent, and that the security of his followers in their new possessions must be bought by the acceptance of conditions onerous in themselves and often foreign to the traditions of their order. At a council held at Salisbury in 1086 he was able to exact an oath of fealty and an act of homage not only from his tenants in chief but from all the landowners of any account in England, "whosesoever men they were."

Fusion of the Two Stocks.—By the end of William's reign, Domesday Book (*q.v.*), the description of England which is the greatest memorial of his rule, shows that the outlines of the medieval English state were beginning to appear. England in 1086 still bore the aspect of a conquered country, administered by men whose political conceptions were French, under a king whose personal interests were centred in France. But a place in the lower ranges of government had always been open to acquiescent Englishmen, and the value of English support against disaffected barons was apparent to every king of the Norman line. The distinction between their French and English subjects was a matter of common form in their chanceries, but intermarriage was rapidly blending the stocks, and a new nation had come into conscious being long before the 12th century was over.

The union of England and Normandy, which all these kings were anxious to maintain, was an artificial creation, advantageous to English traders and a source of power to lords with land on both sides of the channel, but superficial in comparison with the internal forces which were making for the growth of national self-consciousness.

Administrative Developments Under William II (1087–1100) and Henry I (1100–35).—On the Conqueror's death (1087) his dominions were divided in accordance with feudal custom, Normandy, his inheritance, going to his eldest son, Robert II (Curthose) and England, his kingdom by conquest, to his second son, William. Robert's weak rule in Normandy and his claim, as many held, to his father's kingdom, gave rise to inconclusive wars and negotiations which were not ended until 1106, when Henry, his youngest brother, by then firmly established as king of England, defeated and captured him in the battle of Tinchebrai. The forfeitures suffered by Robert's English supporters increased the royal power at the expense of the baronage, and contributed to the virtual despotism achieved in England by Henry I toward the end of his reign. After 1128, when William the Clito, Robert's son, was killed in battle, Henry remained the sole male representative of the Conqueror's line. In Normandy as in England, the troubles which followed his death were due in part to reaction against the arbitrary government of a ruler who outmatched all his foreign enemies and had no rival within his own house.

Under the Conqueror, the men who witnessed royal acts most frequently were magnates such as his half brother Robert, count of Mortain, Richard, son of Count Gilbert of Brionne, or William de Warenne. Under William II and Henry I the king was more often attended by men whose lands were the reward of service in the government of the country or in the royal household. A new official class from which judges, royal agents and financial administrators were drawn had come into existence and, with the clerks of the chapel who formed the royal chancery (*see* Lord High Chancellor), had provided the country with an administrative system of marked efficiency. Its most permanent achievement was the establishment of the financial bureau known as the exchequer (*q.v.*). Created in order to receive and account for the money collected for the king by his sheriffs throughout England,

it had political as well as fiscal importance, for the custom which required each sheriff to come twice a year to Westminster to account for the money which he had collected kept the men responsible for public order within the shires in constant touch with the central government. From its foundation until 1139, the head of the exchequer was Roger, bishop of Salisbury, who in addition to his financial duties acted as the king's representative in England during his frequent absences abroad and at all times functioned in a manner entitling him to be regarded as the chief minister of state. He filled in fact the position afterward occupied by the chief justiciar (*q.v.*). The judicial organization of the country had not as yet assumed a permanent shape; Henry I and his successor, Stephen, appointed local justiciars in each shire to hold their pleas and make their judgments, and from time to time they sent commissions from the omnicompetent *curia regis* (king's court) to dispatch every sort of outstanding business throughout the land.

To deal with matters of the greatest moment each of the Norman kings would send out a special commission, which would use the Norman procedure of the sworn jury for the settlement of disputed issues. The information needed for the Domesday survey had been collected by a series of such commissions, which are characteristic of the English judicial system before the legal reforms of Henry II. Regarded as a whole, Norman England, before the outbreak of civil war in 1139, seems to have been a firmly governed, perhaps an overgoverned, kingdom and it is significant that some among the king's own justices were themselves uneasy about the harshness with which they were compelled to carry out their duties.

Church and State.—In secular affairs the drift of events after 1066 had tended to intensify the power of the crown. In the ecclesiastical sphere, the Conqueror had exercised a supremacy which his successors could not maintain. Edward the Confessor had acknowledged the ultimate authority of the pope over the English church, and sent a tax of a penny on every hearth, known as Peter's pence, to Rome at frequent intervals, but had kept the appointment of bishops and abbots in his own hands as far as the vicissitudes of national politics would allow. William I, claiming to exercise the ecclesiastical authority that had belonged to the Confessor, used it with a sense of responsibility and co-operated with the papacy in the reorganization of the English church after the Conquest. But although he entrusted its government to an archbishop of Canterbury expert in the canon law of Rome (*see* LANFRANC), he steadily opposed all attempts to extend papal influence in England and rigidly supervised the contacts of his bishops with the Holy See. His intransigence provoked from Pope Gregory VII an occasional reprimand; but the pope, deeply involved in his struggle with the emperor Henry IV, was on the whole content to let William have his way, justly regarding him as an ecclesiastical reformer.

William II, himself no statesman, was confronted with a papacy rapidly advancing to effective sovereignty over the whole Western Church and the association of a king who grossly misused his ecclesiastical prerogatives with an archbishop (*see* ANSELM, SAINT) for whom obedience to the pope was the first law ended in the archbishop's exile (1097). He was still abroad when King William died. This struggle settled nothing, but it had shown that the barriers which the Conqueror had set up against papal intervention in England were beginning to give way, and the kingdom to which Henry I succeeded had already become a field for the clash of ideas about the proper relationship of church and state which is commonly called the Investiture contest. The compromise (1107) by which it was ended in England was an act of statesmanship in which each party abandoned what some of his supporters regarded as matters of principle. The archbishop agreed that the homage done to the king by bishops and abbots designate should not be considered a bar to their subsequent consecration. The king surrendered the right to confer on the higher ecclesiastics of his kingdom the symbols of spiritual authority, the ring and pastoral staff. His control of their appointment was unaffected, but he had been defeated on a point which his father had taken for granted.

STEPHEN (1135–54)

The Disputed Succession.—In the next generation the whole question of the relationship between church and state in England suddenly entered upon a new phase, for the pope was requested to intervene in the succession to the crown itself. After 1120 Henry's only surviving legitimate child was his daughter Matilda (*q.v.*), who had married the emperor Henry V in 1114. In 1127, when she was a widow, he had caused the magnates of England and Normandy, lay and ecclesiastical, including his sister's son Stephen (*q.v.*), count of Boulogne, to swear that they would receive her as their lady if he should die without male issue. But in 1128 he had married her to Geoffrey Plantagenet, son of Count Fulk V of Anjou, and this marriage to the head of a family traditionally hostile to Normandy was considered by many barons to release them from their oath. Immediately after King Henry's death (1135) Stephen crossed to England, and within a month had gained enough support to justify his coronation. Before the empress had set her own claims in legal order, Stephen, to supply the defect in his title caused by the oath he had taken to Matilda in 1127, requested Pope Innocent II to confirm his election and consecration as king. The success of his application was publicized at Easter 1136 in a charter of liberties which describes him as king of the English, chosen with the assent of clergy and people, consecrated by William, archbishop of Canterbury, legate of the Holy Roman Church, and confirmed by Innocent, bishop of the Holy Roman See. Three years later when the case between Stephen and the empress was pleaded out before the Lateran council of 1139, Pope Innocent declined to revoke his previous confirmation, which continued to form the ultimate sanction of Stephen's authority as king.

Civil War.—Six months after the failure of her suit, Matilda invaded England, and a civil war began which did not end until 1153. Its course cannot be followed closely, for the central issue between king and empress was obscured by the ambitions and rivalries of great magnates, the activities of unattached adventurers and the uncertain loyalties of lords who could buy lands, offices, money rents and grants of enfeoffed knights by judicious changes of allegiance. Geographically the empress was strongest in the Severn valley. Robert, earl of Gloucester, an illegitimate son of Henry I, was her chief supporter, and Bristol castle was her headquarters. Stephen's resources lay chiefly in the east of England; in Kent, and above all in London, which gave him military assistance at more than one crisis of his fortunes. Two initial advantages of the first importance, the control of the national administration and the support of the church, he forfeited by the sudden arrest (1139) of Bishop Roger of Salisbury, and several members of the powerful clerical family which he had founded. Whatever may have been the reasons for this dramatic violence, its results were disastrous. Stephen's own brother, Henry, bishop of Winchester, the pope's representative in England, deserted him for the empress and won over many important barons to her party. Early in 1141 she came near to victory when Stephen was taken prisoner while attempting to recover Lincoln castle from Ranulf de Gernons, earl of Chester, who had entered the war on her side. But her chance was lost only a few months later when the men of London, where she was awaiting her projected coronation, revolted and expelled her from the city. Late in the year Stephen, who had been kept a prisoner in Bristol castle, was exchanged for Robert of Gloucester, who had been captured by the king's supporters, and the country fell into a state of endemic war.

After the summer of 1145, when Stephen, with an army from London, broke communications between Matilda's base in the Severn valley and her outposts along the middle reaches of the river Thames, it became clear that Stephen would never be dispossessed by violence. The earl of Gloucester died in 1147 and Matilda herself left England in 1148. Henry, her eldest son, born in 1133, was as yet unequal to an effective campaign against a king in possession. Stephen was once more provided with a chancery, and probably with at least the framework of an exchequer. But his power was the faint shadow of the authority that had belonged to the kings before him. He had neither the material resources nor the consistency of purpose needed to suppress the private wars

which had broken out in the previous years of real anarchy. Large tracts of country were terrorized from castles hastily thrown up by adventurers who lived by brigandage. The peace of the realm, which was the first responsibility of any king in normal times, was coming more and more to turn on the observance of treaties concluded by great lords with one another. The restoration of general order could not be expected from a king like Stephen, an aging man hampered by irreconcilable enmities. His elder son, Eustace IV, count of Boulogne, who would never have let his father's crown go by default, died in 1153. The preservation of his vast personal estates for his unadventurous second son, William, became Stephen's first concern. It was secured late in 1153 in a treaty with Henry, son of the empress, who had recently conducted a successful but far from conclusive campaign for his English inheritance. The treaty provided that Stephen should remain king for the rest of his life, and that Henry should be recognized as his heir and do him homage. On Oct. 25, 1154, Stephen died. On Dec. 19 Henry was crowned king at Westminster. (F. M. S.)

III. EARLY PLANTAGENETS (1154–1216)

Henry II (1154–89)

The accession of a young man, well educated and carefully trained for kingship, already duke of Normandy and Aquitaine and count of Anjou, brought England back to the centre of European politics. From the beginning of his reign Henry II ruled an empire stretching from the borders of Scotland to the Pyrenees. Although he failed to incorporate Wales into the English kingdom, his expeditions of 1157–58 and 1165 were impressive enough to keep the Welsh princes in subjection. His desire to add Ireland to his dominions was at least partially fulfilled after the Anglo-Norman invasion of the island (1169–70) and Henry's own expedition of 1171–72. From 1177 the title "Lord of Ireland" was borne by John, Henry's youngest son. William the Lion, king of Scotland, who had invaded England in 1174 in support of the rebellion of Henry's sons, was forced to acknowledge his subordination to the English king. Continual vigilance and close diplomatic relations with the rising kingdoms in Spain were necessary to keep the English hold on Aquitaine. Henry II spent roughly half his reign out of England, sometimes leaving it for several years together (1158–63, 1166–70, 1182–84). But the only political danger of his last years came from the determination of the French king Philip II (1180–1223) to extend his rule throughout France and from his success in exploiting the desire of Henry II's sons to anticipate their father's death. When Henry died (July 6, 1189) he had been driven almost to the point of surrender by the forces of Philip, with whom his sons Richard, duke of Aquitaine, and John, count of Mortain, had made alliance.

This failure at the moment of death in no way affected the position to which Henry II had raised the English monarchy nor did it lessen the respect felt by all chroniclers for the king himself. Save in his attempt to restore the relationship between church and state which William I had established and Stephen had abandoned, Henry had been successful in all his English policies.

The Quarrel With Thomas Becket.—In the early years of the reign as Henry's chancellor, Thomas Becket (*q.v.*) had been his inseparable companion and fellow worker in the essential task of restoring royal authority and reducing unlicensed castles. Appointed (1162) as the successor of Archbishop Theobald at Canterbury, he became the unyielding upholder of ecclesiastical privilege. Disappointment and anger at Becket's seeming defection made Henry show less than his usual skill in negotiation. Conscious that the majority of English churchmen were in sympathy with his own views, he forced Becket to promise acceptance of the relationship between church and state which had existed in Henry I's days. But the Constitutions of Clarendon (*q.v.*) presented at a council held there early in 1164 proved far more comprehensive than Becket had expected and he refused to seal the document. Before the end of the year he had fled from England and remained in exile until late in 1170. His murder in his own cathedral (Dec. 29, 1170) was disastrous for Henry, who in order to obtain reconciliation with the church was obliged to concede

BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

MURDER OF THOMAS BECKET, EARLIEST KNOWN REPRESENTATION, FROM AN ENGLISH PSALTER; *c.* 1200, IN THE BRITISH MUSEUM

that direct papal authority over the English church which had been exercised during Stephen's reign. He also had to abandon his attempt to force ecclesiastical persons to answer for their crimes in the lay courts.

Administrative and Judicial Developments.—The Becket tragedy did not divert the king from his long-term projects of reform. The exchequer had been re-established by Nigel, bishop of Ely, in the first years of the reign and the records of its transactions, the Pipe Rolls (*q.v.*), made up under the watchful care of Nigel's son, Richard (*see* FitzNeale), the treasurer, reflect year by year the success of Henry's internal policies. At the exchequer table, presided over by the justiciar (*q.v.*), all the king's ministers had their seats and there was room also for men specially summoned by the king. The sessions of the exchequer were meetings of specialists not merely in finance but in the art of government.

Nevertheless, the king himself was the initiator of the great developments in law and justice which mark his reign. The Assize of Clarendon of 1166 established the grand or presenting jury to inform the king's judges of the most serious crimes committed in each district, and of local suspicions and accusations made against their perpetrators. The Assize of Northampton (*q.v.*) of 1176 enlarged the list of crimes to be reported by the juries and increased the penalties which could be incurred. Regular visits of royal judges enforced these assizes and made the use of a jury (*q.v.*) familiar in criminal cases, although the ancient methods of trial by battle and ordeal remained in force (*see* Duel: *The Judicial Duel*).

Simultaneously the king was attracting to his court an increasing volume of civil litigation. He enforced the rule that no one need answer in his lord's court touching his free tenement without the king's writ, the writ of right. In the grand

assize he provided an alternative procedure to the duel for the settlement of these actions in his court by the verdict of a jury of local knights. To shorten litigation, parties were encouraged to make agreements, final concords, under the licence of the king's court. More popular still were the new forms of action which provided quick remedies for common wrongs. Eviction, whether by a lord or a neighbour, and the refusal of a lord to allow an heir to enter his inheritance were the commonest of all grievances and the most productive of violence and disorder. The writs of *novel disseisin* (1166) and *mort d'ancestor* (1176) offered victims the possibility of rapid settlement by verdict of a jury in the royal court, and put the burden of bringing an action for right on the defeated party if he thought it worthwhile. From these two writs a vast family of actions stemmed which dominated the law of real property well into the 19th century (*see* ASSIZE).

It would be inaccurate to say that Henry II aimed at limiting the activities of feudal courts of justice. Rather, he supplemented them by offering new procedures so attractive in themselves that barons and freemen alike sought the king's justice. In the long run the royal policy severely curtailed feudal jurisdiction, but this was a slow process which had hardly begun in Henry's reign. The king consulted his great council before issuing each of his assizes, whether it introduced new judicial processes or determined the arms that all must possess for the defense of the land (1181). Even the stringent inquiries into the operations of the sheriffs and the holders of feudal courts of justice, known as the Inquest of Sheriffs (1170), and the heavy penalties laid on delinquents did not provoke immediate reaction, though it is possible that the resentment that inquest caused was responsible for some of the support received by Henry's sons in their rebellion of 1173–74.

Intellectual and Artistic Achievement.—Henry II's last years saw the brilliant flowering of the 12th-century renaissance in architecture and art, letters, law and learning. Many cathedrals and parish churches were built about this time, a small university was growing up at Oxford and there were centres of learning at Exeter, at Winchester, at Northampton, at Lincoln and elsewhere. Despite the quantity and the distinction of contemporary poetry, letters, treatises and histories, it is possible to maintain that the most remarkable literary works of Henry's reign were the *Tractatus de legibus et consuetudinibus regni Angliae* ("Treatise on the Laws and Customs of England"), the first classical treatise on the common law, usually attributed to Henry's justiciar Ranulf de Glanville; the *Dialogus de scaccario,* the dialogue describing in vivid detail the actual working of the king's exchequer, written by his treasurer, Richard FitzNeale; and the chronicles from the pen of Roger of Howeden. Some of this work must have received royal encouragement if not actual direction.

RICHARD I (1189–99)

Philip II of France was obliged temporarily to abandon his attack on the continental possessions of the English king when his ally, Richard, duke of Aquitaine, inherited his father's kingdom. Both kings agreed that the crusade, to which they had pledged themselves in 1187, should be their first duty. Richard I spent a mere four months in England arranging for his absence and raising money for his crusade. For the first time an English king intended to be out of touch with the government of his country, but Richard cheerfully left the problems to his representatives, the chancellor, William Longchamp (*q.v.*), bishop of Ely, who was also justiciar and papal legate, and the judges who formed an intimate council of advisers. The king had given a great territorial endowment to his brother John (*q.v.*), count of Mortain, but had designated as his heir not John but their nephew Arthur of Brittany. Discontented with his position, John took advantage of the general dislike felt for the chancellor to stir up a rebellion against him before Richard had left Europe for the east. Richard therefore sent Walter of Coutances, archbishop of Rouen, who had started on the crusade with him, back to England, instructing him to support the chancellor, but, if necessary, to take control of affairs in association with the council of judges. Richard was captured while returning from his crusade in Dec. 1192 by his enemies Duke Leopold of Austria and the emperor Henry VI, and John, in alliance with the French king, Philip II, made a bid for the crown. It was an astounding testimony to the stability of the English government that these troubles shook neither the loyalty of those who sat round the exchequer board nor that of the mass of the English people. It was only within his own lands in England that John found any considerable support.

Before the emperor freed Richard in Feb. 1194, a large proportion of the 150,000 marks promised as ransom had been paid and John's rebellion had been almost quelled. It collapsed at once on the king's return. Richard stayed in England barely two months, enough for a military demonstration in John's country, a great council at Nottingham in which John and his supporters were punished by forfeiture and a second coronation to wipe out the shame of captivity. The rest of the reign passed for Richard in the defense of Normandy. In a decade of rising prices the crusade, the ransom, the French war and the purchase of alliances in the Low Countries meant unprecedented taxation. Feudal dues were increased, a scutage (or payment taken instead of military service) was accompanied by fines to exempt tenants in chief from personal attendance in the field, towns were encouraged to buy charters, the royal tax known as tallage (*q.v.*) was increased, the ancient land tax or "hidage" was reimposed and a new tax on income and movables was devised. Even these expedients did not suffice and individuals were requested to help the king in his hour of need and churches to lend him their treasures. The financial stress of Richard's closing years undoubtedly prepared the way for the rebellion in John's reign.

Nevertheless, England was well governed through these years by Hubert Walter (*q.v.*), archbishop of Canterbury, who was justiciar from 1193 to 1198, and Geoffrey FitzPeter (*q.v.*), his successor in that office. Throughout the reign the impetus given to legal development by Henry II maintained its force. The Articles of the Eyre (instructions to justices on circuit) were elaborated for the judicial visitation of 1194 and remained the standard to which inquiries could be added in later years until the general eyre was abandoned in the middle of the 14th century. The office of coroner was established and a beginning made in the organization of the exchequer of the Jews, to supervise their financial affairs in the king's interests (*see* EXCHEQUER).

As in the reign of Henry II administrative changes, legal innovations and the imposition of taxes were the responsibility of itinerant judges whose frequent appearances in all the shires made all free men aware of the virtues of the king's justice and kept the land in peace. Although no further steps were taken to incorporate Wales under English rule, successive justiciars by brief military demonstrations and the maintenance of border castles preserved the balance of power between the Welsh princes and the lords of the marches (border region). King William of Scotland remained loyal to Richard throughout the reign. John retained his lordship of Ireland despite his rebellion and before Richard died, on April 6, 1199, he had been acknowledged as his heir.

JOHN (1199–1216)

The Loss of Normandy.—When Richard died he was on the point of securing a favourable treaty with Philip II, but the existence of a rival for the English throne in Arthur, duke of Brittany, forced John to submit to the ignominious treaty of le Goulet (1200), by which he abandoned Richard's German alliances, promised 20,000 marks to the French king and yielded two disputed lordships in Berry (Berri) as dowry for his niece, Blanche of Castile, who was to marry Philip's heir, the future Louis VIII. John's hasty marriage (Aug. 1200) to Isabella, heiress of the county of Angoulême, undoubtedly strengthened his territorial position in Aquitaine, but roused the enmity of the powerful Lusignan family, since Isabella had been betrothed to one of its members, Hugh, count of la Marche. In the course of the quarrel which ensued they appealed to Philip II as overlord, who summoned John to answer their complaints in his court. A sentence of deprivation for contempt of court automatically followed John's failure to appear and the king of France attacked John's lands. The barons of the English lands in France were tired of war and the English baronage was unwilling to serve abroad. John's one

success was the capture of Arthur and the raising of the siege of Mirabeau (1202) and the one heroic episode in the war was the defense of Château-Gaillard, the castle built on an island in the Seine river by Richard I to defend Normandy. By the time it fell in March 1204 the rumours that John had murdered his nephew Arthur had spread over Europe, many of the baronage of Maine, Anjou and Normandy had deserted John and he could only accept defeat.

Nevertheless, he refused to face the permanent loss of the English lands in northern France and kept his hold on the Channel Islands as a base for reconquest. He directed his policy to the rebuilding of Richard's continental alliances and the creation of a transport fleet for his armies. He was prevented by Hubert Walter, archbishop of Canterbury, and William Marshal, earl of Pembroke, from undertaking an expedition in 1205 for which great preparations had been made. No large-scale effort could be attempted thereafter until 1214. The defeat of his allies, the emperor Otto IV and various princes from the Low Countries, at the battle of Bouvines (1214) and his own failure to bring the French to battle in Poitou ended his dreams. But already the loss of Normandy had concentrated the attention of the Anglo-Norman baronage on the country of their choice and those who had preferred their English to their Norman lands were perforce becoming Englishmen at last.

John's Quarrel With the Church.—The death of Hubert Walter in 1205 was followed by a disputed election to the see of Canterbury, which gave Innocent III an unexpected chance to extend papal influence in England. He persuaded the electors whom the dispute had brought to Rome to accept as archbishop Stephen Langton (*q.v.*), a Lincolnshire man eminently suitable for the position, save that, as John complained, he had been brought up in Paris among the king's enemies. During the subsequent quarrel between king and pope, England was laid under interdict in March 1208 and John himself was declared excommunicate in Nov. 1209. The only bishop who remained in England after this was Peter des Roches (*q.v.*), bishop of Winchester. The departure of many clerks overseas made the conduct of government and the maintenance of the courts of justice difficult. Nevertheless, by skilful negotiations John held out until May 1213 when, faced by a baronial rebellion supported by a French invasion, he agreed to receive Stephen Langton as archbishop and to hold England and Ireland as a fief of the papacy. He thus won a powerful ally in Pope Innocent III who gave him unquestioning support for the rest of his reign and prevented Philip II of France from intervening effectively in English affairs even when his son Louis was invading England. The annual tribute of 1,000 marks was a small amount to pay for the advantage which papal aid secured for John.

Baronial Opposition and Magna Carta.—Already in 1212 John had been forced to abandon a projected attack on Llewelyn ap Iorwerth of north Wales when he learned that his barons were plotting to deliver him into Llewelyn's hands. The leaders of the conspiracy, Robert FitzWalter (*q.v.*) and Eustace de Vesci, fled. After the collapse of the king's foreign policy at Bouvines, open opposition developed in England. Stephen Langton from the moment of his return had shown his sympathy with the barons and his willingness to act as a mediator between them and the king. It was he who initiated the demand for a charter of liberties based on that issued by Henry I at the beginning of his reign. He hoped in this way to avoid civil war. That John's rule had been oppressive is beyond doubt. His constant travels in England gave him and his able servants a clear understanding of where the wealth of the country lay and they used this knowledge to collect it. Feudal dues had been heavily and erratically increased and crown debtors, who in the past had been given easy terms of payment by installments, were forced to produce large sums regularly to escape prison or dispossession. Churchmen of all ranks had suffered heavy losses during the interdict and only the bishops had received compensation. The drafters of Magna Carta (*q.v.*), sealed at Runnimede in June 1215, tried to provide a cure for the ills of every class—church and barons, chartered towns and the mass of Englishmen, free and unfree. Even the baronial allies of Scotland and Wales were included. It is not surprising that such a comprehensive program failed in its immediate aim. Both king and recalcitrant barons repented of their concessions. War broke out and the barons invited Louis of France, husband of Henry II's granddaughter, to be their king. Although the fortunes of war fluctuated, there is little doubt that had John lived he would have overcome his enemies. His death (Oct. 18–19, 1216) meant that moderation could triumph.

Although the reign ended in political disaster it had seen massive development in English law, in trade and commerce, and in the political thinking of educated Englishmen, whose numbers were increasing with the growth of universities in England and in Europe. While Hubert Walter was chancellor (1199–1205) the practice of keeping copies of royal letters was begun, and Hubert's influence was probably responsible, at least in part, for the better preservation of judicial records, noticeably with the beginning of the feet of fines in 1195 and the increasing number of the surviving rolls of judicial proceedings thereafter. John's young son inherited a kingdom which a foreign prince was trying to conquer with the aid of English barons, but it was a rich kingdom full of promise for the future.

IV. ECONOMIC AND SOCIAL DEVELOPMENT IN THE EARLY MIDDLE AGES

The inquiries authorized in the Christmas council of 1085 and resulting in the compilation of Domesday Book provide an inevitable opening to the social and economic assessment of early medieval England. Apart from the omission of accounts for London and Winchester, the returns cover the whole of England except the four northernmost counties, but Cumberland had not yet been incorporated into England and the far north had not yet recovered from the upheaval of the Conquest. Everywhere a pause for the readjustment of the population and for the acceptance of the new political situation by other European powers followed the events of 1066. The revival of external and internal trade did not have full effect until Henry I's reign (1100–35), when county towns which had suffered devastation and depopulation began to advance again. Thirty years of peace under Henry I strengthened the economy to withstand the results of the disputed succession which followed his death. The Norman kings inherited from their Saxon predecessors an elaborate system of taxation and a managed national currency, both of which they took over and developed. If the actual coins in circulation deteriorated in artistic merit from the beginning of William II's reign, their weight and silver content remained stabilized. The word "sterling" came into use about this time and has been shown to mean a "tough" or "strong" piece as distinct from the lighter English silver penny which preceded it.

The Growth and Organization of Trade.—Apart from the 19 years of Stephen's reign, the 12th century was a time of rapid development in every aspect of national life. The population was increasing and prices were rising. Similar trends can be seen in other European countries, making for genuine international trade. Before the end of the 12th century surplus grain was being exported to Flemish towns. The greater security provided by strong feudal princes encouraged international trade, although it might be interrupted by the outbreak of war. At the end of the century it was no longer necessary for traders to bring their goods inland for storage and near the coast new towns grew wealthy at the expense of the old inland markets. Hedon and Hull were growing rich at the expense of York. Boston was drawing trade from Lincoln and Lynn from Norwich. Presumably the ships which brought wine to the channel ports carried back grain to the south-west of France as they did later in the middle ages. A ship wrecked on the Cornish coast at the end of the century had a cargo of wine, grain and salt. Sheep farming was increasingly profitable, as the new Cistercian monasteries founded in the first half of the century were finding. Flemish merchants were coming to England to buy wool before the end of the century, but by the middle of the century fulling mills were here and there appearing for the local manufacture of cloth. Individual merchants from the Scandinavian north (whence came the much-prized Norway hawks [gyrfalcons]), as well as from France and Flanders, were

BY COURTESY OF (1, 9) THE SOCIETY OF ANTIQUARIES, LONDON. (2) THE DEAN AND CHAPTER OF WESTMINSTER. (3, 4, 7) THE BRITISH MUSEUM. (5-8, 10-16) THE NATIONAL PORTRAIT GALLERY (8, MANSELL PHOTO)

KINGS AND QUEENS OF ENGLAND

1. William I (1066–87). From the Bayeux Tapestry. **2. Henry III (1216–72).** From his tomb in Westminster Abbey. **3. Edward I (1272–1307).** From a 14th century MS. **4. Edward III (1327–77).** From a 14th century MS. **5. Richard II (1377–99).** **6. Henry IV (1399–1413).** **7. Henry V (1413–22).** **8. Henry VI (1422–71).** **9. Edward IV (1461–83).** **10. Henry VII (1485–1509).** **11. Henry VIII (1509–47).** After Holbein. **12. Mary I (1553–58).** By Joannes Oorvus. **13. Elizabeth I (1558–1603).** **14. James I (1603–25).** By Paul van Somer. **15. Charles I (1625–49).** By Daniel Mytens. **16. Charles II (1660–85).** By Joseph Michael Wright

ENGLISH HISTORY

BY COURTESY OF (1–9) THE NATIONAL PORTRAIT GALLERY, (10) THE WALLACE COLLECTION, LONDON, (11) THE ARTIST'S EXECUTORS; PHOTOGRAPHS, (12) VANDYK, (13) CAMERA PORTRAIT BY HUGH CECIL, (14) ACME, (15) BRITISH INFORMATION SERVICES

KINGS AND QUEENS OF ENGLAND

1. James II (1685–88). By Sir Godfrey Kneller. 2. William III (1688–1702). From the studio of Wissing. 3. Mary II (1689–94). After William Wissing. 4. Anne (1702–14). By Michael Dahl. 5. George I (1714–27). After Sir Godfrey Kneller. 6. George II (1727–60). After Sir Godfrey Kneller. 7. George III (1760–1820). By Allan Ramsay. 8. George IV (1820–30). By Sir Thomas Lawrence. 9. William IV (1830–37). 10. Victoria (1837–1901). By Thomas Sully. 11. Edward VII (1901–10). By Sir Luke Fildes, K.C.V.O., R.A. 12. George V (1910–36). The official Jubilee portrait. 13. Edward VIII (Jan. 20, 1936–Dec. 11, 1936). 14. George VI (1936–52). 15. Elizabeth II (1952–)

BY COURTESY OF (ALL EXCEPT TOP ROW, THIRD AND FOURTH LEFT; BOTTOM ROW, THIRD AND FOURTH LEFT) THE NATIONAL PORTRAIT GALLERY, LONDON, (TOP ROW, THIRD AND FOURTH LEFT) MARQUIS OF SALISBURY, (BOTTOM, THIRD LEFT) REPRODUCED BY GRACIOUS PERMISSION OF HER MAJESTY THE QUEEN, (BOTTOM RIGHT) THE DEAN, CHRIST CHURCH, OXFORD UNIVERSITY

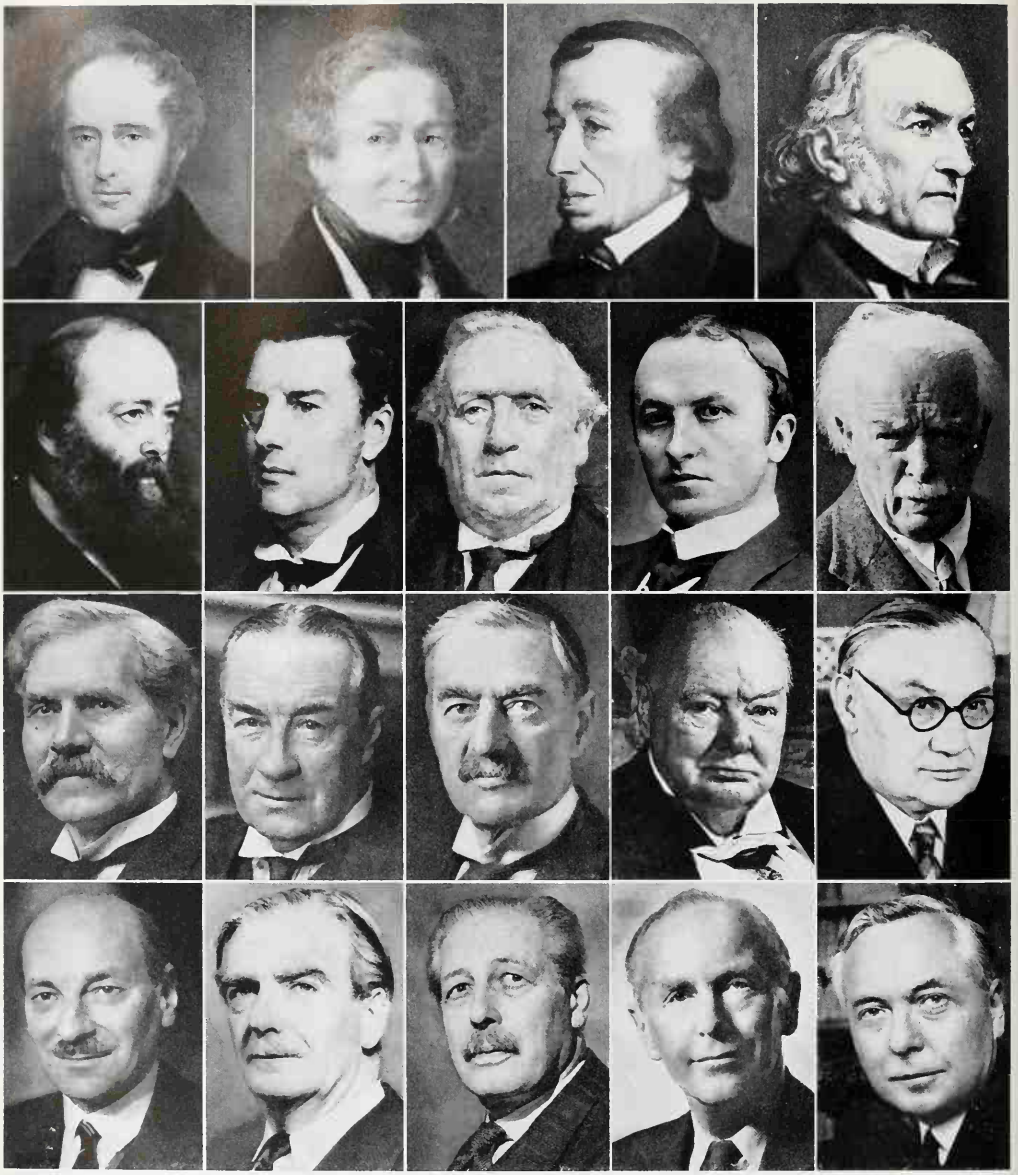
ENGLISH STATESMEN

Top row, left to right: Thomas Wolsey (*c.* 1473–1530); Thomas Cromwell, earl of Essex (1485?–1540); William Cecil, Baron Burghley (1520–98); Robert Cecil, 1st earl of Salisbury (1563–1612)

Second row, left to right: William Laud (1573–1645) archbishop of Canterbury; Thomas Wentworth, 1st earl of Strafford (1593–1641); Oliver Cromwell (1599–1658); Edward Hyde, 1st earl of Clarendon (1609–74)

Third row, left to right: Robert Walpole, 1st earl of Orford (1676–1745); William Pitt, 1st earl of Chatham (1708–78); Edmund Burke (1729–97); Charles James Fox (1749–1806)

Bottom row, left to right: William Pitt the Younger (1759–1806); Arthur Wellesley, 1st duke of Wellington (1769–1852); Robert Stewart, Viscount Castlereagh, 2nd marquess of Londonderry (1769–1822); George Canning (1770–1827)



BY COURTESY OF (TOP ROW, LEFT, SECOND LEFT, THIRD LEFT; SECOND ROW, LEFT, SECOND LEFT, THIRD LEFT) THE NATIONAL PORTRAIT GALLERY, LONDON; (TOP ROW, RIGHT) VICTORIA AND ALBERT MUSEUM, PHOTOGRAPHS; (SECOND ROW, FOURTH LEFT) RADIO TIMES HULTON PICTURE LIBRARY; (SECOND ROW, RIGHT) "ILLUSTRATED"; (THIRD ROW, LEFT) PRESS PORTRAIT BUREAU; (THIRD ROW, SECOND LEFT; BOTTOM ROW, LEFT) J. RUSSELL & SONS; (THIRD ROW, THIRD LEFT) BASSANO—CAMERA PRESS; (THIRD ROW, FOURTH LEFT) KARSH OF OTTAWA—CAMERA PRESS; (THIRD ROW, RIGHT) BARRATT'S PHOTO PRESS LTD.; (BOTTOM ROW, SECOND LEFT) WIDE WORLD; (BOTTOM ROW, THIRD LEFT) TOM BLAU—CAMERA PRESS —PIX FROM PUBLIX; (BOTTOM ROW, FOURTH LEFT) © REX COLEMAN; BARON STUDIOS; (BOTTOM ROW, RIGHT) CENTRAL PRESS PHOTOS

ENGLISH STATESMEN

Top row, left to right: Henry John Temple, 3rd Viscount Palmerston (1784–1865); Sir Robert Peel (1788–1850); Benjamin Disraeli, Earl of Beaconsfield (1804–81); William Ewart Gladstone (1809–98)

Second row, left to right: Robert Gascoyne-Cecil, 3rd Marquess of Salisbury (1830–1903); Joseph Chamberlain (1836–1914); Herbert Henry Asquith, 1st Earl of Oxford and Asquith (1852–1928); George Nathaniel Curzon, 1st Marquess Curzon of Kedleston (1859–1925); David Lloyd George, 1st Earl Lloyd George of Dwyfor (1863–1945)

Third row, left to right: James Ramsay MacDonald (1866–1937); Stanley Baldwin, 1st Earl Baldwin of Bewdley (1867–1947); Neville Chamberlain (1869–1940); Sir Winston Churchill (1874–1965); Ernest Bevin (1881–1951)

Bottom row, left to right: Clement Attlee, 1st Earl Attlee (1883–); Robert Anthony Eden, Earl of Avon (1897–); Harold Macmillan (1894–); Sir Alexander Frederick Douglas-Home (1903–); (James) Harold Wilson (1916–)

buying the royal permission to trade throughout the country. Luxury goods from the east—sugar, almonds, figs and dates—were reaching the tables of the rich before the end of John's reign.

London had always attracted a cosmopolitan population and as early as Stephen's reign its inhabitants were meditating the possibility of setting up a commune and winning the independent rule of their city. But under Henry II the governmental machine was too efficient for a commune to survive, or even to be necessary for the welfare of the city itself. A further attempt in Richard I's reign also failed, but the Londoners won the right to have a mayor as their chief officer. In the early years of the 13th century mayors began to appear in other cities and towns.

Henry II had been unwilling to extend existing privileges of towns, but his sons Richard I and John found towns good customers, ready to pay high prices for charters allowing them to pay their own dues direct to the exchequer instead of through the sheriff, to hold a market and fair, to be quit of toll, and free to regulate their own trading affairs through their merchant guild (gild) and have their own court of justice. Merchant and craft guilds both needed royal licence, but local traders often set up guilds without seeking it. A guild was an association for mutual help of men with common interests. The trade guilds of the 12th and 13th centuries were simply extensions of an ancient Anglo-Saxon practice. A merchant guild was made up of leading citizens in the various trades found in a town and although it was a body distinct from a town council the existence of a merchant guild made the appearance of a town council in the 13th century a natural development. Such councils generally consisted of 12 or 24 members. The theory that residence in a chartered borough for a year and a day made an unfree man free prevailed in England as on the continent. Already in the 12th century lords were beginning to grant charters to the towns on their land and by the middle of the next century such grants were numerous. Sometimes a lord tried to convert a country manor into a town by laying out plots of land and inviting men to come and take them up. Such a speculation might well fail if neighbouring towns complained that the new market harmed their own. Many small towns created in this way failed to survive the middle ages as flourishing trading centres. The survival of many others shows conclusively that in the 13th century trade was increasing sufficiently to justify the creation of many more markets than had sufficed for the needs of previous generations.

The Pattern of Agricultural Society.—Domesday Book records the existence of slaves in most English shires, but genuine serfdom had disappeared from the English rural scene before the 12th century was far advanced. The villeins, cottars and bordars who formed the mass of most village populations were farmers or cottagers, whose holdings varied in size as much as those of modern small farmers and small holders. (For additional details on the rules and customs mentioned below *see* SERFDOM AND VILLEINAGE.) Cottagers might have only a garden, or one acre, or five acres, or more. Much depended on past history, local custom and the nature of the land. The same was true of the villein with his virgate or half virgate of land. The acreage of the virgate varied, but it generally approximated to 20 ac. of land. For this the tenant did week work, generally one, two or more days' work on the lord's demesne farm, and helped him at busy seasons. As a general rule the services of only one man were required from each holding for the performance of the week work. Where the village lands were tilled in great open fields the villein's land would lie scattered in strips in each field and throughout the 12th and 13th centuries individual farmers were arranging exchanges with others so that their strips might lie more conveniently together. The villein farmers were in general personally unfree, unable to leave their lands without the lord's permission, paying merchet when they married off a daughter, giving their lords Christmas presents of hens and eggs. But there were many gradations of unfreedom. The records of the royal courts of justice show that many peasants were prepared to assert that they were personally free while local juries often seemed uncertain what precisely connoted freedom. Throughout these years there was a strong current moving toward the freedom of the rural population. Some men became prosperous enough to buy freedom from their lord. When a lord converted a rural manor into a chartered borough he automatically freed the inhabitants. When a lord accepted a rent instead of labour services manorial ties were weakened. The existence of the shire and hundred courts and the growing strength of the royal courts held before itinerant justices all tended to weaken the hold of the manor courts on the free and unfree farmers of the countryside.

No useful generalizations can safely be made about the economic position of the freemen and sokemen who appear in Domesday Book. Some in both classes were large farmers with tenants under them. Many had holdings smaller than those of prosperous villeins. Sokemen, who appear in largest numbers in the shires where Scandinavian settlement had followed the 9th-century invasions, were personally free, generally paying a rent to their lord, giving him help at the busy seasons of the agricultural year and attending his court of justice. Very little can be said about the village craftsmen, the millers, the smiths, the ironworkers and the masons.

The rapid development of castle-building, the churches and cathedrals springing up over England in the 12th and 13th centuries made demands on the services of such men which drew them from the farms and small holdings of their ancestors. Furthermore, there was still much land in fens, forests and marshes which could be won for the plow. As at every other period in English history able men could rise from the lowest ranks of society and the inefficient and foolish could sink from plenty to poverty.

V. THE 13th CENTURY

HENRY III (1216–72)

The Minority.—At the death of King John in 1216 London and much of eastern England were in the hands of the rebels led by Louis of France, later Louis VIII, but the removal of John's vivid personality from the political scene meant that the war would soon be ended. His son, who became Henry III, then a child of nine, had aroused no hatred; Henry's friends encouraged the view that he should not suffer for his father's sins. The papal legate, Guala, supported his cause with all the authority of the church and Henry was crowned at Gloucester on Oct. 28, 1216.

Regency of William Marshal.—The oldest and most respected English earl, William Marshal, earl of Pembroke (*q.v.*) and hereditary marshal of England, became, as *rector regis et regni* (governor of the king and of the kingdom), the first regent to rule the land on behalf of a minor. The charter was reissued in 1216 in the young king's name (*see* MAGNA CARTA) and rebels were urged to return to their allegiance. The French invaders who were besieging Lincoln castle were defeated in the city. Their reinforcements were destroyed in a naval engagement off Dover. But the marshal, who still held lands in Normandy and hoped that the old Anglo-Norman relations could one day be restored, was glad to offer Louis easy terms of peace in the treaty made on an island in the Thames near Kingston on Sept. 12, 1217. A general amnesty made it possible for many rebellious barons to go on crusade until the troubles had blown over. The council was free to rebuild the fabric of English central and local government, and in this work the justiciar, Hubert de Burgh (*q.v.*), took a leading part. The new treasurer (1217) was one of John's judges, Eustace de Fauconberg, and the chancellor was still John's old servant, Richard Marsh, the newly elected bishop of Durham.

Peter des Roches, the bishop of Winchester and one of King John's executors, could not be excluded from the young king's council, despite his unpopularity with the baronage and the church; nor could William Brewer, another faithful servant and also an executor of King John. Decisions which were to shape the policy of the new reign came, therefore, not from the rebel barons but from the wisest supporters of the old regime.

During the regency the charter was twice reissued, in 1216 and again in 1217, each time with important changes. Although the exchequer was at work again in 1217 and pleas were being heard at Westminster it was not until Nov. 1218 that the council could send out a general eyre (judicial circuit) to deal with the innumerable pleas awaiting settlement in the shires. It marked the end of the

regency, for the marshal fell ill while it was in progress and died in May 1219. Although he would have liked Pandulph, who had succeeded Guala as papal legate in 1218, to succeed him as regent, it was the justiciar, Hubert de Burgh, who, with the full support of Stephen Langton, acted for the king thereafter. Pandulph's recall to Rome in 1221 made their position simpler since it removed from the kingdom the representative of the king's overlord and left Langton head of the church in England. But Henry III never forgot how much he owed the pope for his protection and the help of his legates in those difficult early years.

The Ascendancy of Hubert de Burgh.—The war had left many problems. Innumerable cases of dispossession were dealt with by the "common writ" authorizing a local jury to inquire whether the plaintiff had been seized (in possession) of the land at issue before the war and disseized because of the war. More serious was the strong position of mercenary captains holding royal castles strengthened for war by King John; such as Philip Mark, sheriff and castellan of Nottingham, or Engelard de Cigogné at Odiham and Windsor, or Peter de Mauley at Corfe. Still more serious was the habit of mind which war had engendered in some of the king's greatest tenants. It led William de Forz, count of Aumale, to defy the king's itinerant judges in Lincolnshire and stand a siege in his castle of Bytham in 1221, and Falkes and William de Bréauté to defy the royal judges in Bedfordshire and raise a rebellion there in 1224. In each case the feudal levy was summoned and in each case the rebellion was quickly suppressed. The justiciar's success in building a marcher principality around the castles of Grosmont, Skenfrith and White castle between the rivers Usk and Wye did little to prevent Llewelyn of north Wales from encroaching further on the south and central marches and contributed, as did his marriage (1221) to Margaret, sister of Alexander II of Scotland, and his new title, earl of Kent (1227), to the growing hostility felt against him by long-established baronial families. No credit came to the justiciar from English affairs in France. Isabella of Angoulême, King John's widow, returned to her inheritance in 1218, and in 1220 married Hugh de Lusignan, son of the Hugh to whom she had once been betrothed. The union of Angoulême and Lusignan altered the balance of power in southern France and successive seneschals of Poitou, as well as the individual cities, continually wrote to the king stressing the dangers to his interests there.

The personality of the young king began to influence events as the first full decade of the reign drew to its close. The pope had declared him of age to have a great seal in 1223 and the reissue of the great charter under the new great seal in 1225 marked a turning point. After the king had proclaimed himself of full age (1227) and Stephen Langton had died (1228), the justiciar found it progressively more difficult to retain his complete authority. The death of Louis VIII in 1226 had left Blanche of Castile to face the problems of her son's minority and seemed to offer an opportunity for a revival of English influence in France. But the chance was missed. Henry found that his forces were still unready to cross the channel when he reached Portsmouth in Oct. 1229 and he accused Hubert of treachery. When the expedition took place in 1230 it was no more than a costly demonstration, securing nothing. The return (1231) from crusade of Peter des Roches brought back to English politics a powerful personality, bound to exert an influence on an impressionable young king who had once been his pupil. The fall of Hubert de Burgh was not long delayed.

The Fall of Hubert de Burgh.—The fall of a minister so strongly entrenched as Hubert de Burgh could only be secured by secret and complex action. Without doubt the Poitevin Peter des Roches was the prime mover, with his son or nephew, Peter des Rivaux, who returned in 1230 to the royal household where he had served from 1218 to 1223. That Hubert foresaw his danger is suggested by the solemn pact sworn in July 1232 between the king, the justiciar and his wife, and the great officers of state and household, by which the king solemnly promised to observe the charters granting to all these officers their positions, castles and perquisites for life. Nevertheless, within a month the process of stripping Hubert de Burgh of power had begun. The new justiciar was Stephen of Seagrave, the senior member of the bench, whose appointment indicates a growing feeling that the justiciar should be first of all a judge, not a politician. Peter des Rivaux became treasurer of the exchequer as well as of the wardrobe, in place of Walter Mauclerc, bishop of Carlisle, but the chancellor, Ralph de Neville, bishop of Chichester, remained in office.

Peter des Roches held no office of state, but his influence and that of his Poitevin associates with the king, and Peter des Rivaux's overhaul of local government finance, aroused strong baronial opposition led by the younger Richard Marshal, earl of Pembroke, who found that Hubert de Burgh's marcher lordships had passed into Peter des Rivaux's hands. In alliance with Llewelyn of Wales the earl of Pembroke opened war in the marches and then, fearing that his Irish lands were in danger, went to defend them. His death in Ireland in April 1234 ended the brief episode of the so-called Poitevin authority, for under the leadership of the new (1234) archbishop of Canterbury, Edmund Rich, the great council forced the king to denounce the bishop of Winchester and his associates, including the justiciar, Stephen of Seagrave, as traitors. With the fall of Peter des Roches the last of the great figures of John's generation passed out of English politics. Henry III appointed no justiciar to succeed Stephen of Seagrave. The king's personal rule had begun.

The Personal Rule of Henry III.—Henry's personal rule ended 24 years later in rebellion in 1258; but it was in no sense a period of stagnation or failure. It was the central period of a reign which saw rapid development in learning, in Latin and vernacular literature, in architecture, both ecclesiastical and domestic, and in all the arts which serve it. (*See* UNIVERSITIES: *English Medieval Universities;* ANGLO-NORMAN LITERATURE; LATIN LITERATURE; EARLY ENGLISH PERIOD.) By his ready support of scholars and artists the king himself did much to make his reign the most brilliant age of English medieval civilization.

The very rapidity with which English society was changing in response to increasing wealth, improved agriculture, the rapid development of the common law (*q.v.*) and of political thinking made for the troubles which culminated in rebellion. The most striking achievement of these years is the majestic progress of the common law exemplified in Henry de Bracton's massive treatise *De legibus et consuetudinibus Angliae* ("The Laws and Customs of England"). Written before 1256, it was based on the judgments of the great judges of the previous generation and on the ever-increasing mass of proceedings in the royal courts of justice and before the justices in eyre. Before the end of this period textbooks were being compiled to train the rapidly increasing numbers of students preparing for careers in the law. Of necessity, legislative activity followed on this rapid advance. The Statute of Merton of 1236 made a number of important changes in matters frequently before the courts: actions of dower, the rights of widows in their dower lands, the enclosure of common land and the protection of infant heirs from usury. In the same council at Merton the barons made their famous stand against altering the law of England to allow the subsequent marriage of their parents to legitimize children born out of wedlock. In 1256 the council put out a definite direction about the calculation of a year and a day in leap years.

The king's marriage in 1236 to Eleanor of Provence brought to England many Provençals and also Eleanor's uncles from Savoy: William, to whom Henry III wished to give the see of Winchester; Peter, who ultimately became count of Savoy; and Boniface (*q.v.*), who succeeded Edmund Rich at Canterbury. They brought with them foreign clerks, as anxious for money and position as they were themselves. The king was ready to employ them and they served him well. He gave a warm welcome to his own half brothers from Lusignan who came to England with their dependent clerks in 1247. His household organization was dominated from 1240 to 1258 by a succession of foreign keepers of the wardrobe, which had become a department of state as important as the exchequer and chancery. Financial developments, set in motion when the Poitevins were in power, were carried to their logical conclusion. The sheriffs became salaried officials, accounting as strictly for the king's manors in their shire as for the profits of the shire and hundred courts. Much of what was done was well

done, but Henry's generosity to his kinsmen and their dependents led him into difficulties long before 1258.

Not only the king had welcomed the young Simon de Montfort (*q.v.*) as heir to the earldom of Leicester in 1231. The charm which won Simon general acceptance won him also the king's sister, Eleanor, widow of William Marshal. Their marriage in 1238, of which few but the king were at first cognizant because Eleanor had at the age of 16 taken a vow of chastity when she was widowed, brought resistance to the king's policies to a head. The baronial opposition was led by the king's brother, Richard (*q.v.*), earl of Cornwall. The trouble passed, but it was possibly at this time that the barons drafted a plan of reform strikingly similar to that evolved in 1258. They wanted four conservators of the liberties to see that the Charters of Liberties and of the Forest were faithfully observed, more speedy and impartial justice, and a justiciar and chancellor chosen by all the baronage and removable only by them. The precise date of this plan, which has been disputed, is immaterial, for it expresses baronial opinion held throughout these years. The Charters of Liberties were often referred to in the courts of justice, and the royal minority had given a new status to the king's council, making the barons deeply concerned to preserve their full rights as the king's counselors. The need for a council smaller than the great council was not new, but it was not until Henry III's reign that the practice of its members taking a councilor's oath was formally established. The chronicler Matthew Paris records meetings of the great council in words which suggest that its character was changing. Meetings of the magnates "to treat of the affairs of the kingdom" had become occasions when the king asked for money and the barons tried to refuse it (1242), or consented grudgingly after the king had agreed to confirm the charters (1244). The occasional notice of a summons of knights of the shire before the council (1254), the recurrent mention of a "great parliament" or a "parliament" on occasions when the word council was used in former days, and the suggestion that parliament means not only a place for discussion but also a place for justice to be administered were also signs of change.

The Barons' War: *Proximate Causes.*—Henry III's own ineptitude, the ability of the French king Louis IX and the success of his brother Alphonse in Poitou and Toulouse ended any prospect of recovering the English lands in France. Gascony alone remained, and that was held only tenuously after Henry's costly expedition in 1242. Simon de Montfort had to admit in 1252 after four years as seneschal that he, like his predecessors, had failed. Gascony became part of the appanage of Henry III's young heir, Edward (*see* EDWARD I), and in 1253–54 the king himself spent 15 months there trying to make peace by lavish expenditure instead of force. Edward's marriage (1254) to Eleanor of Castile was part of the settlement, but no permanent treaty was secured with France. In the same year the king accepted from Pope Innocent IV for his infant son Edmund the offer of the crown of Sicily. Richard of Cornwall had refused this two years before, but in 1256 he allowed himself to become one of the candidates for the imperial title. The English clergy were already restive under demands for tithes for the recent papal crusade against Frederick II. The bishop of Lincoln, Robert Grosseteste, had already voiced English resentment at the papal patronage which set Italian-speaking foreigners in English benefices. The earl of Cornwall's expenses in Germany were his own affair, but the barons and clergy of England had no mind to see the pope fight his enemies in Sicily with English money. In April 1258 a confederation of barons took the first steps toward an overhaul of the government of England.

The Provisions of Oxford (1258).—The magnates presented the king with a plan of reform in the Provisions of Oxford (*q.v.*) at the parliament held there in June. Their list of matters requiring amendment covered the whole field of local government from the appointment of aliens as castellans to the difficulty of holding a grand assize because of exemptions from jury service. Foreign affairs were not included although they had stung the barons to action. A new justiciar was appointed, Hugh Bigod, sworn to do his work justly and made answerable at the year's end to the king, the council and his successor. The new treasurer and the chancellor were likewise made answerable at the year's end. Four knights were chosen in every shire to prepare for the coming inquiry by the justiciar into abuses throughout the land. The knights were to be a standing local committee. The barons showed their distrust of the wardrobe by insisting that "all the issues of the land" should come into the exchequer "and in no part elsewhere." The admonition "Let the charter of liberty be firmly kept" indicated the baronial purpose. A standing council of 15 barons was appointed with whom in three annual parliaments a committee of 12 was to work, in order "to spare the cost of the commonalty." Important legislative changes, many of them long overdue, were made in the Provisions of Westminster on Oct. 13, 1259.

Baronial Government and the Outbreak of War.—English affairs were left to the justiciar while the king with Simon de Montfort was in France seeking a settlement with Louis IX. By the treaty of Paris, ratified after many delays in Dec. 1259, Henry abandoned English claims in Normandy, Anjou, Maine and Poitou, but Simon de Montfort's insistence on his wife's right to compensation delayed the settlement. Throughout the time that the baronial committees were at work Henry was seeking to recover his lost power. By 1261 he was the more anxious because his heir, now of full age and influenced by Simon de Montfort, listened with sympathy to the advanced reformers and the complaints of the unestablished feudal elements known as the "bachelry of England." His father seems even to have feared that Edward might be ready to replace him as king. Richard of Cornwall returned from Germany (Oct. 1260) in time to reconcile father and son. A papal bull of April 13, 1261, absolved Henry III from his oath to observe the provisions and the death of Richard of Clare, earl of Gloucester, in 1262, brought Simon de Montfort to the leadership of the baronial party. But the king's position was far stronger than it had been in 1258. Many of the older barons were out of sympathy with Simon and the extreme reformers. Both parties agreed to put the question of the Provisions of Oxford to the arbitration of Louis IX, whose decision, the "Mise of Amiens" (Jan. 1264), in favour of the king inevitably led to war. In the battle of Lewes in May 1264 the king was defeated and captured and the barons were back in power.

Simon de Montfort's Rule.—Simon was at last the dominating spirit, but his followers were becoming restive under his arbitrary authority. His association with Llewelyn ap Gruffydd of Wales and his establishment in the castles of Chester and Bristol, part of Edward's appanage, alienated supporters in the marches. The parliament of Jan. 1265 with which Simon's name will always be associated was no real landmark in parliamentary history. Although in his desire for more general support he summoned two representatives from each borough as well as two knights from each shire to be before the council in parliament, this was a precedent which was not immediately or uniformly followed. When this parliament met it showed that Simon's influence was waning. He quarreled publicly with Robert de Ferrers, earl of Derby. The new earl of Gloucester, Gilbert de Clare, previously his disciple, deserted him to aid Edward to escape from supervision (May 1265) and raise forces in the marches. The end of Simon's career followed swiftly. His army was annihilated and he himself was killed at the battle of Evesham in Aug. 1265. Isolated garrisons of rebels held out desperately in the Isle of Ely, the Isle of Axholme (Lincolnshire), and in Kenilworth castle and it was not until 1266 that a settlement was at last achieved in the *Dictum de Kenilworth.* A long series of judicial visitations determined the penalties laid on individual rebels and dealt with outstanding problems left over from the war.

The Close of the Reign.—The period of baronial rule and rebellion made little permanent change in the government of England. It pressed home the need for a permanent standing council of sworn members to advise the king and keep him in touch with national feeling. It demonstrated the need for continuous watch to prevent the army of minor officials in the central and local government from preying on the countryside. It showed how much legislation was necessary to direct the inevitable development of law. That Edward was aware of all these things is shown by the speed with which he ensured the issue of the Statute of

Marlborough of 1267, collecting and ratifying the legal changes made in the provisions issued by the baronial committees, and by the great spate of statutes that marks his own reign. The king's household was little affected by the rebellion, for the wardrobe had become a necessary third department of state, supplementing the work of exchequer and chancery. The last years of the reign passed so quietly that Edward could leave England (1270) to fulfill his crusading vow conscious that his succession was secure. His father died in his absence on Nov. 16, 1272. (D. M. S.)

Edward I (1272–1307)

Apart from Edward I's own doings, which were in themselves of great significance, the years of his reign saw the climax and the end of the material prosperity of the medieval English kingdom. Soon after 1300 there were signs of this on every side. Population, which had risen slowly but markedly since the Norman conquest, began to decline. The agricultural prosperity of the 13th century tailed off into the beginnings of that long period of agrarian depression destined to last until the third quarter of the 15th century. English society in the years just before 1300 had its full share of economic troubles and social tensions, but they were neither so malignant nor so intractable as those which were to plague it for the next 150 years. Nor had men yet begun to question the spiritual and intellectual assumptions to which the medieval social order itself was anchored. Compared with what was to come soon afterward the time of Edward I was a secure, self-confident and flourishing age.

The King and His Ministers.—Edward himself was typical of his time. He was content with the established order, and his mind moved easily within the limits of the ideas of his age. He was neither rebel nor innovator. That did not prevent him from being a soldier and statesman of outstanding capacity, who set a deep and lasting mark on the kingdom he ruled, who was alert to keep abreast of the times and who could neither be trifled with nor safely ignored. His tall, straight figure and the extraordinary physical toughness, which did not desert him even in old age, were matched by an iron will and a sternness of purpose that left little room in his nature for mercy or pity. As king he saw his duty in simple, practical terms and he never flinched from it, although in later life he made some grievous mistakes. His business was to maintain the rights belonging to his crown and dignity and, if possible, to enlarge them. Other people had rights too, particularly his great subjects both lay and ecclesiastical, and these he was also prepared to maintain provided they were not exercised to the detriment of his own authority. Again, Edward recognized the fact of social change, and he did his best to make the legal and political adjustments which were needed to bring his authority up to date and to keep it efficient. This was the main purpose behind the law reforms of the years before 1290, and it also determined his use and development of parliament as an institution.

Finally, Edward had the knack of choosing capable and devoted servants, and much of the credit for his political achievements really belongs to them. He relied chiefly on a small group of ecclesiastics, trained in his household and rewarded in the conventional manner with bishoprics. The most able of them was his trusted friend Robert Burnell, bishop of Bath and Wells, who served as chancellor until his death in 1292. It is certain that Burnell had much to do with drafting the great statutes and it is likely that, had he lived, Edward might have been saved from some of the political blunders he later made. Other men of this type, less well known but hardly less important, were William of Louth, bishop of Ely, Walter Langton, bishop of Lichfield, and John of Droxford, bishop of Bath and Wells. On occasion Edward also secured the services of able foreigners like the Burgundian noble Otto of Grandson and the Italian lawyer Franciscus Accursius, while he did not scorn the help of such great English nobles as the earls of Pembroke and Lincoln.

Edward's reign divides neatly into two parts. The first and the most successful was the period from 1272 to 1290. In these years the great statutes were made and the conquest of Wales was carried through. The second period reached from 1290 until the king's death in 1307. It witnessed his costly and unavailing efforts to stretch his authority over Scotland, his troubles with his French relatives about his rights in Gascony, his long and bitter quarrel with the English magnates and his ruthless efforts to scrape together sufficient money for his wars.

Methods of Government.—Edward I governed efficiently, perhaps more efficiently than any other medieval English king, but he was not, as scholars have sometimes asserted, the founder of English representative government. He used established methods of administration in a highly intelligent way, elaborating and adapting them to reinforce his personal authority. Under Edward I, as in his father's time, the seat of power was in the king's household. There it lay with a small council of trusted officials, prominent among them the chancellor, the treasurer and the senior wardrobe clerks. It was these men who, on the king's behalf, reached out to control the chancery, the exchequer and the benches, which were now independently organized to carry the heavy routine burden of government. From the centre the tentacles of royal power were still spread over the face of the kingdom in the accustomed manner by the regular journeys of the king's justices and through the continuous, if not always well-disciplined, activity of sheriffs, escheators, coroners and bailiffs in the shires and hundreds.

But royal government, on the scale conceived and practised by Edward I, required the active co-operation of his subjects to an ever-increasing extent, and this the king contrived to secure. In the shires more and more of the king's business was regularly loaded onto the shoulders of the knights. Collectively, as the community of the shire, they answered before the justices. Individually they were assigned to keep the peace, to raise money and troops, to gather information the king required and to execute the judgment of his courts. It was in the 13th century, and particularly in the time of Edward I, that the foundation was laid of that tradition of unpaid public service in local government by the gentry which was destined to be a notable feature of English life for many centuries. In the towns, too, the same thing was beginning to happen, as wealthy burgesses exercised their dearly bought privileges of municipal self-government under the watchful eyes of the king's justices.

Parliament.—Nor was the co-operation of Edward I's subjects enlisted only in the shires and the towns. Its most spectacular and enduring expression was in those assemblies, each lasting for a few weeks at most, which men were slowly learning to call parliaments. Edward did not deliberately create parliament (*q.v.*) as a means of summoning the nation to share in the business of government; what he did was more prosaic and in keeping with the ideas of his age. From the beginning of the reign the ancient practice of summoning the magnates, lay and ecclesiastical, to treat with the king on the affairs of the realm was combined with a newer device whereby the king, the judges, the barons of the exchequer and the regular members of the council came together at intervals to decide judicial problems with which the ordinary courts were unable to deal effectively. Such meetings were held when the king wanted them; sometimes there were two or three in a year; sometimes, especially toward the end of the reign, there were none for several years. They were attended by those Edward asked to come, and his council had pride of place there. Indeed, when all the others had been sent home, the members of the council remained to complete the business, and the proceedings of the parliament did not formally end until this task was accomplished. To about one in every seven of these meetings Edward, following precedents from his father's time, summoned knights from the shires and burgesses from the towns to appear with the magnates. They were to come armed with *plena potestas*, that is, with full power from those who sent them so that these latter might be legally bound to pay such sums as were agreed in parliament. Many different things might be done in parliament at this period—aids demanded, homage rendered, laws promulgated and high politics discussed—but there is no doubt that its most characteristic activity was to dispense the highest kind of justice.

Under Edward I parliament was still very malleable. Justice, politics and finance were jumbled together among its business because the king and his principal subjects found this convenient, and for the same reason its members were not present in accord-

ance with any clear or predetermined scheme. As yet nobody had bothered to ponder where such practices might eventually lead.

The Great Statutes, 1275–90.—It has been truly said that a thorough commentary upon the statutes of Edward I would be in effect a history of the common law from the 13th century down to the close of the 18th. Nevertheless it should be recognized that Edward and Burnell worked with more immediate purposes in view, and it is those purposes that call for emphasis here. There is also the fact that some of the most important legal changes have to be sought elsewhere than in the statutes themselves; they are to be found in administrative orders and judicial rulings. For instance Edward's definition of the powers of the church courts, which remained effective until the Reformation, is contained in the writ *Circumspecte Agatis* (1285).

Among the statutes only two, Mortmain (1279) and *Quia Emptores* (1290), are well constructed; the others are badly arranged and loosely drafted, while individually they often deal with a variety of different matters. The texts were apparently the work of the judges in the council and some of them at least were formally approved in assemblies where lords, knights and burgesses were present. These various enactments were preceded by a searching inquiry into the state of the kingdom carried out by special commissioners immediately after Edward's return from the east in 1274. The evidence they collected is gathered into that great compilation familiarly known as the Hundred Rolls. Armed with this knowledge the king and his helpers addressed themselves to a major program of reform and reorganization, introduced by statutes of four main types.

Administrative Statutes.—First, public order was to be strengthened, the administration of justice improved, the excesses of royal officers curbed and military arrangements overhauled. The Statutes of Westminster I and *Districtiones Scaccarii* (1275) struck at the malpractices of local officials in judicial matters. Westminster II (1285) established the practice of holding assizes three times each year and initiated the more convenient nisi-prius system, whereby many civil actions could be heard locally and not in London.

The Statute of Winchester (1285) was the most comprehensive of these administrative measures. It extended the use of collective fines on hundreds which failed to present suspects to the justices. Henceforth every man under 60 was to arm himself at his own expense and according to his means to defend the kingdom and to assist in maintaining order. Constables in every hundred and liberty (*qq.v.*) were to make regular inspection of arms and compile muster rolls of those liable to serve.

Statutes in Defense of Crown Rights.—There was also need to reassert and define many of the traditional rights of the crown usurped by others during the troubled years of the Barons' War (1263–67), and to prune even further the franchisal jurisdictions which still inhibited the operation of the king's justice in many districts. The sharp-edged writ of quo warranto bit deeply into this tangled problem. It required those who claimed franchises to produce satisfactory titles or to suffer confiscation of their privileges. Under the Statute of Gloucester (1278) only a royal grant of the rights claimed was acceptable, but by the Statute of Quo Warranto (1290), in certain circumstances, the king recognized long use, that is, undisputed exercise of such rights stretching from a time before the accession of Richard I (1189), as a sufficient title. Edward I did not destroy private jurisdiction in England; what he did do was to bring very near its conclusion the process begun in the 12th century, whereby jurisdictional privileges like view of frankpledge (*q.v.*), return of writs and possession of a gallows ceased to be a source of power to those who had them and became no more than a steadily diminishing source of profit.

Statutes Reorganizing Land Law.—Much existing law needed bringing up to date. Not only had the 13th century been an age of great legal progress, when many new remedies and better forms of action were discovered, but it had also witnessed such a growth of complication in the feudal scheme of land tenure that the system was threatening to break down. Therefore, besides authorizing many necessary procedural innovations as in the Statute of Westminster II, Edward thoroughly overhauled the law relating to land. This was done not to destroy but to maintain feudal rights and duties in land, although by Edward's time these rights had become no more than valuable property rights regulated by feudal conventions. The eroding action of social, political and economic change had already stripped away the characteristically feudal notions of personal loyalty and dependence. The problems with which Edward had to grapple here were very practical and of great social consequence. How could a lord compel a recalcitrant tenant to perform the services in return for which he held his land? How could a tenant prevent his lord from exacting services larger than those which were really due? Even worse, what was to happen when a lord, by the deliberate action of a tenant, found himself deprived of his right of escheat (*q.v.*) or of marriages and wardships, that is, of the temporary but profitable control he was accustomed to exercise over a tenant's estate during the minority of a tenant's heir? Such a deprivation happened either when the tenant granted his fief to a religious corporation, which was technically immortal and childless, thus voiding the right of escheat, or when he subinfeudated it, that is, sold it to another for a substantial sum, which the lord could not touch, in return for some nominal service. As for subinfeudation, that was prohibited by the statute *Quia Emptores*, although it could not stop the slow extinction of the lord's rights in other ways. Alienation in mortmain, that is, the granting of land to the church, was forbidden by the Statute of Mortmain (1279), though its effects were weakened by the regular issue of licences to alienate in mortmain in return for a money payment to the crown. On the other hand new forms of action, authorized by judicial interpretation of the Statute of Marlborough (1267) and the Statute of Westminster II, protected tenants from the worst effects of seignorial distraint, that is, the temporary confiscation by which some lords sought to exact additional services. (*See* REAL PROPERTY AND CONVEYANCING, LAWS OF.)

Statutes Regulating Commerce.—Finally, vigorous efforts were made to grapple with a new set of legal problems which arose from the growth of towns and trade. Commercial activities were not easily accommodated within the framework of a society still fundamentally feudal and agrarian in character. This raised a host of practical issues, prominent among them the problem of efficient debt collection. Edward's most systematic efforts to deal with such matters were made in the Statute of Acton Burnell (1283) and the Statute of Merchants (1285), and they were sufficient to shape the practice of the common law toward business transactions for the next five centuries.

Edward I used to be called the English Justinian, but that title flattered him. Nevertheless his legal changes left a deep and lasting mark, not that he particularly intended they should but because, on the whole, they were well made and they were impressed on English law at an important formative stage.

The Conquest of Wales.—Of his various interests outside the English kingdom proper, Wales caused Edward I most difficulty before 1290. Llewelyn II ap Gruffydd, the old ally of Simon de Montfort, had profited from the civil wars of 1263–66, and for three years after Edward's return from the east in 1274 he refused to do homage for his principality. In 1277 Edward launched a systematic attack upon Llewelyn, cut off his food supply from Anglesey and penned him in the mountains of Snowdonia, where famine compelled surrender. In consequence Llewelyn lost the lands he had gained between the rivers Conway and Dee and on the upper reaches of the Severn and the Wye. Nevertheless, he was allowed to marry Eleanor de Montfort and Edward, it seems, was content. But the steady pressure of English power on Welsh society from the palatinate of Chester and the marcher lordships did not relax and in 1282 Llewelyn and his brother David (III) rebelled again. This time Edward acted decisively, even summoning his Gascon vassals to his aid. Llewelyn was killed near Builth in Dec. 1282 and with his death the main Welsh resistance collapsed, although David struggled on a few months longer, until he was taken and beheaded as a traitor at Shrewsbury.

Faced with the task of making a new settlement of Welsh affairs, Edward I acted in characteristic fashion. After Llewelyn's fall two small Welsh ruling houses remained. There was Rhys ap

Maredudd in the Towy valley and the family of Powys in the neighbourhood of Welshpool. They were left undisturbed. But the powerful claims of the marcher lords could not be disregarded. Accordingly their domains were considerably enlarged; for example, Chirkland went to Roger Mortimer and Denbighland to Henry de Lacy, earl of Lincoln. Edward kept Llewelyn's principality for himself, and by the Statute of Rhuddlan (March 1284) it was annexed to the crown. It was straightway organized into three new shires, Anglesey, Caernarvon and Merioneth, and Edward deliberately created what amounted to a great royal march there. Five new castles, Conway, Caernarvon, Beaumaris, Criccieth and Harlech, were raised, each sheltering a new town designed as a centre of English trade and influence. Away to the south the ancient counties of Cardigan and Carmarthen were formally organized in the same way, while Flintshire arose as a dependency of Chester.

After a serious Welsh rebellion in 1294, which was only repressed with difficulty, Edward in 1301 revived the title of prince of Wales in favour of his eldest son, whom he invested with this new royal dependency, which henceforward became the regular appanage of the heirs to the English throne.

Edward effectively destroyed Welsh independence, but he left Wales as a land of small independent communities, where native custom and tradition lingered, if they did not flourish. The barons of the march held fast to their privileges and exceptional jurisdictions, although the need for them was now much diminished. Despite vigorous action, Edward was never able to break the custom of the march, with the private war it permitted and the lawlessness that it produced. It was left for Henry Tudor to complete the disciplinary process by making Wales an integral part of the English kingdom.

The Reorganization of Gascony.—From 1286 to 1289 Edward I was chiefly busied with his interests in France. His queen, Eleanor of Castile, possessed the little fief of Ponthieu, round the mouth of the Somme, inherited from her mother. Edward himself was lord of Gascony. As such he was continuously involved in highly complicated dealings with the French king and with his own Gascon subjects. These matters had an important bearing upon the course of English politics until the outbreak of the Hundred Years' War (1337).

In his dealings with the French king, Edward I had to face the perennial problem of the extent of his Gascon fief. In Gascony, as in Wales, his dominion did not involve the direct rule of a compact block of territory. Instead Edward claimed tenurial and jurisdictional rights in a number of adjoining districts in southwestern France, rights which had been secured by the treaty of Paris (1259). This agreement, which was in part highly ambiguous, had never been fully executed. Edward already had the rights to which he was entitled in the bulk of these southern lands, that is, in "Bordeaux, Bayonne, Gascony and the Islands." He made good his claim in the Agenais in 1279, and in Saintonge, south of the Charente river, in 1286, but he was still deprived of privileges in the three dioceses of Cahors, Limoges and Perigueux.

In Gascony Edward was the vassal of the French king, subject to the jurisdiction of the *parlement* of Paris and amenable to the involved feudal conventions which then governed such relationships; and Gascon affairs were always peculiarly sensitive to discords elsewhere between the rulers of England and France, particularly in the Low Countries and on the narrow seas (the name given to the English channel and the Irish sea). Clearly the proper management of such an involved relationship was no easy matter, but before 1290 Edward was reasonably successful, although afterward it was a different story.

Among Edward's Gascon subjects the nobles were very prominent. Like the Welsh marcher lords they were turbulent and highly privileged. Nor were they averse to making mischief between Edward and his French overlord if that forwarded their own interests. Bribes and threats were constantly needed to keep them in order, but from 1286 to 1289, when Edward was himself in France and was tightening his grip on the duchy, he sought to contrive a counterweight to their influence as well. He looked to the townsmen, not only in Bordeaux and Bayonne but in smaller places, often no more than large villages, to men who were interested in the English connection through the wine trade and to whom public order meant prosperity. He gave them privileges and built fortifications, thus creating the fortified towns, or bastides, which were to become strongholds of ducal power rather than centres of English influence. It is certain that, by the time Edward hurried back to England in 1289 to deal with insubordination and corruption among his servants there, he had done what he could to prop up his precarious authority in Gascony.

The years 1289–92 form the hinge of Edward's reign. In 1290 he expelled the Jews from England. He had already banished them from Gascony. Unlike his predecessors he could dispense with them as bankers and moneylenders, for the great Italian firms were ready to supply his needs. Then also death struck hard among those nearest to the king, for his queen and his friend and chancellor Burnell were plucked away. Above all else the business of the Scottish succession now loomed large and Edward, who had already risked much to assert his rights in Wales and Gascony, was not the man to stand aside when a tempting chance of realizing his ancestral claim to control the Scottish kingdom presented itself.

The Years of Trouble and Decline.—*Scotland.*—The second half of Edward's reign indeed was overshadowed by repeated but unsuccessful attempts to extend his authority over Scotland. This fair prospect, which seemed alluring and proved illusory, was opened by the death of the Scottish king, Alexander III (1286). The course of the Scottish struggle is discussed in articles relating to that subject. (*See* SCOTLAND; BALLIOL, JOHN DE; WALLACE, SIR WILLIAM; ROBERT I.) Its repercussions on English politics, however, were profound and call for mention here. Alone the Scottish enterprise would have strained the resources of the English crown to their limit, but Edward could never deal with it alone. After 1294 he was also at war with Philip IV (Philip the Fair) of France, for disputes between Norman and Gascon sailors had fanned into flame the discords of Edward and his overlord in the Gascon duchy, and the *parlement* of Paris declared the English king contumacious and his French lordship forfeit. Edward did what he could to protect his own, but Bordeaux and Bayonne were soon occupied, and elsewhere his troops made little headway against the invaders. Worse was to follow in 1295, for a Franco-Scottish alliance hooked the two issues of Scotland and Gascony tightly together. In consequence, from 1295 to 1297 the need to fight in Scotland, Gascony and the Low Countries, besides supporting his continental allies against Philip IV, laid on Edward an intolerable military and financial burden. This he could only hope to meet by stretching his rights over his English subjects to their furthest limits.

Clerical and Baronial Opposition.—When Edward tried to rally English support and to gather English money to meet his pressing needs, his troubles increased still further. The English clergy, led by the archbishop of Canterbury, Robert de Winchelsea, and supported by the pope, Boniface VIII, refused Edward's demands on ecclesiastical revenues. By the spring of 1296 the clergy were outlawed, while those who tried to do the king's bidding were threatened with excommunication. The papal explanation that voluntary gifts or contributions necessary for the defense of the kingdom were not forbidden coupled with William Wallace's raids into northern England put the clergy in a more accommodating mood, so that in the autumn of the same year they granted in part what the king had asked. But the domestic crisis took a new and even more serious form in 1297. General discontent was widespread because of the unprecedented prises (or requisitions), customs and taxes the king had already been forced to take; and now the great lay nobles, with whom, taught by the lessons of his father's time, Edward had hitherto been on good terms, turned against him. Faced with a situation more familiar to their fathers than to them, the barons pressed for a royal confirmation of the Charter of the Forest and of Magna Carta, brought up to date with six additional clauses, designed to prevent a repetition of Edward's recent arbitrary acts. In Oct. 1297 at a parliament in London, Edward of Caernarvon, on his father's behalf, swore to observe this *confirmatio cartarum* as it was called.

Nor did the matter rest there. Although peace was not made un-

til 1303, actual fighting with the French ended in the autumn of 1297 and that eased matters for Edward I. Nevertheless he persisted with his task in Scotland, and piled up a great load of debt for his successor. Edward fretted under the baronial restraints of the *confirmatio* and tried every dodge he knew to escape them. With papal help he finally managed this in 1305. But it was a hollow victory, for it provoked real disillusionment among the English magnates. They lost faith in the value of their traditional relationship with the crown and, when opportunity offered in the next reign, they sought new and more effective ways of maintaining themselves at the centre of affairs. Edward I died on July 7, 1307, at Burgh by Sands on his way north to yet another Scottish campaign. There is no doubt that he was in a measure responsible for the troubles which befell his son.

VI. THE 14TH CENTURY

Edward II (1307–27)

The King and the Baronial Opposition.—Edward II inherited a great burden of trouble from his father, and he himself was ill-fitted by training and inclination to bear it. The Dominicans benefited considerably from his undoubted piety, but the new king's other contemporaries were quick to mark his faults, his taste for low company and odd recreations, not to mention the inordinate affection he lavished on the Gascon knight Piers Gaveston. The real trouble, however, lay deeper. A 14th-century king was expected not only to reign but also to rule. When Edward II succeeded his father in July 1307 he had to take upon himself a heavy and exacting load of day-to-day responsibility in council, and this monotonous grind was not to his taste. It seems he neither could nor would give the unremitting labour his office demanded. Here Edward II had not been helped by his upbringing, for Edward I had not troubled to provide his son with anything like such a good political apprenticeship as he himself had received. All this was painfully evident in the first few months of the new king's reign.

Edward II did not press on with the Scottish campaign which his father had been about to begin when death overtook him. Instead he turned back and left Robert I the Bruce to gather fresh strength for reducing the numerous English garrisons in the great castles of southern Scotland. It was not to be expected that the new king, or anyone else for that matter, could then have seen that the Scottish enterprise was already a lost cause, as in fact it was. Where Edward II failed badly was that he straightway contrived to put himself in a situation which, in the course of the next few years, made him powerless even to defend the English border against Scottish raiders, let alone to continue the fight in Scotland itself. This happened when he imprudently provoked a long and bitter quarrel with some of his most powerful subjects including his cousin, Thomas of Lancaster. Their patience had already been tried to the limit by Edward I, but in his time baronial opposition was no more than a nuisance and a hindrance. Under Edward II it quickly became a canker which sapped the strength of the king's government, undermined public order and hopelessly enfeebled military action in Scotland and Gascony.

Those who opposed the king were not moved by any very high purpose. Self-interest strengthened by mutual suspicion and rivalry usually governed their actions. Thus great men like the earls of Arundel, Gloucester, Warwick, Lincoln, Pembroke, Hereford and Richmond could never manage to hold together for long; they only contrived to do so for four short periods in 1311, 1314, 1318 and 1321. They got rid of royal favourites and forced other concessions from a reluctant king, but their coalitions soon fell apart. Thomas of Lancaster, the most typical of them, was also the strongest. On Edward's accession Thomas was earl of Leicester, Lancaster and Derby and, in 1311, he acquired the earldoms of Lincoln and Salisbury through Alice, his wife, heiress of Henry de Lacy. To these great possessions he added royal birth (for he was a grandson of Henry III) and a particularly sullen, quarrelsome and faithless disposition. In the long run Edward II and Lancaster between them wrecked the baronial cause and compassed their own ruin, but in so doing they inflicted terrible damage on the peace, security and prestige of the kingdom.

Piers Gaveston and the Ordinances.—Trouble began soon after Edward's accession. Gaveston, who had been exiled by the old king, was recalled to receive the royal earldom of Cornwall and the new king's intimate confidence, while others, who had served Edward as prince of Wales, were given virtual control of the royal household. Early in 1308, when the king went to France for his ill-fated marriage with Isabella, daughter of Philip the Fair, Gaveston remained behind as keeper of England. Deeply affronted by these proceedings and particularly by Gaveston's tactlessness, a group of magnates forced Edward to exile his favourite for a second time (June 1308). He went as the king's lieutenant to Ireland, but he was back again the following year, after Edward, in the so-called Articles of Stamford, had promised to behave more circumspectly. Things, however, were but little improved and after July 1309, when Lancaster became chief leader of the opposition, a serious crisis was clearly impending. By Feb. 1310 he, together with the earls of Warwick, Hereford and Pembroke, had decided on drastic action and they openly accused Edward of wasting his inheritance and of ruining the kingdom. The king then had to agree to the appointment of a committee of eight earls, seven bishops and six barons, who, before Michaelmas 1312, were to prepare ordinances for reforming the government of the realm. This body was known as the lords ordainers. Weakened by yet another failure in Scotland, Edward met the ordainers at Westminster in Aug. 1311, where about 40 ordinances were presented.

The Ordinances (1311).—These were well-meaning and strictly traditional in tone. The ordainers looked back to the precedents of Henry III's time, and they had "the righteous earl," Simon de Montfort, as their model. The king must rid himself of his evil advisers and get some better ones, and the ordainers were in no doubt where these could be found. Edward must look to his "natural counselors," the baronage, and especially to the whole body of them in parliament, where policy ought to be decided and all important appointments in the royal service made. All the king's officers, including the steward of the household and the keeper of the wardrobe, should swear to observe the ordinances, while in all future parliaments a baronial committee should hear complaints against royal servants. In the 20th ordinance Gaveston was singled out for special mention. He was to be permanently exiled from all the king's dominions. The ordainers also cherished the illusion that, if only the royal revenue were properly managed, the king could live on his own without continuous financial demands upon his subjects.

The king accepted the ordinances because he had no alternative, but he seems to have had no real intention of observing them. He failed to dismiss many of those who had been proscribed, and even Gaveston was recalled after a few weeks' exile. Thereupon Lancaster and the others took up arms. Gaveston was cornered in Scarborough castle, and he surrendered on the curious understanding that if Edward's negotiations with the ordainers about him failed he should be returned to Scarborough again. His fate was brutally different. Aymer de Valence, earl of Pembroke, escorted him south to Deddington in Oxfordshire, where he was seized by Guy, earl of Warwick, and beheaded on Blacklow hill near Warwick (June 1312). Edward had to pardon this crime, although he neither forgot nor forgave. In return the strict bondage of the ordinances was somewhat relaxed. By now, however, Edward's long neglect of his Scottish interests was threatening calamity there. In Aug. 1311 Bruce ravaged the Tyne valley and by the spring of 1313, in Scotland itself, he had taken the castles of Linlithgow, Perth, Edinburgh and Roxburgh, while Stirling, the strongest of them all, was closely besieged. Edward hurried north and on June 24, 1314, he fought with Bruce at Bannockburn, only to sustain a crushing defeat and a tremendous slaughter of his followers. At one stroke Scottish independence was virtually secured and Edward was put at the mercy of the baronial opposition in England.

The Ascendancy of Thomas, Earl of Lancaster.—Thomas of Lancaster had taken no part in the Scottish campaign, but he busily exploited its failure to make himself the real master of the English kingdom. His supporters became the king's ministers and he dominated the council. It is true that he faced great difficulties.

From 1314 to 1316 harvests were exceptionally bad and cattle disease was widespread. This led to a great scarcity of food in which all classes suffered; at times the king himself went short of bread. Deaths from starvation and disease multiplied and there was no real improvement until after the harvest of 1317. While the shortage was at its worst, the Scots carried fire and sword as far south as Richmond in Yorkshire, feuds among the Welsh marcher lords threatened a general insurrection there, and north of the river Mersey Lancaster's own tenants fought among themselves. Farther afield Edward Bruce, brother of Robert I, had landed near Carrickfergus, Ulster, in May 1315, and soon reduced northern Ireland to anarchy. In face of all this Lancaster was supine and incompetent. He never tried to use the power he had, and he allowed it to be taken from him without a struggle by some of his former associates, headed by the earl of Pembroke, in the treaty of Leake of Aug. 1318. During the next two years this moderate group genuinely tried to mend matters. Wheat came down to sixpence a bushel. It had been more than three shillings. In Dec. 1319 a not unsatisfactory truce was made with the Scots, an accumulation of difficulties with the French king over Gascony was tidied up and a thorough overhaul of the administration, begun by the Household Ordinance of York (1318), was carried through. In part this work was continued after those who began it were displaced, for the last years of Edward II's reign were a time of administrative improvement, particularly at the exchequer. There two able ecclesiastics, Walter of Stapledon, bishop of Exeter, and William de Melton, archbishop of York, made considerable progress toward clearing off the mass of debt left by Edward I and gave the office its final medieval shape.

Political differences, however, were not to be permanently laid aside. By 1321 Pembroke and the moderates were deeply divided among themselves, while the king had come under the influence of two able but ambitious men, Hugh le Despenser and his son and namesake. They belonged to a respectable if pushing baronial family, and the younger Despenser, who had long experience as a household official, was also husband of one of the Clare heiresses, through whom he secured the great marcher lordship of Glamorgan. His ambitions there brought a renewal of party strife for, with the king's connivance, he contrived to stir up the resentment of the Welsh marcher lords and the infection quickly spread.

In July 1321 Lancaster and his friends appeared at a parliament in force and, with the now familiar menaces, compelled Edward to banish the Despensers. This time the king was not so easily intimidated. In October of the same year, when Bartholomew de Baldesmere, a member of the league against the Despensers, insulted the queen, Edward attacked his castle of Leeds in Kent, hanged the garrison and drove off the barons who had come to Baldesmere's aid. Thereafter the Despensers returned, while Lancaster and some of his associates made ready to fight the king. They displayed great incompetence, were roughly shepherded northward and thoroughly beaten at Boroughbridge (March 1322), after which Lancaster surrendered. He was immediately beheaded on Edward's order. Baldesmere and a number of others were subsequently executed.

The Statute of York (1322).—The king's triumph was followed by a parliament at York (May 1322) in which the famous Statute of York, drafted by the council, was promulgated. This declared the ordinances of 1311 void, but asserted that all matters touching the king and the realm should be "treated, accorded and established in parliament by the king and by the consent of the prelates, earls, barons and of the community of the realm." Many scholars have seen in this an uncompromising declaration of parliamentary sovereignty, but this is almost certainly an exaggeration. The troubles of Edward II's reign had often hindered the working of the king's government, but they had never undermined its strength. Power was still firmly in the hands of that inner ring of officials who staffed the council and the household. Nevertheless these bitter quarrels did set a lasting, if unintended mark on the slowly developing institution of parliament. Politics obtruded themselves into the forefront of its business and henceforth they stayed there. The baronage firmly established their privileged influence as the lords of parliament, an influence which the king might resent but which in future he could not disregard. In somewhat ambiguous language the Statute of York recognized these facts. Nor is it an accident that soon after 1330 parliament achieved its characteristic medieval shape; the ground had been well prepared under Edward II.

The Rebellion of Isabella and Mortimer.—Although Edward was now rid of his private enemies, the last five years of his reign were marked by fresh political disasters abroad. In 1323 the Scottish war blazed out again to be stifled by a truce, which could not become a peace because Edward stubbornly refused to recognize the royal title Bruce bore. But it was with his French relatives that the worst trouble developed and it was largely of Edward's own making. Charles IV, who succeeded Philip V in Jan. 1322, had waited patiently and in vain to receive Edward's homage for his French lands. Also, when Charles ordered the building of a bastide on disputed land at St. Sardos in the Agenais, the English seneschal in Gascony attacked the town and killed a number of Charles's officers. Not unnaturally all this led to the formal confiscation of Gascony and Ponthieu in July 1324, and their invasion by the French followed quickly.

Edward was now foolish enough to send Isabella, with whom he had quarreled, to intercede with her brother Charles IV, and he even allowed Edward, prince of Wales, to follow her to Paris. There she plotted with Lancaster's exiled friends and returned to England in Sept. 1326 at the head of an army commanded by Roger Mortimer of Wigmore, whose mistress she had become. The invaders got substantial baronial help, while the king and his few remaining supporters were soon rounded up. Both Despensers were executed and Edward was deposed in favour of his son. To give it a larger appearance of authority this latter proceeding was carried through in a parliament at Westminster in Jan. 1327. In the following September Edward II died, most probably by violence, at Berkeley castle. He was buried by the monks of Gloucester near their high altar and the offerings from his shrine paid for the rebuilding of their magnificent choir.

Edward III (1327–77)

The Regency of Isabella.—During the next four years Isabella ruled as regent for her son, but the real power rested with Roger Mortimer. For a time he depended on the support of the magnate coalition that had overthrown Edward II, in which Edmund, earl of Kent (Edward II's half brother), and Henry, brother of the dead Thomas, earl of Lancaster, were especially influential. But this group held uneasily together and began to break up when Mortimer shamelessly used his position for his private advantage. He became justice of Wales for life, greatly enlarged his territorial interests in the Welsh marches and added lands in Ireland and the Midlands to his vast patrimony. In Oct. 1328 he became earl of the march of Wales. Now his allies turned against him, but Kent was executed and Lancaster lost half his estates. It was only late in 1330, when with papal encouragement the king took a hand, that Mortimer himself was overthrown. He was condemned for treason and hanged at Tyburn. Isabella lived in retirement until she died in 1358.

The regency made little difference to the routine of government and the general condition of the kingdom, but Mortimer did put an end to the Scottish war. The treaty of Northampton (1328) made the necessary but unpopular renunciation of English suzerainty over Scotland and recognized Bruce as king. No such drastic solution of the Gascon problem was possible. In March 1327 Mortimer made what was optimistically called a "final peace" with Charles IV, by which most of the duchy was to be returned when the English had paid a heavy indemnity. But Charles IV died barely a year later, the French succession was in doubt and Edward III's claim through his mother, the last surviving child of Philip IV, was rather halfheartedly pressed. It is doubtful whether it was meant as anything more than a maneuver to help matters in Gascony. Certainly when Philip of Valois established himself as Philip VI, Edward III proffered rather ambiguous homage for his French lands (Jan. 1329), and after further severe pressure made complete submission in April 1331.

Social and Economic Background of the Reign.—To be

fully understood, the political and military happenings of Edward III's reign need to be viewed against the larger background of those social and economic changes which were then slowly beginning to transform the face of medieval society both in England and on the continent.

The Hundred Years' War itself, begun in Edward's reign, was not without influence on the course of these changes. For the English it was a long and tantalizing war, in which there were many great victories, but no final triumph. In the end it broke the strength of the medieval English monarchy, dangerously loosened the ties of public order and shook the whole fabric of society to its foundations. Nevertheless it seems that the main explanation for the slow decay of medieval society, for that is what was happening, must be sought elsewhere than in the doings of kings and the achievements of armies.

The signs of this decay were increasingly manifest on every side in England during the 14th and 15th centuries. They point unmistakably to a general loss of material prosperity and of moral confidence. In the economic sphere they are particularly obvious, especially between 1350 and about 1470. The vigorous economic progress of the 13th century petered out early in the 14th century. It was quickly followed by a long period of decline, which gathered momentum as it entered the 15th century, to reach its lowest level and its end under the Yorkists. Wheat prices sagged; land went out of cultivation; in some places, in Norfolk and Leicestershire for instance, whole villages were abandoned. On many manors tenants became increasingly hard to find on the traditional terms, and the whole apparatus of villein services was slowly but generally laid aside.

Even the wool trade, that staple of English commerce, suffered, so that in the 15th century its total volume was less than it had been 100 years before. It is true that the native cloth industry prospered in the west country and in East Anglia during the 14th century, but this was at the expense of the Flemish towns; it was not a genuinely new economic growth. With the notable exception of London, most of the English towns felt the pinch, especially in the 15th century; trade declined and population shrank.

Signs of strain and confusion showed themselves in other departments of life as well. Widespread heresy appeared in England through the Lollard movement in the closing years of the 14th century (*see* under *Henry V*, below), while the poetry of the orthodox William Langland (*q.v.*), who wrote about the same time, has been described as "the exceeding bitter cry of the socially disinherited." It was also a fervent affirmation of the ancient faith in an age which was growing ever more secular in outlook. Then there were the movements of peasant discontent, of which the revolt of 1381 is the best known. These were matched in the towns by the struggles of the journeyman guilds. The established social and economic order was breaking up, and the process was painful for the great majority.

The main causes of this decay, which was apparent throughout Christendom, have not yet been established with certainty. One possible explanation is that soon after 1300 medieval society reached the extreme limit of its capacity for economic development. Its resources were strained to the utmost and after that they began to contract. A less speculative cause is to be found in the effects of the great pandemic of plague, which began with the Black Death (*q.v.*) of 1348 and smoldered in England with slowly decreasing intensity for the next three centuries until it finally became extinct about 1700. The mortality in 1349 was tremendous—at least one-third of the population died in a few months—but, by itself, that was not a fatal blow. The real damage was done by the steady recurrence of the disease in national epidemics and local outbreaks until, toward the end of the 15th century, its virulence was somewhat abated.

Plague was certainly a main cause of the marked fall in population, which is registered by the general decline of economic activity, and of that loss of moral confidence which plainly showed itself in social and religious discontent. In any case it is into this background that the doings of Edward III and his successors, together with the Hundred Years' War itself, must be fitted.

Character and Ability of Edward III.—After Mortimer was overthrown Edward III began the long period of his personal rule, which lasted until his death. The new king had authority thrust upon him when he was almost too young to bear it (he was 18 at the time of Mortimer's death), and he lived long enough to be overtaken by a creeping senility which, in his last years, deprived him both of the will and the ability to govern.

From the first he exhibited a curious mixture of qualities. He was, and always remained, an impulsive, passionate man, who lived for the present and took little thought for the morrow. Soldier and huntsman by inclination, he also loved the pomp and pageantry of power. In his prime he was a great and splendid figure who dazzled his contemporaries as the living embodiment of those knightly virtues then generally admired. He was also endowed with great personal charm; an easygoing and companionable man.

But Edward was in no sense a statesman; he was not even a competent political craftsman. An improviser, he was always ready to purchase present convenience at the price of future difficulties. He was not interested in the steady exercise of power or in the promotion of orderly government. All Edward's enterprises, though begun with zest and prosecuted with vigour, had an aimless quality about them. They were the work of a man who never quite knew where he wanted to go and was content that this should be so. The politics of his reign were largely shaped by the Hundred Years' War (*q.v.*), which he began in 1337 and bequeathed as a miserable, debilitating legacy to his successors, until it spent itself and the English were totally defeated in the middle years of the 15th century.

When he assumed the management of affairs in 1330 Edward gathered up the threads of the three great, and now familiar, political problems—Gascony, Scotland and the discontents of the English magnates—which his father and grandfather had manipulated with varying success during the previous 40 years. Like them he had to work on this material, but he soon contrived to fashion it into a new and distinctive pattern. Baronial discontents quickly ceased to be of first importance. For a few years Scottish affairs claimed him, but then they were merged into the larger undertaking of the French war.

The Problems of Scotland and Gascony.—Robert Bruce died in 1329, and Scottish government passed to his young and inexperienced son, David II. In 1332 a pretender appeared in Edward, son of John de Balliol, who collected the "disinherited," the members of the old English faction expelled by Robert Bruce, and invaded the realm. At the battle of Dupplin near Perth he beat the regent, the earl of Mar, and was soon crowned at Scone. But knowing his hold was precarious, Balliol offered himself as Edward's vassal. At first the English king had striven to appear as a disinterested, although shocked, spectator of events in Scotland, but the temptation was too great. He accepted the proffered homage, went to Balliol's assistance and by his victory at Halidon hill (July 1333) established his vassal securely, or so it seemed. English garrisons reappeared in Perth and Edinburgh, while English troops penetrated as far north as Inverness. Edward established his military reputation, collected a valuable group of trained soldiers and practised the tactics (a combination of archers and dismounted men-at-arms) which were to win the battles of Crécy and Poitiers.

But there was a less satisfactory side to the matter. Like his grandfather, on whose experience he relied heavily, Edward III overreached himself in Scotland. He compelled Balliol to cede all his rights in the counties from Linlithgow to the Solway (June 1334), an action which aroused much resentment in Scotland and led to the collapse of effective English control there. Meanwhile, in accordance with a now traditional policy, Philip VI of France had espoused the cause of David II and, as had happened before, this stirred up the smoldering discontents over Gascony between the French and English kings. In 1334 Philip put an end abruptly to some most promising negotiations with Edward, designed to settle outstanding differences in the duchy, because David II and the Scots were not to be included in the agreement. From then on the crisis widened and deepened as others, such as Robert of Artois, a cousin whom Philip VI had expelled from France, took

a hand in the business. By the autumn of 1336 Edward III and his advisers realized the full extent of the mischief; they saw that a fresh confiscation of Gascony was impending and that papal diplomacy was powerless to heal the breach.

Yet there is no reason to suppose that those concerned regarded this as anything more than another of those recurring periods of acute tension that had marked Anglo-French relations for the past 70 years. To their eyes it bore a close resemblance to the clash between Edward I and Philip IV at the end of the 13th century, and in handling their crisis Edward III and Philip VI looked back to their predecessors' experience. They proceeded by well-tried methods. Gascony was confiscated in May 1337. Edward III cast round for allies among the princes of the Low Countries and the Rhineland, began his political manipulation of the English export trade in wool and reasserted his claim to the French throne, to quiet the consciences of some of his allies as well as to cause embarrassment to his opponents. Finally, at the beginning of Nov. 1337, the bishop of Lincoln arrived in Paris to present his master's defiance to Philip VI, and the conflict was formally begun.

The Hundred Years' War From 1337 to 1360.—Those who began the war neither desired nor foresaw that it would spread so widely or last so long. The extraordinary military incapacity of Philip VI and his successor John II encouraged the English to commit themselves so deeply that they could not draw back. Edward III, for his part, never seems to have realized that, although he might ravage the French kingdom, he could not break the strength of the French monarchy. He made the same mistake that had already ruined the English cause in Scotland.

The war started slowly. In 1339 and again in 1340 Edward III with his continental allies tried unsuccessfully to break into northeastern France. Philip VI refused battle and they accomplished nothing. At sea the English were more fortunate. In the engagement off Sluys (June 1340), the French fleet was destroyed, and Edward III had command of the English channel for the next 20 years. This did not prevent him from being driven home in the same year by shortage of money. In England he faced a ministerial crisis, provoked in part by his dissatisfaction with the efforts of his subordinates to meet his financial needs. For the next five years the struggle languished. Edward drifted along the path of financial ruin until he actually went bankrupt in 1345, when he repudiated his debts and wrecked several of the great Italian banking houses. The Flemings were also hard hit by this collapse of the king's credit. Their enthusiasm for the English cause waned noticeably, and Jacob van Artevelde, its mainstay, was murdered by his own townsmen in 1345.

Meanwhile the English were busy trying to force another side door into the French kingdom through Brittany, where Edward supported John de Montfort, a claimant to the ducal title. This enterprise, begun in 1342, was hindered by fresh papal efforts at mediation, of which the most notable were the conferences held by Clement VI at Avignon (Oct.–Dec. 1344). These failed but they are important in that they lay bare what were now the aims of each side. Edward asked for Gascony in full sovereignty. Philip would do no more than offer to restore it on the old terms.

The Crécy Campaign.—In 1346 Edward III changed his military policy. He ceased to rely on continental allies and organized a frontal assault on the French kingdom with his own army. He landed in the Cotentin peninsula, ravaged Normandy and penetrated up the Seine valley almost to Paris. This brought the French king hurrying against him with an army four times as large as Edward's own. The English retreated northeast, toward Flanders, but after crossing the Somme they halted at Crécy, near Abbeville, and fought with their pursuers. Using the tactics already perfected in Scotland, Edward cut the French host to pieces (Aug. 1346). He then moved to Calais, which surrendered after a long and bitter siege (Aug. 1347). Prostrated by his defeat, Philip VI made no attempt to interfere and the English took the port which was to serve them as a secure and convenient bastion in northern France for the next 200 years. Elsewhere other victories were added to this great English success. By his triumph at Auberoche in Périgord (Oct. 1345), Henry of Lancaster restored Edward's suzerainty over lands lost to the Gascon duchy since the beginning of the century, while at Neville's Cross, near Durham (Oct. 1346), the border lords defeated and captured David, king of Scots. This greatly weakened the help the Scots were able to give their French allies who, after the fall of Calais, hurried to make a truce with their English foes.

For the next nine years there was much intrigue but little serious fighting. Successive truces were not always scrupulously observed, but other considerations contrived to hold the combatants apart. There was the Black Death in England and France, which slowed down the pace of life and caused great, if temporary, disorganization in both kingdoms. Then Philip VI died in 1350, to be succeeded by his mediocre son, John. The English, greedily anxious to profit from his weakness, used every kind of pressure to harass and annoy him. Finally negotiations for a definitive peace were begun in the spring of 1353 at Guines under papal auspices. They dragged on until April 1354, when John's representatives accepted proposals which gave to Edward III not only Gascony but Poitou, Touraine, Anjou, Maine and Normandy as well, all in full sovereignty. It was a harsh settlement which probably could not have been enforced, but in any case it was never ratified. John and his ministers, infuriated at flagrant English violations of the truce in Brittany, went back on their word, and in the autumn of 1355 fighting started again in Artois and Aquitaine.

In the next two years there followed a series of great plunging English raids into the heart of the French kingdom, launched from Calais and Bordeaux. Those in the south were commanded by the king's son Edward, the Black Prince. In 1356 he was returning after a great circular sweep through Périgord, Limousin and Berry, when he was intercepted by King John near Poitiers.

Poitiers and the Treaty of Calais.—At the great battle of Poitiers (Sept. 1356) the English triumphed against even larger numerical odds than at Crécy, and took John himself prisoner. He, bemused by visions of a crusade, was willing to buy his freedom, and he made two successive agreements with Edward III which gave such advantageous terms to the English that the dauphin Charles, as regent, refused to accept them. Accordingly in 1359 Edward in person led an army which devastated all northern France as far as Reims and the borders of Burgundy. Both parties then became anxious to make peace and at Brétigny, near Chartres (May 1360), they agreed to proposals which were ratified in the treaty of Calais (Oct. 1360).

Although not so great as he had hoped, Edward's territorial gains were still immense. Gascony was enlarged into a great sovereign Aquitaine, stretching from the Loire to the Pyrenees and reaching eastward to the central plateau. It was buttressed by channel bridgeheads in the counties of Ponthieu and Guines, together with Calais. Edward renounced his claim to the French throne and John undertook to pay 3,000,000 crowns, in seven installments, for his freedom. (In 1357 David of Scotland had bought his liberty on equally onerous terms. He ceded Roxburgh and Berwick, also promising to find a ransom of 100,000 marks.)

This spectacular mutilation of the French kingdom was not destined to last, for it was the outcome of highly exceptional circumstances. It had not come about because of Edward's military genius; his strategic thinking began and ended with large-scale devastation. It happened because the great material strength of the French kingdom, beside which that of England was puny, had not yet been brought into play. When, in Charles V (1364–80), the French acquired a competent leader, the English conquests melted away. Moderation in 1360 could have ended the struggle on terms notably advantageous to the English. Edward was blind to this, for he took all he could get regardless of whether it could be held.

Domestic Affairs.—At the beginning of the French war it seemed that Edward might adopt the strict form of household government favoured by his predecessors. Indeed the ordinances made at Walton, Suffolk (July 1338), sought to re-establish such a system, but the ministerial crisis of 1340–41 put an end to the experiment. Henceforth the king turned mainly for advice to his hereditary counselors, the lords of parliament; while executive responsibility remained with the clerks of chancery, the exchequer

and the other offices, who were slowly being shaped into a rudimentary civil service.

A parliament met for a few weeks in most years, generally in London, now becoming the administrative capital. The clergy, except the bishops, abbots and priors, ceased to attend, although they were still summoned. After 1340, in face of continuous royal demands for money, the knights and burgesses gradually drew together as a homogeneous body of commons in parliament, although as yet they made no attempt to assert their corporate power. Elsewhere the office of justice of the peace was slowly defined and those who held it shouldered an increasing burden of local responsibility on the king's behalf.

The Statute of Labourers (1351), which sought to fix wages and prices after the plague, together with that multitude of enactments which led to the establishment of the staple system as a means of regulating the export trade in wool, show the increasing control of the central government in economic matters. The Statute of Labourers fixed rates for all wages practically identical with those prevailing before the Black Death. Workmen who refused to accept them were to be imprisoned, while employers who secretly paid higher sums were to be punished by heavy fines. Later stripes and branding were added to the labourer's punishment if he remained recalcitrant or absconded and landowners were empowered to seize all vagrant able-bodied men and compel them to work at the statutory wages. At the same time parliament tried to bring down the price of commodities to their former level. The scheme proved impossible to carry out. If driven too hard, labourers absconded to the towns, where labour was needed as much as in the country, or migrated to districts where the statute was laxly enforced. In consequence many more lords commuted for money rents the agricultural services customarily due from their tenants, and let out their demesne lands as separate holdings. (*See* also *Economic and Social Development in the Later Middle Ages* below.)

The Statute of Provisors (1351), designed to restrict papal use of English ecclesiastical patronage, and the Statutes of Praemunire (1353 and 1365), which bridled appeals to Rome, were a direct consequence of the papacy's entanglement in the politics of the French war. (*See* PRAEMUNIRE; PROVISION.)

Edward's Last Years.—After 1364 Charles V, with the help of men like Bertrand du Guesclin, slowly contrived to raise the French monarchy from its abasement. Meanwhile Edward, the Black Prince (*q.v.*), as ruler of Aquitaine, had involved himself (1367) in a successful but expensive campaign in Castile to support its ruler, Pedro I, against his francophile rival, Henry of Trastamara. This enterprise was paid for by Aquitanian taxes, which were much resented. The count of Armagnac and others appealed to Charles V, who was ready to interfere, and the war was started afresh. But Edward III could no longer take his share. His queen, Philippa, died in 1369, and he entered his dotage to fall under the malign influence of his mistress Alice Perrers and her accomplices. The Black Prince was in no better case. In 1371 he returned home a chronic invalid, to predecease his father by a few months.

Power passed to John of Gaunt, duke of Lancaster (*q.v.*), Edward's next surviving son, and for 20 years he was to have a predominating influence in the management of affairs. By his first wife, Blanche, he had acquired the vast Lancastrian inheritance; his manors, it was said, covered one-third of England. His second wife, Constance, gave him a claim to the Castilian throne. Gaunt was loyal, although somewhat grasping, but he was not particularly able. Throughout his career he had to shoulder responsibilities heavier than he was capable of bearing. He was a clumsy politician, and in war he adhered to established methods long after their usefulness was spent.

This was painfully evident in his handling of the renewed struggle with the French. The great English raids on the French kingdom were begun again, but, although they still inflicted grievous damage, the enemy had taken their measure. Now they avoided pitched battles and steadily pulled the Aquitanian principality to pieces while the English armies were marauding elsewhere. Gaunt himself commanded the largest and the most futile of these raiding enterprises in 1373. He marched from Calais to Bordeaux through Picardy, Champagne and Berry, leaving half his army dead by the way. When a truce was made at Bruges in 1375, the English dominion in France had shrunk to Calais and a strip of the Gascon coast from Bordeaux to Bayonne.

At home, these years when the old king's grasp had loosened saw the reappearance of aristocratic factions like those that had dominated the politics of Edward II's time. Gaunt, whose wide connections were a party in themselves, headed the strongest group; another formed itself round the Black Prince's widow, the princess of Wales and their son, the future Richard II; while a third included powerful bishops like William of Wykeham and William Courtenay. Their chosen battleground was now in parliament and a foretaste of what was to happen under Richard II is provided by the doings of that best reported of all 14th-century parliaments, the so-called Good parliament, which met in the early summer of 1376. Gaunt's friends had governed with indifferent success and much corruption since 1371, and their opponents were determined to remove them.

The commons were stirred into unprecedented activity under their speaker, Peter de la Mare, and with the dissident lords they forged the handy weapon of impeachment to strike down the ministers. Then, like the ordainers of 1311, they tried to put the king in tutelage to a council nominated in parliament. Their success was short lived, for Gaunt soon reasserted himself and was firmly in control again when Edward died on June 21, 1377.

RICHARD II (1377–99)

The accession of Richard II was of no more than formal significance for it changed nothing. The new king was only ten and, although he was to reign for 22 years, he took an effective part in government only after 1389. Before that he was subjected to various forms of control and coercion, necessary during his legal minority but continued after that ended because the habits of intrigue and lack of discipline among those in high places, begun during his grandfather's last years, were now firmly rooted.

Richard's character has caused much speculation. He has been described as the victim of a progressive mental disease which finally overwhelmed him in the last few years of his life. The facts warrant a simpler explanation. Richard was something of a fanatic about his regality, for he had imbibed the high doctrine of monarchy taught by the Roman lawyers. Otherwise he was a shrinking, fearful creature, gifted with no great intellectual ability nor even with much power of application. Up to a point he could learn, for he learned to be cruel and ruthless, but he never learned when to stop. He never had any political common sense. To crown it all he had to deal with a situation which, from the beginning, was too much for him. It would have taxed the full powers of a much abler man. In quiet times Richard might have managed, but not in his own age. The dominant theme of the history of Richard's reign is this renewed outbreak of political disorder in high places; the same malady which had destroyed Edward II and was now to compass the ruin of his great-grandson. All else was subordinate, even the French war and the great crisis of 1381.

The Hundred Years' War From 1377 to 1396.—The French war began again in 1377 after the desultory peace negotiations that followed the truce of Bruges had failed. The English continued the strategy of devastation, to which they still clung, and, in 1380, Thomas of Woodstock, Edward III's youngest son, captained the last great raid from Calais. In the same year Charles V of France died. From then on both sides drifted into the unenviable position of being unable to wage effective war and unwilling to make an enduring peace. The English became more feeble and intransigent, and for the French the death of Charles V was a disaster.

The youthful Charles VI was ill-fitted to carry forward his father's work of repairing the realm and of chastising the English. Moreover, both kingdoms were now caught in the toils of the Great Schism of the West, which began in 1378, and their differences were further acerbated by the remorseless feud between Rome and Avignon. In the summer of 1383 the English, who were adherents of the Roman pope, Urban VI, staged what was called an Urbanist crusade in Flanders, an enterprise which was grossly

mismanaged by Henry le Despenser, bishop of Norwich. Meanwhile the French and the Scots, who supported Clement VII, the pope at Avignon, made a new alliance, intended to promote a simultaneous invasion of England. This enterprise was twice attempted, in 1385 and again the following year. On both occasions it failed despite the fact that a French expeditionary force operated in Scotland and a great invasion fleet was collected at Sluys. Nevertheless the business caused great alarm in England and demonstrated that Richard's government did not lag behind that of his adversary of France in its capacity for mismanagement.

After this, fighting gradually died down and from Aug. 1388 long truces became the rule. Negotiations for a permanent settlement began in 1391 and dragged on until March 1396. Then Richard, whose first wife, Anne of Bohemia, had died in 1394, was betrothed to Isabella, daughter of Charles VI, and the truce was prolonged for 19 years. Despite Richard's obstinate desire to end the struggle, no territorial settlement was made, for even after nearly five years of haggling the parties could not agree on details. The conflict was simply suspended for the time being by mutual consent.

The Peasants' Revolt.—Meanwhile, at home the simmering discontent in town and countryside, produced by social and economic changes which have been described, had momentarily boiled up. The peasants' revolt (*q.v.*) of 1381 was a movement of desperation, spectacular but with little or no permanent result. Although some of the poorer clergy were involved, it does not seem that the religious radicalism of John Wycliffe and the Lollards (*qq.v.*) had any influence on the movement. The immediate cause of the outbreak was the government's attempt to uncover numerous evasions of the flatrate poll tax of one shilling a head, authorized for war expenses by a parliament at Northampton in Nov. 1380. It was in the areas chosen for particular scrutiny that most of the trouble occurred—in the shires from the Wash to the channel and in isolated places as far away as the Wirral peninsula and Yorkshire. It was a movement of town and countryside; rich and poor peasants, town workers and disgruntled parsons banded together to hit out at lawyers, manorial lords, wealthy monasteries, oppressive merchant oligarchies and politicians like Gaunt and Simon of Sudbury, the archbishop of Canterbury, whom they detested. The local risings, which seem to have been spontaneous and not planned, happened at intervals throughout May and June 1381. There was no national leadership, nor was there any uniform program.

London was the scene of the most serious trouble. The men of Kent and Essex, led by Wat Tyler and John Ball (*qq.v.*), entered the city with help from sympathizers within and occupied it for four days, looting, burning and killing. After Tyler was killed at Smithfield and Richard accepted their demands, which included the abolition of villeinage and the partial disendowment of the church, the rebels dispersed. Outside London, for instance in Norfolk where the bishop put down the revolt with an iron hand, the disturbances were more easily suppressed. Many ringleaders were severely punished by the ordinary process of justice, but there was none of the indiscriminate barbarity which marked the suppression of similar contemporary movements on the continent.

BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

PEASANTS' REVOLT OF 1381: (LEFT) DEATH OF WAT TYLER; (RIGHT) KING RICHARD II ADDRESSING THE PEASANTS: FROM A MANUSCRIPT OF JEAN FROISSART, CHRONICLE; 15TH CENTURY. IN THE BRITISH MUSEUM

The King's Enemies.—During the king's minority the government was carried on by a council drawn mainly from the cliques which had jostled for power in the years just before Edward III died, and among them the Lancastrian interest still predominated. The king did not begin to play an active part until after his mother's death in 1385. Then he too built up his personal following, which included his father's old companion in arms, Simon Burley, another able knight, Baldwin Raddington, his young friend, Robert de Vere, earl of Oxford, and Nicholas Brembre, grocer and mayor of London. There were others also busy establishing themselves at this time, prominent among them Thomas of Woodstock and Edmund of Langley, uncles to the king, the earls of Warwick and Arundel, the young Thomas Mowbray, earl of Nottingham, and Gaunt's heir, Henry Bolingbroke, who became Henry IV.

However, the ambitions of all these, the king included, were held in check by Gaunt's influence until the summer of 1386. Then, despite the pressing danger of a French invasion, he departed to Spain, where he remained for three years, leaving Richard at the mercy of men who were far less scrupulous and far more self-seeking even than Gaunt himself. He went to Spain to press his claim to the Castilian throne, a scheme which had been his chief concern for some time past, but for which until that time he had been unable to obtain sufficient financial support.

Gaunt's departure temporarily suspended his political influence and Richard, with the help of his friends, tried to manage for himself. Others were equally anxious to profit from the occasion, notably Thomas of Woodstock, with the earls of Arundel and Warwick. They quickly gathered a considerable party, which included Thomas Mowbray and Henry Bolingbroke, and also numbered in its ranks that rather sinister trio, Sir John Bushy, Sir William Bagot and Sir Henry Green, who belonged to the influential group of country gentry which provided the sheriffs, the justices of the peace and the members of the commons in parliament from the shires.

It is not quite clear how far this group was determined to go. Some of its leaders, like Woodstock and Arundel, wanted to get rid of Richard altogether. They openly threatened him with the fate of Edward II. As a body, however, they desired to make the king their tool, and in a parliament held in 1386, the so-called Wonderful parliament, they contrived to put Richard, his council and his household under the close supervision of a committee appointed by themselves.

Richard reacted characteristically and unwisely. In Aug. 1387 he asked the judges whether such acts were treasonable, and they said yes. This stampeded the opposition into precipitate action. In their turn they "appealed" Richard's principal associates of treason, hence the name appellants, by which the opposition group was henceforth known, and under threat of force compelled the king to summon a parliament to Westminster in Feb. 1388.

In that assembly, called the Merciless parliament, the appellants used every trick known to them, every abuse of due process of law, to carry through a systematic judicial massacre of Richard's closest friends and supporters. It took from February until May to complete the business. Then, just before the parliament broke up in June, the instigators of all this violence arranged to reward themselves handsomely at the public cost for what one of the chroniclers calls "their great expenses in procuring the salvation of the realm and the destruction of the traitors."

A few months later Gaunt returned and, on the surface at least, things settled down again. Largely as a result of Gaunt's steadying influence, the king now assumed full authority without much outside restriction. Richard was badly shaken by the events of 1388 and for a time he was thoroughly cowed, but he neither forgot nor forgave the destruction of his friends. He also quickly learned the lessons in political gangsterism which Woodstock,

Arundel and Bolingbroke had obligingly taught him. For nearly seven years he bided his time, gathering new friends, assembling an armed retinue and winning over many of those, including Bolingbroke, Mowbray, Bushy, Bagot and Green, who had formerly supported the chief malcontents, Woodstock, Arundel and Warwick.

By the summer of 1397 they were plotting again, for they may well have been alarmed by the multitude of Richard's supporters. Mowbray knew sufficient to betray them to the king, by whose order they were seized. Woodstock was hurried off to Calais, where he was murdered, if not at Richard's command certainly with his knowledge. Warwick and Arundel were brought before a parliament at Westminster in Sept. 1397 where, with Richard's Cheshire bowmen standing by, they were condemned through Gaunt's mouth by a procedure almost exactly similar to the one they had used against Richard's friends nearly ten years before. This was done at the king's behest and it casts a curious light on his character, as does the promotion of Bolingbroke and Mowbray, who got dukedoms to mark the occasion.

After this none of those prominently concerned in the activities of the Merciless parliament felt safe, despite the pardons and favours they had received. Richard's own attitude contributed to this situation, and it is certain that the famous quarrel between Mowbray and Bolingbroke in 1398 was one outcome of the general feeling of insecurity among highly placed folk. The readiness with which Richard exiled both of them was ominous.

Deposition of Richard II.—Richard's dethronement was largely his own fault. Gaunt died in Feb. 1399 and the king thereupon confiscated Bolingbroke's inheritance. For this there was some justification, but many reasoned that if the vast Lancastrian properties were not safe then nothing was safe. Thus in July, when Bolingbroke returned, ostensibly to claim his possessions, he found a host of supporters he would otherwise have lacked. Richard further smoothed his cousin's path by undertaking an expedition to Ireland in the spring of 1399. English power there had fallen into ruin, but that was as nothing compared with the damage done to the king's authority at home by his absence during three crucial months. When he returned at the end of July his cause was lost and although he might have escaped abroad he feebly surrendered to Bolingbroke at Conway.

It seems likely that Bolingbroke came back determined to take the crown, for he had at his elbow Thomas Arundel, brother of the executed earl, who had himself been deprived of the archbishopric of Canterbury by Richard in 1397. Arundel was the most able and the most bitter of Richard's enemies.

It took nearly two months to effect Richard's deposition. Men could only with difficulty bring themselves to regard a king as a mere man. They found it equally hard to abandon the notion of hereditary succession according to the strict rule of primogeniture. Much time was wasted by ingenious if unsuccessful efforts on Bolingbroke's part to prove his claim by hereditary right. Eventually he had to act without this support. At the end of Sept. 1399 Richard, under pressure, was persuaded to abdicate and to nominate Bolingbroke as his successor. Even this was not thought sufficient. The king was deposed in parliament, and without awaiting formal authorization Bolingbroke stood forth and took the throne as Henry IV. He thus acquired the dubious title that was to prove the undoing of his house.

Richard II came to an obscure end in Pontefract castle early in 1400. He was probably starved to death. All that is known with certainty is a minute of the council, dated Feb. 8, 1400, which directed that Richard if alive should be placed in safe custody, if dead should be shown to the people. It has been rightly characterized as a murderous suggestion.

VII. LANCASTER AND YORK (1399–1485)

At first sight the accession of Henry IV, opening a new century and bringing a new dynasty, seems to be a landmark in English history. Yet on closer examination this is not so. There was no break in the strong continuity of political life, nor did the coming of the Lancastrians have any obvious effect upon that process of social and economic change through which English society was slowly passing in the later middle ages. It is true that there is much to be learned about the history of 15th-century England before the advent of the Tudors, but some outstanding facts have been established with reasonable certainty and call for particular emphasis here.

The Lack of Effective Government.—The most obvious feature of this age was its political failure—what contemporaries called lack of governance. By this they meant that public order was not properly maintained, that justice was not impartially administered and that the powerful went unchecked when they sought private advantage by force and fraud. The root of the matter was simply that the strong royal administration, so laboriously built up from the 12th century onward, was faltering in operation and under Henry VI was visibly tottering to collapse. It was the enfeeblement of the monarchy, not the exhaustion of the kingdom, that put an end to the French war in 1453, and ushered in that long period of civil disturbance, known as the Wars of the Roses, which lasted until 1485 (*see* ROSES, WARS OF THE).

This state of affairs was already developing in the time of Richard II, and the same causes which stimulated it then worked even more powerfully to foster its larger growth in the 15th century. These causes are to be found in those political troubles that have already been described: that is, in the strain of the French war, in the chronic poverty of the crown, in the manifest incapacity of monarchs like Richard II and in the restlessness of the great families. All these things thrust themselves forward into the new century and soon came to dominate English political life completely.

The French war, broken off in 1396, was renewed in 1415. Henry IV's son and successor, Henry V (1413–22), probably outshone Edward III in military skill, but he was just as lacking in political acumen. The English were again quickly and deeply committed to the vast enterprise of dismembering the French kingdom and the familiar pattern of events was repeated. Great successes were followed by spectacular reverses on the continent. At home there was no lack of militant enthusiasm, but there was an increasing reluctance to pay for the war, while the crown, always poverty stricken, was eventually reduced to virtual bankruptcy in its efforts to sustain the fighting across the channel. In England itself this seriously affected the strength and efficiency of the king's government, which was already weakened by the turbulence of the great families.

The bitter strife of the various aristocratic cliques was not stilled by the overthrow of Richard II. Henry IV only maintained himself with difficulty, and even his son was never quite secure. Under his grandson Henry VI (1421–71), first a minor and then an imbecile, the great factions accumulated unprecedented strength. This was only dissipated by the Wars of the Roses. In this conflict the monarchy was abased, the Lancastrians were overturned and their Yorkist successors suffered a like fate, while many of the great families were hopelessly enfeebled by death and confiscation. The personal deficiencies of Henry VI, while they did not cause the trouble, were not without influence on the course of events. At a time when the authority of the monarchy was brought low, it was no light matter that the monarch himself was incapable.

Nor were contemporaries unaware of what needed to be done to set matters right. Sir John Fortescue, sometime chief justice of the king's bench, writing about 1470, probed to the root of the matter. The crown was weak because it was poor. Its revenues had to be increased, while the wealth of its great subjects was limited by every possible device. Then a king who could at last pay his way and buy the services he needed must see that all the threads of his authority were gathered into the hands of a professional council in which he could repose implicit trust.

It has often been noticed that in substance this was the program the Tudors followed to construct their strong, authoritarian regime. Yet it is also true that the Yorkist kings fumbled with the same devices, although they were denied the opportunities and the capacity vouchsafed to Henry VII. Edward IV (1461–70, 1471–83) was tireless and shameless in his efforts to make the crown financially secure and independent. Gifts, forced loans and the pension he got from Louis XI at Picquigny in 1475 all helped, and enabled him to dispense for long periods with parliament and its grants.

There were also efforts to make the king's council the real seat of power, and through it to develop new procedures for hastening the process of justice. The Tudor prerogative courts, the Star Chamber and others, were clearly foreshadowed by the judicial activity of the council under the Yorkists.

Yet to dwell solely on the political shortcomings of this age, serious although they were, is to put the period in false perspective. In other directions there were signs of vigorous growth, clear intimations of the shape of things to come.

The Growth of National Consciousness and Maritime Enterprise.—National self-consciousness, that feeling of solidarity as a people, widespread among Englishmen in Tudor times, was already gaining ground in the 15th century. One sign of this is the spread of English as the language of all ranks in society. Between 1400 and 1450 the vernacular established itself firmly among the upper classes, ousting French which for generations had been in common use there. The same thing happened, although more slowly, in government departments, while the courts ceased to conduct their proceedings in Anglo-Norman jargon. Under the Yorkists this linguistic change had progressed so far that "secretaries in the French tongue" had to be employed in the service of the state.

This same self-consciousness shows itself, although less markedly, in other ways. The English representatives at the Council of Constance (1414–18) strove to promote there what can not inaccurately be called national self-interest. The anonymous author of the *Libelle of Englyshe Polycye* (1436) gave clear, if somewhat premature, expression to those ideas of militant economic nationalism which dominated English commercial policy from the Tudors to the beginning of the 19th century.

Again, the great maritime enterprise of the Elizabethan age was clearly foreshadowed in the 15th century. The Cornish pirates who then infested the English channel were the forebears of the west country seamen of the next century, while under the Yorkists English merchants were already trading with places as far apart as Iceland and Portugal. Political weakness and civil disturbance grievously hampered both trade and industry. Nevertheless, part of the foundation of English mercantile strength was laid by the enterprise of great families such as the Cannynges at Bristol with their fleet and the Celys trading with the Low Countries in wool, not to mention others like the Forteys, the Busheys and the Boleyns. These people kept clear of politics if they could, and busied themselves with the future, not the past.

Art and Learning.—Although the achievement of 15th-century England in art, learning and literature (*see* ENGLISH LITERATURE: *From Chaucer to the Renaissance*) is not to be compared with that of the succeeding age, nevertheless it has an interest and significance of its own. It is true that in the 15th century the humanist movement, which was first nourished on the literature of the ancient world, and which had already transformed the intellectual and artistic life of Italy, had no wide influence in England. At first it was left to a handful of individuals to make themselves acquainted with the new learning and through their patronage to raise up a succession of enthusiastic disciples.

Humphrey, duke of Gloucester (*q.v.*), was the chief pioneer of humanism in England. He knew Italy and was a patron of scholars there. Leonardo Bruni and Decembrio worked for him, and for his age he possessed a great library of Greek and Latin literature, from which he made valuable gifts to Oxford university. Thomas Beckington, bishop of Bath and Wells (d. 1465), John Tiptoft, earl of Worcester (d. 1470), and William Grey, bishop of Ely (d. 1478), were, like Duke Humphrey, patrons of foreign scholars and collectors of manuscripts; they were not teachers. It was not until the close of the 15th century that William Tilley (d. 1494) and Thomas Linacre (d. 1524) founded a living tradition of classical studies in England, but their task was made far easier by the less-enduring achievement of their humanist predecessors there.

While humanism made slow headway in England, other more elementary and traditional forms of education flourished there in the 15th century. There was no educational movement to compare with that sponsored by the Brethren of the Common Life in the Low Countries, but in England many new grammar schools were founded. Besides exceptional places like Eton (1440), which was linked with King's college, Cambridge, on the model of Winchester and New college at Oxford, erected 50 years before, these schools included many humbler foundations made by individuals and guilds. They were designed not for the poor but for the sons of merchants and landowners. Not less important is the fact that such people were coming to value at least a modicum of education for their children. Literacy was spreading gradually among the upper and middle ranks of lay society and the consequences of this change were to be far-reaching both in politics and religion.

Fifteenth-century Englishmen, particularly the merchants, were not less ready than their forebears to pour out money on art and architecture. The great Perpendicular parish churches of Norfolk, Suffolk and the west country were the chief fruits of this benevolence, but elsewhere a few fine houses, some magnificent bell towers, rood screens and wall paintings also remain. These things were the work of provincial craftsmen, largely unaffected by outside influences. In painting the English had fallen behind their continental neighbours, particularly the Flemings, but despite its flamboyance 15th-century architecture in England was of sustained excellence.

In short it is difficult to catch the spirit or fix the character of the age of the Lancastrians and Yorkists in a definitive phrase. In the pattern of its life, enterprise and violence, promise and failure, hope and frustration were entwined in bewildering complexity.

HENRY IV (1399–1413)

Henry IV had hardly established himself before the political problems which had plagued Richard II returned to vex him. This was because, from the first, the new king was determined to rule as his predecessor had done; that is, with the help of his own trusted servants and without the interference of his natural counselors, as the baronage still liked to describe themselves. Henry was tireless in pursuit of this aim. After a desperate struggle, and despite his shaky title and the illness which dogged him from 1408 until his death on March 20, 1413, Henry succeeded in maintaining what he considered the rights of the crown intact against all comers. The history of his reign is largely the record of this achievement. But, though dealing energetically with his personal enemies, Henry IV was strangely indifferent to the menace of those impersonal forces—administrative weakness, judicial corruption and fiscal improvidence—which steadily sapped the strength of the monarchy he strove to protect and eventually destroyed the dynasty he had struggled to establish. Against these dangers he took no action. Indeed in financial matters the crown was more deeply embarrassed at his death than before his usurpation.

In his political troubles Henry IV had one great advantage: the opposition never made common cause against him. The outcome might have been very different if that had happened. Instead the great families practised armed rebellion, for that weapon had worked well against Richard II, while the majority, who were more timorous, clamoured and schemed in parliament for the king's undoing. When he had crushed the rebels, Henry found it a less difficult task to bridle parliamentary opposition.

The Rebellions.—The first armed challenge to Henry's authority came in Jan. 1400 from a group of Richard II's friends. They were the earls of Huntingdon and Kent, related to the former king on his mother's side, the earl of Salisbury, a noted Lollard, and the lords Despenser and Lumley. After an unsuccessful attempt to seize the king at Windsor, they retired to the west, where Kent and Salisbury were slain at Cirencester, while the others were captured and executed with a number of their supporters. This outbreak almost certainly sealed Richard's own fate (*see* above).

The unsuccessful rising of the earls was soon followed by a movement of a far more formidable kind. This was a rebellion in Wales, where sympathy for Richard II was widespread and where his natural heirs, the house of Mortimer, descended from Richard's eldest uncle, Lionel, duke of Clarence, had wide estates. It was led by Owen Glendower (Glyn Dwr), a substantial landowner, descended from the native princes, who had quarreled with Henry's friend, Lord Grey of Ruthyn. Glendower proclaimed himself

prince of Wales in Sept. 1400, and although the king in person led two expeditions against him before the end of 1401 these only served to demonstrate that long and difficult fighting would be required to suppress what had now become a genuinely national uprising.

Meanwhile Henry had been threatened by a revival of the Franco-Scottish alliance and a renewal of the continental war. This danger was stifled for the moment by a truce with the French made in Aug. 1401 and by the decisive defeat of the Scots at the battle of Homildon hill just a year later. This victory was the work of the earl of Northumberland and of his son Henry Percy, nicknamed Hotspur. They expected their recompense from the ransoms of the prisoners they had taken. When the king refused them this they conspired against him and engineered the most serious of all the plots he had to face. They came to an arrangement with Glendower and enlisted the support of Northumberland's brother, Thomas, earl of Worcester. In July 1403 Hotspur raised the men of Chester and of the Welsh march, but before he could be joined either by his father or Glendower, the king brought him to action at Hately, near Shrewsbury (July 21, 1403). There Hotspur was slain, the earl of Worcester was taken and their army destroyed. In this, the crisis of his reign, Henry had acted with extraordinary vigour to prevent what might well have been a fatal coalition of his domestic enemies.

All the same he felt the repercussions of these events for several years longer. In 1405 Charles VI of France sent help to Glendower and the invaders penetrated almost to Worcester before they were compelled to withdraw. Thereafter Henry of Monmouth, the king's son, took charge of the Welsh war and slowly sapped Glendower's strength. Glendower was driven out of south Wales in 1406 and his castles of Aberystwyth and Harlech, the linchpins of his power, were taken in 1408. From 1409 onward the Welsh rebellion gradually died down. Glendower, who was never caught, became once again a guerrilla chief, a nuisance but not a menace to the king.

In Feb. 1405 Northumberland, who had been pardoned for his share in the events of 1403, organized a new coalition against the king in the north. He was joined by Henry's old enemy, Thomas Mowbray, the earl marshal, by Lord Bardolf and by Richard Scrope, the inoffensive archbishop of York, who was included to give the movement some popular appeal. Mowbray and Scrope were betrayed to the king, who had them executed at York after a summary trial. Henry then proceeded to a systematic reduction of the Percy castles, and the earl of Northumberland fled at his approach into Scotland. He returned after nearly three years of exile only to be caught and slain by the sheriff of Yorkshire at Bramham moor in Feb. 1408.

Opposition in Parliament.—In parliament Henry had to face a different but not less tenacious kind of opposition. The causes of the trouble were familiar enough; the king's determination to rule with ministers of his own choice and his constant need of parliamentary subsidies. Nor was there anything very new in the demands of the opposition in parliament. They asked that Henry render a strict account of the monies granted him and show that they had been used for the purposes for which they were intended. Further he was required to be guided by advisers known to and approved by his subjects in parliament. What was unusual was the persistence with which the commons, and the knights especially, pressed their case. They were feeling their way toward a doctrine of ministerial responsibility to parliament enforced by the power of the purse. It is true that this promise of constitutional growth was not immediately fulfilled, but it was nevertheless a significant precedent. This new-found determination on the part of the commons was best exemplified in the parliament of 1406, which lasted, with two adjournments, from March to December. Henry rid himself of the restraints it had imposed a few months later in a parliament held at Gloucester but, although he repeated these tactics on several occasions, besides interfering with elections and browbeating members, he never succeeded either in curbing the critical spirit of the commons or in slipping entirely free from their growing financial power.

During the last years of the reign, when Henry IV was often ill and there were rumours that he might abdicate, his eldest son, Henry of Monmouth, gained steadily in influence. His relations with his father were never easy, for he grasped greedily after power. It is significant that the prince was even then contemplating military action in France, where the quarrels of the Burgundians and the Armagnacs, after the murder of the duc d'Orléans in 1407, presented a tempting opportunity.

Henry V (1413–22)

Henry V succeeded his father in March 1413 and died of fever at Bois de Vincennes in Aug. 1422, when he was only 35. As king his achievement was spectacular. It caught the imagination of his own age and it lived on after him. He is best remembered as a soldier of genius, but he also proved himself a strong and capable ruler. Side by side with his French victories there must also be placed the fact that for a few short years he gave the English kingdom the resolute leadership it so rarely got in the later middle ages. Under him there was no lack of governance. He succeeded, as no one else ever did, in rousing a general and sustained enthusiasm for the French war, and when he died his people had achieved a commanding influence in Christendom.

But there was a less satisfactory side to the account. Henry V died before the French could sink their differences in face of the common peril and muster their superior strength for the inevitable counterblow. His very success was, in some measure, the result of his skilful exploitation of these differences, and when they were healed English power in France at once became very precarious. There never was any permanent prospect of establishing the great continental dominion of which Henry dreamed, and it is no tribute to his political sagacity that he committed his people so deeply to this fruitless and costly enterprise. Again, in money matters Henry was notoriously improvident, even for a prince of that age. The war was financed mainly by reckless borrowing and even before the king died there were ominous rumblings of parliamentary discontent. Finally, as a man Henry V was harsh, crafty and cruel, something of a hypocrite and pitiless to his enemies. He had much in common with the Italian despots of the period.

For most of Henry's reign his strong hand, together with the demands of the war, smothered the domestic troubles which had spilled over from his father's time. Only two outbreaks disturbed this unwonted tranquility. On the eve of the king's departure for France in Aug. 1415 his cousin Richard, earl of Cambridge, together with Lord Scrope and Sir Thomas Grey, a kinsman of the Percies, plotted to kill him. The earl of March betrayed the conspirators and they were promptly beheaded. The second disturbance was of a different kind. It was the rebellion inspired by the Lollard, Sir John Oldcastle (*q.v.*).

The Lollards.—Since about 1380 Wycliffe's teaching, in which doctrinal heresies and anticlerical sentiments were nicely blended, had taken root in the popular movement known as Lollardry. Faced with the unaccustomed task of suppressing a dangerous heresy, the ecclesiastical authorities showed more zeal than efficiency and at first, except for help in banning Lollard writings, they received little assistance from the government. Meanwhile Lollard influence waxed not only among the poor but among the gentry as well. Moreover, Lollard opinions were frequently voiced by the commons in parliament. Under Henry IV things changed, for the king was prepared to lend his aid to the bishops against the Lollards. The statute *De Haeretico Comburendo* (1401) meant that in future death by burning at the stake might await the obstinate heretic and during his reign a number did suffer this penalty. In his time Thomas Arundel, the archbishop of Canterbury, broke the power of the Lollards in Oxford, which had been the chief centre of their influence.

Oldcastle, a fervent Lollard, had been on friendly terms with Henry before his accession. When he took the throne, however, the king quickly disowned him. He was arrested by royal officers in Sept. 1413, and when he steadfastly refused to recant he was condemned in St. Paul's. He soon escaped from prison and busied himself in preparing a Lollard rebellion. The government circulated the rather improbable story that he intended to dethrone the king and make himself protector of the commonwealth. Old-

castle certainly planned a rising in London for Jan. 1414, but this was forestalled by the king's prompt action, although the chief conspirator managed to avoid capture. Thereafter the law against Lollardry was further stiffened, and Oldcastle himself was finally captured in Wales in 1417. His execution and the brutal repression meted out to Lollards in general destroyed the public influence of the sect, but it continued to exist in obscurity. In the middle years of the century people, mostly poor and ignorant, were still being punished for expressing Lollard opinion in dioceses up and down the country. (*See* LOLLARDS.)

Renewal of the Hundred Years' War (1415 to 1420).—The French war was renewed before the Lollard movement had been finally crushed. Henry V began military operations across the channel in Aug. 1415 after careful and deliberate preparation. In Nov. 1414, before a breach with the French was imminent or even certain, parliament gave him the money he needed with surprising readiness, and by the summer of 1415 he had raised and equipped what was, by the standards of the time, the large force of 12,000 men. Meanwhile, by diplomacy Henry strove to force the French either to war or a humiliating surrender. His envoys busily negotiated with the Armagnacs, who had the poor, mad Charles VI as their puppet, and with their bitter rival, John the Fearless, duke of Burgundy. To both the English made staggering demands. They wanted nothing less than a revival of the Angevin empire, together with the marriage of Charles VI's daughter to their king.

Since neither side would buy his aid at such a price, Henry announced that he himself would restore peace and order to the distracted French kingdom. John the Fearless stood aside and Henry busied himself with the siege of Harfleur, which surrendered in Sept. 1415. Then, since the fighting season was near its end, the English withdrew northward, for their operations were more in the nature of a raid than a regular campaign. However, they were hotly pursued by the Armagnac host and forced to battle at Agincourt in Artois (Oct. 25, 1415). Here the French repeated the tactical errors they had made at Crécy nearly 80 years before and were cut to pieces by the smaller English force.

The English gained surprisingly little from their great victory. Three weeks after the battle Henry V sailed home from Calais and it was nearly two years before he returned to France. Meanwhile, Charles VI and his Armagnac masters, whose power though much battered was not broken, sought outside help, and in particular the aid of Sigismund, king of the Romans. He was ready to intervene because the quarrels between French and English, Burgundians and Armagnacs threatened to wreck the Council of Constance then meeting under his patronage with the aim of ending the papal schism and dealing summarily with the adherents of the Bohemian heretic John Huss. Sigismund, however, chose to ally himself not with Charles VI but with Henry V, and by the treaty of Canterbury (Aug. 1416) promised to help him against the Armagnacs. This pledge was of little practical use, but it had the important effect of persuading John of Burgundy to come out openly on the English side.

With the assurance of Burgundian co-operation Henry V returned to France in Aug. 1417 and began the systematic conquest of Normandy. Before Rouen fell (Jan. 1419) the Armagnac party had disintegrated; John of Burgundy was master of Paris and custodian of Charles VI, while effective opposition to the English rallied round the dauphin, later Charles VII, who was soon to be disowned by his parents. After John the Fearless had been murdered (Sept. 1419), his son Philip, nicknamed the Good, made what was intended to be a permanent settlement with Henry V, whose power was now spread far and wide in northern France.

The Treaty of Troyes.—This settlement, formally concluded in the name of Charles VI, was embodied in the treaty of Troyes (May 1420). It provided that Henry V should marry Catherine, daughter of Charles VI, and should then be recognized as heir presumptive of the French king. Until the latter's death Henry was to rule in France as regent with Normandy as his appanage. No change was to be made in the kingdom's institutions and, as regent, Henry was to act together with Philip the Good, whose great domains in northeastern France were not to be subject to Lancastrian interference. In fact what can best be described as a dual monarchy was envisaged. Henry V and his successors were to reign over two separate kingdoms only united by their personal authority. In June 1420 Henry and Catherine were married and the following December Henry and his father-in-law solemnly entered Paris.

The actual situation, however, was far different from that proposed at Troyes. Henry was master of Normandy, where most of the landowners had been turned out to make room for English knights, who were thus quartered on the countryside as a permanent military garrison with a personal interest in maintaining themselves there. With the help of Philip of Burgundy, Henry also controlled Paris and its immediate neighbourhood. But there his direct influence ended. All central and southern France, with the exception of English Gascony, acknowledged Henry's opponent, the dauphin, soon to be derisively called king of Bourges, while much of the eastern part of the kingdom was the personal dominion of Henry's ally, the duke of Burgundy. There is no doubt that in France north of the Loire the treaty of Troyes was widely popular, but only because it was supposed that Henry V would be able to put an end to the civil disturbance and the foreign war which had plagued the kingdom for so long. To do this he had to succeed in the formidable task of breaking the power of the dauphin and of maintaining the Anglo-Burgundian alliance. It was because they never succeeded in doing either of these things that the English eventually lost all they had gained in France. Henry V himself was scarcely able to begin the job before he was overtaken by death on Aug. 31, 1422.

He was succeeded in England by his son Henry, a child eight months old, who on the death of his grandfather, Charles VI, in Oct. 1422, was proclaimed king of France as well.

HENRY VI (1422–61 and 1470–71)

Henry VI reigned for nearly 40 years until he was deposed in March 1461. He was restored in Oct. 1470, deposed again in April 1471, and he died in May of the same year. He was never able to play that part in the management of affairs which was expected of him. His minority formally ended in Nov. 1437, but by then he was already the victim of that nervous instability which made him incapable of shouldering responsibility or of bearing the strain of public life. After 1452 the king's malady worsened, for he then suffered periods of complete mental incapacity. The cause of Henry's chronic illness is not known with certainty but he may have inherited a strain of insanity from his maternal grandfather, Charles VI of France. In any case, since Henry himself was unfit others managed the French enterprise and governed England on his behalf.

The End of the Hundred Years' War.—*Bedford's Regency in France.*—After Henry V's death the control of English interests in France passed to his brother John, duke of Bedford (*q.v.*), uncle of the new king. Bedford was a good soldier and an admirable administrator who deliberately set himself to continue his dead brother's policy. Yet despite his most strenuous efforts he failed, and when he died in Sept. 1435 English power in France was already crumbling.

The main cause of this change lies in the fact that Bedford was unable to maintain the Anglo-Burgundian alliance. Within a week of Bedford's death, Philip the Good finally laid aside his political hesitations and, at the treaty of Arras (1435), committed himself to the cause of Charles VII.

Bedford began well. He tried to make the English occupation tolerable, particularly in Normandy; he was not without success in curbing brigandage and he contrived a form of administration which was both smooth and efficient. In April 1423 at Amiens the Anglo-Burgundian alliance was strengthened by the adherence of John VI of Brittany and by the agreement for the marriage of Bedford himself to Anne of Burgundy. This was to herald a concerted attack on the position of Charles VII south of the Loire. Yet it never materialized, mainly because Philip of Burgundy did not intend to co-operate wholeheartedly with the English unless it was clearly in his own interest to do so. For his part Bedford lacked the solid support of those at home on whom Henry V had been able to rely. Reinforcements and supplies were doled out to

him grudgingly, and they were not sufficient to convince Burgundy that it would be profitable to work actively with his allies. Then Humphrey, duke of Gloucester, Bedford's brother, made matters worse by his liaison with Jacoba (Jacqueline) of Bavaria, the runaway wife of John, duke of Brabant. Gloucester and Jacqueline were married in 1423, and he straightway tried to obtain her continental inheritance, the county of Hainaut, which was within the Burgundian sphere of influence. These irritations stretched the Anglo-Burgundian alliance almost to breaking point, and when operations against Charles VII were at last begun it was by the English alone. Their army invested Orléans, the key to central France, in Oct. 1428, and it was then that Joan of Arc intervened.

Joan of Arc and the Treaty of Arras.—There is little evidence that her contemporaries, including Bedford and the English who compassed her destruction, took Joan of Arc as seriously as posterity has done. Yet it is clear that in the short time vouchsafed her before she was burned in Rouen (May 1431) Joan did achieve two things of capital importance. By raising the siege of Orleans she showed that the Lancastrians were not invincible. By insisting that Charles VII should be crowned at Reims she secured for him a stronger claim on the loyalty of all his subjects than his English rival could ever hope to possess. (*See* JOAN, SAINT.) The immediate result of Joan's work was soon apparent. Philip of Burgundy abandoned the English and at the treaty of Arras (Sept. 20, 1435) made his peace with Charles VII. Thereafter the balance in France tipped suddenly and heavily against the English. Bedford was dead. They were quickly driven from Paris and the Ile-de-France so that they soon found their effective power confined to Calais, lower Normandy and Gascony. Elsewhere Charles VII was universally acknowledged.

These reverses were not without effect in England. There Cardinal Beaufort was the leader of a growing party which desired to put an end to the war, but another powerful group looked to Gloucester, the king's uncle, who was for fighting on. Charles VII, however, did not prosecute his struggle against the English with vigour after his capture of Paris, not least because he too found Philip of Burgundy an unreliable associate. In consequence Richard, duke of York, and John Talbot, earl of Shrewsbury, one of Bedford's trusted captains, were able, by the end of 1436, to restore English authority in Normandy and Maine. Until 1441 the war dragged on without obvious advantage to either side. Then Charles VII bestirred himself. First he destroyed the outposts which Talbot had laboriously re-established in the Ile-de-France. Then he turned to Gascony, where he still further reduced the steadily shrinking area of English power and threatened Bordeaux itself.

In England, where Gloucester's influence had waned, there was now a widespread desire for peace. Accordingly William de la Pole, earl of Suffolk, concluded at Tours in May 1444 a truce with the French, which was cemented by the betrothal of Henry VI to Margaret of Anjou, the niece of Charles VII. Three years of fruitless negotiations followed in which both sides tried and failed to use the truce of Tours as an opportunity for making a permanent settlement without further fighting. The French asked too much while the English would concede too little. Finally, although the truce was still in force, Charles VII pushed the English garrisons out of Maine in the summer of 1448, and by so doing gave notice of his intention to recover his entire kingdom.

The English Loss of Normandy and Gascony.—Suffolk, now dominant in English politics, was thus faced with an intolerably difficult situation. He could not have a settlement on terms his own countrymen would accept, yet he lacked the means to coerce the French. So when early in 1450 Edmund Beaufort, duke of Somerset and cousin to Henry VI, arrived in Rouen as the king's lieutenant and deliberately provoked trouble with Charles VII in Brittany, his actions were no more than empty motions of defiance. Yet they gave the French king the excuse he wanted. At Formigny (April 15, 1450) the English army was destroyed by a smaller, but now technically superior, French force, and soon all Normandy was finally lost. The destruction of the remnants of English power in Gascony was a more difficult and lengthy task, but it was achieved with the surrender of Bordeaux to the French in Oct. 1453. This was the last notable event of the Hundred Years' War. In sharp contrast to its ceremonial beginning, its end was not marked by any formal settlement. Henry VI still held Calais and the county of Guines, but the shadow of civil disturbance hung heavily over the English kingdom, and for the next 30 years its rulers had no time to spare for foreign adventures.

Domestic Rivalries.—During the minority of Henry VI (Aug. 1422–Nov. 1437) the English kingdom was managed by the king's council, now a predominantly aristocratic body. This arrangement, which probably did not accord with Henry V's last wishes, was not maintained without difficulty. Like Richard II before him, Henry VI had powerful relatives eager to grasp after power and to place themselves at the head of factions in the state. The council soon became their battleground.

Gloucester and Beaufort.—Humphrey, duke of Gloucester, the king's uncle, had hoped to be regent, but he had to be content with the empty title of protector, until Henry's coronation in Nov. 1429 deprived him of even that. Although he is justly famous as a patron of scholars, Gloucester was a domineering and untrustworthy politician. His principal rival, Henry Beaufort (*q.v.*), bishop of Winchester, was more competent but not less self-interested. After 1426, when his brother Thomas, duke of Exeter, died, he was the last survivor of the three male Beauforts, children of Catherine Swynford and John of Gaunt. Beaufort had been constantly employed in the royal service since 1403, acting twice as chancellor, but his ambition, which Henry V had curbed, was service in the Curia Romana. He received a cardinal's hat in 1427, when he also led a fruitless crusade on the pope's behalf against the Hussites. From then until 1432 Beaufort was frequently absent from the kingdom, and he only reluctantly abandoned his hopes of a great career in the curia when Gloucester's threat to his English interests became menacing. The cardinal was also notable for his inordinate love of money. For his time he was a man of great wealth, and he lent vast sums to the crown in hope of profit for himself.

It was Beaufort and his friends of the council who prevented Gloucester from becoming regent in 1422, but by 1425 the duke was threatening to use force against them and Bedford was summoned home to mend matters. When, early in 1427, Bedford returned to France Beaufort accompanied him, leaving behind his friend John Kempe, archbishop of York, as master of the council. He and his fellows kept Gloucester in order until 1430, when the need to secure Henry VI's coronation as king of France, because of the growing prestige of Charles VII, took half the council across the channel with the king.

This was Gloucester's opportunity. With unusual care he gathered allies—a group headed by John, Lord Scrope—in the remnant of the council. Steps were taken to charge the absent Beaufort with praemunire, and when the king returned in Feb. 1432 Gloucester and his friends engineered a change of government. Kempe, the chancellor, was removed, Scrope succeeded Hungerford as treasurer and most of Beaufort's other supporters in the council were also dismissed. The cardinal himself retired to Winchester until July 1433 when Bedford returned once more, not to rescue him but to pull the English government together and to raise money now desperately needed for the French war. This he failed to do, but Lord Cromwell, the treasurer, made plain to a parliament meeting at Westminster in Oct. 1433 the government's seemingly incurable financial difficulties, which were slowly undermining its efficiency and bringing its authority into contempt. Ordinary revenue regularly fell far short of ordinary expenditure, and there was a debt of nearly £170,000 which had been swollen by lavish grants made to those with influence. Before he went back to France in July 1434 Bedford secured Beaufort's readmission to the council. During the next four years, when Bedford, the one stout prop of the Lancastrian dynasty, was removed by death, Beaufort slowly gathered power into the hands of his own party.

Beaufort and Suffolk.—The council, which had been formally reappointed when Henry's minority ended in Nov. 1437, was now displaced from the control of affairs, which passed to the officers of the household under Suffolk, the steward, supported by Beaufort's many friends among the territorial magnates. Until be

retired in 1443 the cardinal's ascendancy was complete and unchallenged, not least because of the skill with which he manipulated royal patronage. Even Gloucester ceased to trouble him. The king took little part in politics, and when Beaufort withdrew, Suffolk, his chief henchman, took his place, maintaining himself with increasing difficulty until he was overthrown in 1450.

Although Suffolk (*see* SUFFOLK, WILLIAM DE LA POLE, Duke of) was not the tyrant he has sometimes been called, he had none of the qualities required to steer the kingdom away from the shoals onto which it was fast drifting. Like Beaufort he was grasping and corrupt yet he saw that the interests of the country as well as his own safety demanded an end to the French war. What he failed to realize was that although many of his countrymen wanted peace they were only prepared to have it on their own terms. After the failure of his efforts to make a settlement with Charles VII after the truce of Tours, Suffolk's popularity waned, and it shrank still further because of his treatment of Gloucester and York, two of the king's most powerful subjects. Gloucester, who had begun by supporting the truce of Tours, later became its most bitter opponent, so that Suffolk and his assistants decided he was too dangerous to be left at large. Accordingly he was impeached in a parliament which met at Bury St. Edmunds, where Suffolk's influence was strong, in Feb. 1447. Gloucester died suddenly, probably from natural causes, before his trial began, but Suffolk's enemies gleefully accused him of murdering "the good Duke Humphrey" and the charge was widely believed.

Suffolk and York.—Gloucester's death left Richard, duke of York (*q.v.*), as heir presumptive to the crown, and this was an ominous development both for Henry VI and his dynasty. The duke was the greatest of the territorial magnates. His father and his uncle had been executed as traitors to the Lancastrian house, and he combined in his own person the strong hereditary claims to the crown directly transmitted from Lionel of Clarence and Edmund of York, who were the third and fifth sons of Edward III, whereas John of Gaunt, ancestor of Henry VI, was Edward's fourth son. Since Beaufort's death Richard of York had been much occupied in Normandy, latterly as the king's lieutenant. Rumour had it that he had misapplied public funds there, and he returned in 1446 to clear himself. He suspected that Adam de Moleyns, keeper of the privy seal, bishop of Chichester and Suffolk's principal assistant, was his chief traducer. After an open quarrel between them, it was decided to send York into virtual exile as the king's representative in Ireland (Dec. 1447). From that time he was Suffolk's implacable foe.

In the country at large also there were widespread and menacing signs of trouble. The government grew more feeble and its debts mounted by 1450 to more than double the figure they had reached in 1437. In consequence there was a general loosening of the bonds of order and unity throughout society. Some territorial lords had acquired the habit of settling their private disputes by arms and not in court. Their bands of armed retainers were growing in numbers, and in some places already terrorized the countryside. The gentry too hastened to ally themselves with powerful protectors, and helped to pervert the ordinary course of justice by the intimidation of witnesses and the bribing of sheriffs and jurors. Parliament, which met infrequently, did not rise to the need of the times. The commons were niggardly with their grants for the French war, they failed to protest against the emasculation of the council and clamoured unceasingly against alien merchants in England.

The final loss of Normandy in 1450 brought about Suffolk's fall and his murder off Calais as he went into exile. It also provoked numerous outbursts of popular resentment of which the most serious was the rebellion of the men of Kent under one who called himself John Mortimer, but whose real name was apparently Jack Cade (*q.v.*). He and his followers broke into London (July 1450) demanding the return of Richard of York, the formation of what they called a government of "true" barons, an improvement in the administration of justice and the repeal of the Statute of Labourers. The movement was suppressed with difficulty, and just as it was all over York returned unbidden. In consequence Edmund Beaufort, duke of Somerset, nephew of the dead cardinal and, after the king, the leading member of the Lancastrian house, was hastily summoned from France to be constable of England.

The Beginning of the Wars of the Roses.—Politics now revolved round the quarrel of York and Somerset, until the latter was killed by his rival's men at the first battle of St. Albans (May 1455). After that the king's faction was led by the queen, Margaret (*q.v.*), of Anjou, who since the birth of her son, Edward (Oct. 1453), had taken an active part in politics.

Between 1450 and 1460 Richard of York became the head of a great baronial league, of which the foremost members were his kinsmen, the Nevilles, Mowbrays and Bourchiers. Among his principal lieutenants was his able nephew, Richard Neville, earl of Warwick (*q.v.*), a powerful man in his own right, who had hundreds of adherents among the gentry scattered over 20 counties. Yet until 1460 York was not openly aiming to displace Henry VI. He accepted the loss of the succession after the birth of Henry's heir, and acted as protector during the king's two long fits of insanity without trying to seize permanent power for himself.

All the same he was prepared to use force against the government and to break his oaths when it suited him, as it did in 1455. Thus the long period of intermittent civil disturbance called the Wars of the Roses (*q.v.*) opened almost hesitantly, as politics slowly degenerated into armed bickering between two strong baronial leagues, the one serving the king, the other loyal to York.

A new phase began in 1459 when York, goaded by the queen's undisguised preparations to attack him, rebelled for the last time. After the rout of his followers at Ludlow, York fled to Ireland, and the Lancastrians, in a packed parliament at Coventry (Nov. 1459), obtained a judicial condemnation of their opponents and executed those on whom they could lay hands. From then on the struggle became bitter. Both parties laid aside their scruples and struck down their opponents without mercy. The cold-blooded and calculated ferocity which now entered English political life certainly owed something to the political ideas of the Italian Renaissance, but it was also in part a legacy of the lawless habits acquired by the nobility during the Hundred Years' War.

The Lancastrian triumph was short, for in June 1460 Warwick returned from France whither he had fled, and at Northampton (July 10, 1460) inflicted a crushing defeat on the king's army. York came back and in a parliament at Westminster (Oct. 1460) it was decided that Henry should retain the crown for life but that York should succeed him. But there was still the queen to reckon with. Marching with a small force to expel her from the north, Richard of York was surprised and slain at Wakefield (Dec. 30, 1460).

The Lancastrians gave no quarter, and they repeated the same tactics two months later, when their northern army advancing on London met and defeated Warwick at the second battle of St. Albans (Feb. 1461). But Edward of York, Richard's heir, was approaching from the west. He entered London with Warwick on Feb. 26, 1461, was acclaimed king, pursued the retreating Lancastrians northward and destroyed them at Towton (March 29, 1461). Henry VI and his family fled into Scotland.

EDWARD IV (1461–70 and 1471–83)

At his accession Edward IV was a handsome youth of 19, able, impulsive and not lacking in military skill. But before he died, on April 9, 1483, he had enjoyed 12 years of relative security, for his enemies were all dead, and this drew out his worst side. He showed himself cruel, grasping and somewhat careless of his duties. He was not wanting in political skill and subtlety but, like his brother and successor, Richard III, he lacked both the will and the purpose to carry through the changes needed to re-establish the power of the crown firmly and to restore the unity and strength of the kingdom.

The Ascendancy of Warwick.—Warwick was the true architect of the Yorkist triumph. He was the statesman of the group and a worthy representative of his class. Until 1464 he was the real ruler of the kingdom. He ruthlessly put down the survivors of the Lancastrians who, under the influence of Margaret of Anjou and with French help, kept the war going in the north and in Wales. The wholesale executions that followed the battle of Hexham (May

1464) practically destroyed what was left of the Lancastrian party, and the work seemed complete when, a year later, Henry VI was captured and put in the Tower of London.

Warwick made an equally vigorous effort to put the government of the realm in better shape, to restore public order, to improve the administration of justice and, by confiscations and economies, to make the crown solvent. At the same time both Warwick and his master were caught in the diplomatic schemes of Louis XI of France, the astute man who had succeeded Charles VII in 1461. He was still preoccupied with the power of Burgundy and the English were to be the pawns in the game he intended to play for the humbling of Charles the Bold.

Yet Edward IV was not prepared to submit indefinitely to Warwick's tutelage, efficient and satisfactory though it proved to be. It was not that he deliberately tried to oust Warwick, rather he found the earl's power irksome. Edward's hasty and secret marriage to Elizabeth Woodville in 1464 was the first overt sign of his impatience. The Woodvilles, a family with strong Lancastrian connections, never achieved real political influence, but they climbed into positions of trust near the king, thus estranging Warwick still further.

The open breach between the king and the earl came in 1467. Edward dismissed Warwick's brother, George Neville, the chancellor, repudiated a treaty with Louis XI which the earl had just negotiated and concluded an alliance with Burgundy against which he had always protested. Warwick then began to organize opposition to the king. He was behind the armed protest of the gentry and commons of Yorkshire that was called the rising of Robin of Redesdale (April 1469). A few weeks later, having raised a force at Calais and married his daughter Isabel without permission to the duke of Clarence, Edward's brother, Warwick landed in Kent. The royal army was defeated at Edgecote and the king himself became the earl's prisoner, while the queen's father and brother, together with a number of their friends, were executed at his command. But the experiment in coercion failed through lack of support and by April 1470 Warwick and Clarence were back in Calais. Then, probably at the instigation of Louis XI, Warwick came to terms with Margaret of Anjou. In Sept. 1470 he was back in England with a fresh army and barely a month later Edward IV sought refuge in Holland. Henry VI was restored and for six months Warwick ruled as his lieutenant.

Warwick's power was insecure, however, for the Lancastrians found it difficult to trust one who had so lately been their scourge, while many of the earl's Yorkist followers found the change more than they could bear. There was thus little real opposition to Edward IV when he returned from Flushing to land at Ravenspur (March 1471) in a manner reminiscent of Bolingbroke. His forces met those of Warwick on April 14 in the battle of Barnet (*q.v.*), and there the earl was slain. On the same day Margaret of Anjou and her son, who had hitherto refused to return from France, landed at Weymouth and gathered the Lancastrians of the west for a final rally. Less than a month later (May 4, 1471) her army was crushed at Tewkesbury, her son slain and she herself taken.

Edward IV marked his victory with the wholesale execution of the more important of his prisoners, including the last Beaufort duke of Somerset. Henry VI died in the Tower on May 21, 1471; his end was so opportune that many believed the king and his brother Richard of Gloucester were responsible.

Edward's Personal Rule.—For the last 12 years Edward held his own without difficulty, but the methods he used did no more than prop up his personal authority; they did not restore the strength of the monarchy. Of course it required more time than he had to cure the gentry, on whom local government had now come to depend, of ingrained habits of lawlessness, but he could have applied the known remedies far more vigorously than he did. These were the development of the power of the council together with the modernization of the common law and the welding to it of a more speedy and flexible type of conciliar justice. Instead at the end of his reign men were still clamouring for "a good rewyll and a sad (strong) in this contre."

Edward's government paid its way, but largely through temporary expedients. These were the profits of many confiscations and the annual pension Edward received from Louis XI after 1475 for cynically abandoning his Burgundian ally, his brother-in-law Charles the Bold. The king also traded in cloth and tin on his own account. Curiously enough Edward's much condemned benevolences or forced loans probably represent his one constructive effort at financial reform. They were a deliberate attempt to spread the burden of taxation over the growing mercantile wealth of the community, must of which escaped the net of parliamentary subsidies.

In other directions Edward's doings were equally opportunist and hand-to-mouth. He took Louis XI's money and meekly allowed his paymaster's suspicions to involve him in a Scottish war. This was fomented by the French king to keep Edward busy after 1477, when the great duel between the French and the Habsburgs began. Finally in 1478 Edward and Gloucester were almost certainly involved in the murder of their brother Clarence. He was denounced in parliament by the king as incorrigible and died by violence soon afterward. It seems likely that, with such an example before him, Gloucester, who was Edward's disciple in political method, did not shrink from murdering the king's own sons, Edward and Richard, when that seemed expedient.

Richard III (1483–85)

Richard III, the last of the Yorkists, remains a shadowy figure. He was grossly caricatured by Tudor propagandists so that it is now difficult to see him as he was. He seems to have been a hard-working, businesslike prince but, though personally pious, he was, like many of his contemporaries, devoid of scruple in political matters.

Richard came to power by stages. In April 1483 he was acknowledged protector for Edward V, a boy of 12. Then, on the pretence that they were conspiring against him, he caused a number of the queen's relatives and some of the late king's closest friends to be done to death. Next, on the evidence of an alleged precontract between Edward IV and Lady Eleanor Butler, which made the Woodville marriage illegal, Richard declared his nephews illegitimate. At the end of June 1483 Edward V was deposed. He and his nine-year-old brother were murdered in the Tower, probably only a few weeks later.

Yet from the moment of his coronation Richard's power was precarious. Henry Stafford, duke of Buckingham, his closest associate, quickly turned against him, and by the autumn of 1483 a widespread although as yet ill-organized conspiracy was in being. One of its moving spirits was Margaret Beaufort, now married to Lord Stanley, head of a great northern family. Margaret contrived to draw together Buckingham, the surviving Woodvilles and Edward IV's widow into an association on behalf of her son by a former marriage, Henry Tudor, earl of Richmond, then an exile in France. It was agreed that he should marry Elizabeth, the last surviving child of Edward IV, and thus at last draw together the warring descendants of Edward III's many sons.

The plot issued in rebellion in Oct. 1483 but it failed, for Buckingham, who had mustered in Wales, was hunted down and killed, while Henry Tudor's little force was prevented by storms even from landing on the English coast. The respite thus gained by Richard III was no more than temporary for his position worsened in 1484, when his only son died, and his hirelings, Lord Lovell and William Catesby, maintained his authority only by naked force. The conspirators took heart again and Henry Tudor gathered a formidable company of his friends in France. On Aug. 7, 1485, he landed at Milford Haven with a powerful following of exiles. The Welsh joined him and when he reached Shrewsbury Richard's English subjects flocked to his banner. When he met Richard's army at Bosworth Field (Aug. 22, 1485) the king's force was much the larger, but the greater part of it was determined not to fight. The issue was not long in doubt, although Richard and his more desperate partisans fought to the last. When it was over the victorious army saluted their leader by the new title of Henry VII.

VIII. ECONOMIC AND SOCIAL DEVELOPMENT IN THE LATER MIDDLE AGES

At intervals in the foregoing paragraphs mention has been made

of economic matters, particularly as they influenced the pattern of domestic and foreign politics in England during the 13th, 14th and 15th centuries. Now it is possible to glance briefly at the economic history of this period as it were in its own right.

Decline of Manorialism

In this period England was still a country in which agriculture was the predominant economic activity, as it continued to be until the first impact of modern industrialism was felt in the second half of the 18th century. Nevertheless, between *c.* 1300 and *c.* 1500 English agriculture changed its character in some important respects. In technical matters there was little improvement. Th changes came in methods of land holding and in the relations between peasants and their lords. Nor was this solely an English phenomenon, for it was a movement common to the whole of Christian Europe, and sufficiently uniform for the same features to be discerned everywhere below the superficial variations imposed by differing local circumstances. In England the general nature of the change can be simply stated. Lords turned into landlords and serfs became tenants in the modern sense of that term. Money rents replaced labour services and dues previously paid in kind. This economic change was the instrument of far-reaching social changes, which had the general effect of conferring universal legal freedom, while mitigating to a small extent the rigours of social dependence and economic subjection which had characterized the feudal order in its heyday. It is much less easy to chart the progress of this general change even in so small an area as England, and to uncover the causes which brought it about.

At first historians thought that the process of change or commutation, the technical phrase for the complicated agrarian transformation just mentioned, started on a broad front in the second half of the 14th century as a direct consequence of the widespread social and economic disorganization brought about by the Black Death of 1348–49. Until then, it was held, the feudal order in the countryside represented by the classical manor and the nexus of economic and social ties it involved between peasants and lords, between the many and the few, was more or less universal and intact. Modern work has shown that this account is inaccurate because it is too simple. We now know that manorialism was never universal in England; it was strongest and most fully developed in the mixed farming areas of the Midlands and the south, and particularly on the great ecclesiastical estates there. But in the hilly districts of the north and the west and in the Fens, even in the high feudal age, the grip of the lord was never so tight and the peasantry retained far more personal and economic freedom.

Moreover, in the 13th century the growth of regional markets for foodstuffs and of international markets for commodities like wool had already transformed English estate management. Production for the market was now competing with production for subsistence. For this purpose wage labour was more effective than customary labour service, and the large domain was preferable to a multitude of small plots. Thus it is now clear that long before 1300 these developments had brought many changes, even in those parts of England where manorialism was most highly developed and firmly entrenched. Then, as has already been mentioned, soon after 1300 the economic climate began to change all over Europe; population ceased to grow and soon began to decline, demand contracted spasmodically and prices tended to sag. This period of stagnation lasted until almost the end of the 15th century, for it was intensified and prolonged by war and plague. Its exact causes are still unknown, but its effect in the English countryside is plain. It completed the destruction of manorialism, already begun under the different conditions of 13th-century prosperity. In the 14th and 15th centuries fixed money rents were more attractive to those possessing estates than the direct exploitation of their lands. Labour was scarce and tended to get dearer, and customary services were difficult to enforce, particularly after the plague of 1349. So the yeoman, the peasant farmer and the day labourer replaced the villein and the serf. Land was now held on lease or by copy of the court roll, legal servitude disappeared and the enclosure of arable pasture and waste, still generally held in severalty, began to be widely practised.

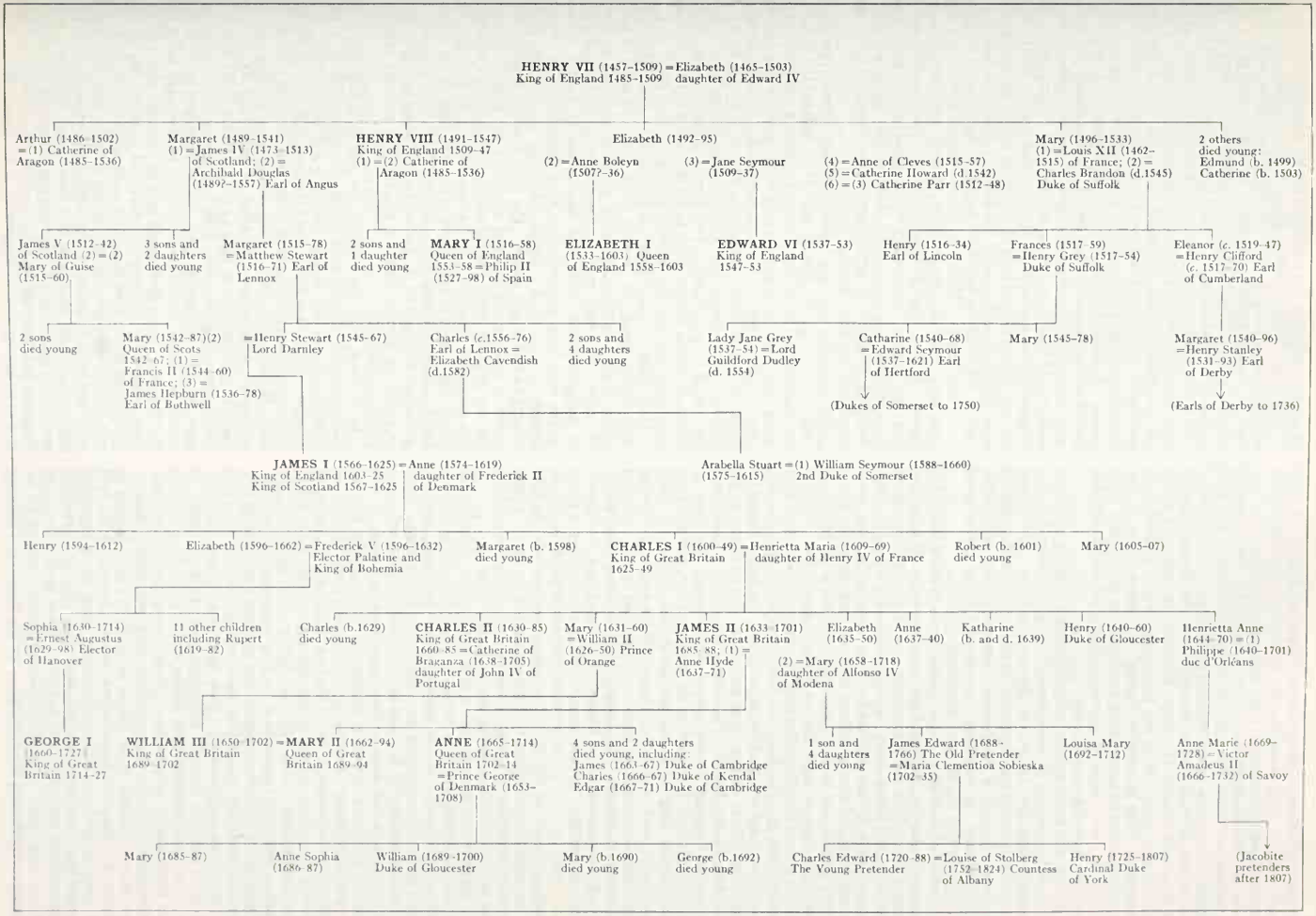
Developments in Trade

In the 13th century England already had an important stake in the trade of northern Europe, being an essential source of primary products, particularly of food and wool, for the urbanized and industrialized parts of the Low Countries and the lower Rhineland. Coal, tin and lead were also sent abroad, while alien merchants armed with foreign capital from the Low Countries, northern Italy and the Hanse towns were well entrenched in London, Boston, Lynn, York, Southampton and some smaller towns. Their particular concern was to develop the export trade in wool on which western Christendom largely depended for its supplies. Trade was then concentrated on the towns, which, although small by modern standards, had acquired a high degree of local self-government and strictly regulated their industry and trade through an elaborate system of guilds. In the 14th century, however, this situation began to change in a fundamental way. Native merchants slowly ousted the aliens from the control of the export trade in wool, both in England and on the continent, an achievement that was considerably helped by Edward III's manipulation of this trade for diplomatic and financial purposes in the early stages of the Hundred Years' War. Even more important was the fact that in the second half of the 14th century the cloth industry in England suddenly grew first to rival and finally to overshadow every other branch of English trade and industry. For this the French war provided the stimulus. The export trade in wool was often disrupted, Flemish weavers seeking asylum brought their skill to England and a heavy export tax on every sack of English wool gave the English cloth makers a large advantage over their continental competitors. Thus encouraged, the English industry shook itself free from the corporate towns with their guild restrictions, and established itself in the countryside. There district specialization developed in particular types of cloth, and the use of water power for fulling helped to keep costs down. The rapid growth of English cloth manufacture in the 15th century revolutionized the character of the bulk of the country's export trade now in the hands of English merchants. For the first time they had to venture far afield in search of customers for their cloth. By 1500 they had broken the monopoly of the Hanseatic league in northern and eastern Europe. The same need and the same spirit took their Elizabethan successors to America, to Africa and to India. (Gv. T.)

IX. THE TUDORS (1485–1603)

Henry VII (1485–1509)

The Wars of the Roses did not end on Bosworth Field, for a considerable number of Richard's supporters were not prepared, after all the turns of fortune that the previous 30 years had witnessed, to accept as final the verdict of a single battle. It is true that Richard's usurpation had driven many Yorkists into an alliance with the Lancastrians—an alliance that Henry VII strengthened by marrying Edward IV's elder daughter, Elizabeth (Jan. 1486), as soon as parliament had recognized him as king in his own right. Now Richard's death had left the rest without an obvious leader. His heir, the earl of Lincoln, had but a poor claim and a weak following. Clarence's son, the earl of Warwick, was Henry's prisoner and but ten years old. Edward IV's two sons were dead and though, as their fate was still a popular mystery, they might be impersonated, their impersonator must again be a mere boy of 12 or 15 years. Yet the cause was not hopeless. The dynastic quarrel of Lancaster and York had encouraged and been nourished by a multitude of lesser family rivalries. Admittedly, the wars and Edward IV's policy had weakened the more ambitious and overmighty of these families; even the Stanleys could not now make and unmake kings quite as the Nevilles had done. But so many had learned to take the law into their own hands, so many retained the influence and retinues to corrupt or coerce local government and justice, that dynastic faction might still hope to draw strength from local feuds. Those who longed for strong and settled government were probably far more numerous, but too often they were more concerned with profits than policies, more ready to cheer the victor than to risk life and fortune for a new king who had yet to prove his ability to hold what he had won. Thus, if the irreconcilable Yorkists could obtain enough aid from Henry's



(1) (2) Next to Name Indicates Order of Marriage

jealous neighbours to win an initial victory on English soil, they might rally to their cause not only the lawless and factious minority but also the war-weary and apathetic majority.

It was this that made the pretender Lambert Simnel so serious a menace to Henry VII. Simnel, put up by Lincoln and Lord Lovell to impersonate the young earl of Warwick (1487), obtained money and 2,000 German mercenaries from the dowager duchess of Burgundy, Edward IV's sister Margaret, an inveterate enemy of all Lancastrians. With this small force he won Ireland to his obedience. Then he invaded England and marched for 12 days through Lancashire before being crushed in the hard-fought battle of Stoke (June), where it is perhaps significant that barely half the royal levies came into action.

Henry, who seems to have underestimated his danger, was quick to learn his lesson. Henceforward all his policy showed a clear understanding that, to prevent the irreconcilable Yorkists from reversing the verdict of Bosworth, he must enrich and strengthen the monarchy until it could master domestic faction and lawlessness, and until his neighbours abandoned hope of embarrassing him by abetting the plots of his dynastic rivals.

Financial Recovery.—Even before Simnel's rising Henry had begun to tackle the problem of the crown's poverty. His first parliament (Nov.–Dec. 1485) passed the sweeping Act of Resumption which restored to the crown all lands alienated since 1455. Commissions of local men stiffened by royal councilors at once began to search out and take over these lands, to such effect that Henry during his first year on the throne received a clear profit of £13,633 from his estates, whereas Edward IV had received only £6,471 in the last year of his reign. The next big change followed close upon Simnel's defeat. The fines and forfeitures levied upon his supporters were not collected by the exchequer but by a hitherto insignificant official of the royal household, the treasurer of the chamber, Thomas Lovell. This private servant of the king could receive and pay out money upon a mere verbal order from his master. He kept his books in the simplest form, brought them daily for the king's approval, and could then explain any items quickly by word of mouth. These methods had such manifest advantages over the cumbrous procedures of the ancient exchequer that Henry soon began to use Lovell, and his successor, John Heron, even more extensively than Edward IV had used their predecessor, Thomas Vaughan. So the treasurership of the chamber, a minor court office, grew to be the greatest receiving department for the king's revenues, and the chief issuing department as well. By 1505 it was receiving more than £131,000 a year and paying out almost as much, while the exchequer accounted for no more than £45,000 and some of this was actually handled by the chamber.

This expansion brought with it a new machinery of audit—the general surveyors, royal councilors before whom particular collectors and spenders of the revenues "declared" their accounts, much as Lovell and Heron presented theirs to the king. Some councilors, notably Sir Reginald Bray and Robert Southwell, came to specialize in this work. Others gave more and more of their time to particular branches of financial business—Sir Richard Empson and Edmund Dudley (*qq.v.*) to taking and enforcing recognizances for payments due to the king; Richard Foxe, bishop of Durham, from 1497, to supervising prerogative rights in the north; Sir John Hussey, from 1503, to dealing with wardships; Sir Thomas Belknap, as surveyor of the king's prerogative from 1508, to controlling the various commissions of inquiry and dealing with the lands and rights which they reclaimed for the crown. These developments made possible a great tightening of royal control in every branch of financial administration. Largely thanks to this, the king's revenues from crown lands, feudal dues, customs and the profits of justice—his ordinary revenues as distinct from extraordinary income from parliamentary and clerical taxes—rose from £52,000 a year in 1485 to £142,000 in 1509. As early as 1490 he had become a lender rather than a borrower. After 1496 he could do without those parliamentary taxes which invariably yielded badly and sometimes provoked whole counties to rebellion (Yorkshire in 1489; Cornwall in 1497). When he died the crown was no longer poor. The king could "live of his own," on his ordinary revenues; and his wide estates gave him the local influence of a great landowner as a solid basis for his authority as king.

Conciliar Government.—Upon this solid basis Henry had also been developing effective instruments of justice and administration. Simnel's rebellion seems to have made him aware of the deficiencies of the king's council and common-law courts as well as of those of the exchequer. Within a few months of the battle of Stoke, his second parliament passed the act (1487) which a century later acquired the misleading title *pro camera stellata*. This act did not create the court of Star Chamber; but it did make possible the gradual evolution of that court out of the undifferentiated jurisdiction of the king in council. It gave statutory recognition to the legality of conciliar procedures, often challenged and restricted by statute during the Lancastrian period. It thus enabled the government to act against unruly or suspected persons without the formalities and hazards of indictment and trial by jury or the other safeguards that the common law provided for the liberty of the subject. It also offered hope of redress to suitors whose opponents were mighty enough to take the law into their own hands or to pervert the ordinary course of justice by coercing or corrupting juries and sheriffs. Moreover, the act placed upon half a dozen councilors a special responsibility to hear and determine such causes. At least three of these councilors (the chancellor and the two chief justices) were tied to Westminster by their other duties. Some councilors at least would therefore always be available to deal with these suits at regular times at the council's ancient meeting place in the Star Chamber (*q.v.*).

The next step was to get these facilities generally known and used. Accordingly the Justices act (1489) ordered the justices of the peace to read out at every quarter sessions a proclamation outlining their own duties and the procedure by which any aggrieved person might appeal to the king in council. In 1495, under the threat of a new Yorkist pretender, Perkin Warbeck (*q.v.*), the Perjury act laid down procedure for bringing corrupt juries before the council for punishment. Two other acts even allowed the local courts, in certain defined cases, to bring persons to trial upon mere information or suspicion, without indictment by a grand jury.

Meanwhile the organization of the council's own judicial work went further ahead. In the early 1490s it became usual to appoint a few of the councilors attending on the king to act as masters of requests for poor men's suits—a development or revival of an earlier practice, which also shows signs of French influence. At the same time, the councilors were ceasing to sit in chancery, leaving the chancellor there as sole judge for the ever-growing volume of civil equity causes. The same process of specialization which in financial matters was producing the general surveyors, the master of the wards and the surveyor of the king's prerogative was gradually separating the various aspects of conciliar jurisdiction which were to take shape in the next reign as the courts of Star Chamber, requests and chancery. Similarly, local and less specialized offshoots of the central council began to appear again, as under the Yorkist kings, in the troubled outlying regions of Wales and the north.

In all these ways the common law and local authorities were reinforced and brought more under the supervision of a king's council made effective in both authority and organization. Moreover the organization which fitted the council to control the kingdom also enabled the king more easily to control the council. For duties so many and various, many councilors of varying capacities and powers were needed—we know of more than 200 who were called king's councilors in Henry's time. Henry could thus safely employ even the great lords. They lent dignity to his council in its more formal manifestations, but few of them had to be included among his "more inward councilors" or be given key posts in his government. Those posts could now be filled by men who owed their influence not to lineage or estates but to ability and the king's favour, men drawn from the gentry and merchants, the clergy, lawyers and those upper servants of the nobility who now found a wider and more lucrative employment in the expanding service of the crown. Of the same type, on a lower plane, were those gentlemen who in each county formed the quorum among the justices of the peace and who were leading figures in the special commissions now so frequently used to supplement or investigate the normal ac-

tivities of local authorities. Around the reorganized council Henry thus mobilized the men of middling fortune who were rising to wealth and influence on the profits of land, law and business and who needed security against "lack of governance" and noble factions. With his own growing wealth and lands and the support of these men, he could make the royal power respected as it had seldom been before. He could fine the earl of Northumberland for ignoring royal rights of wardship and the earl of Oxford for disregarding the laws against livery and maintenance, the wearing of a badge or other insignia by all pledged to support one of the great barons in return for his promise of maintenance, or protection, the abuse of which practice led to the formation of lawless private armies as during the Wars of the Roses. He could even execute the great Sir William Stanley for dabbling in treason with Warbeck. The increasing use of conciliar justice, especially in Henry's later years, certainly encroached upon the liberties of Englishmen. Nevertheless, its drastic procedures were still used to reinforce, not to replace, the common law and still depended for their effect upon the tacit consent of subjects whom the king had little armed force to coerce.

Foreign Policy.—At first Henry's neighbours were more ready to intrigue with his Yorkist rivals than to seek his friendship. The Irish and Margaret of Burgundy aided Simnel. The Scots clamoured for Berwick. The French sent an army against Henry's old friend Francis of Brittany, and when Francis died (Sept. 1488) sought to compel the young duchess Anne to marry their king Charles VIII. This drew Henry into war. He could not risk his yet precarious popularity by letting England's old enemy annex Brittany, with its seafaring population and its ports so conveniently placed for harrying England's southern shores and southward trade. Yet his conduct of the war showed his determination not to jeopardize his work at home by overtaxing his strength abroad. When alliances with Maximilian, regent of the Netherlands, and with Ferdinand and Isabella of Spain (treaty of Medina del Campo, 1489) failed to check the French, he recognized defeat and promptly cut his losses. He did invade France in 1492, but he only began operations in October and had clearly come to bargain rather than to fight. As Charles was eager to be off on his Italian expedition, peace was made at Étaples in November. Brittany remained French; but Henry obtained an indemnity of £159,000.

This treaty came just in time to deprive Perkin Warbeck of French support; and the beginning of the Italian wars (1494) ensured that neither France nor Spain would wish to provoke Henry by dabbling in Yorkist plots. Maximilian and Margaret of Burgundy, however, wove round Warbeck a conspiracy involving James IV of Scotland, the earl of Kildare in Ireland and Sir William Stanley in England. This compelled Henry to a remarkable development of his foreign policy. First he struck at Maximilian and Margaret. He transferred the English cloth staple from Bruges to Calais and forbade all exports to the Low Countries (1493). It is doubtful how much effect these measures had; but by Feb. 1496 the Italian situation induced the Netherlands government to abandon Warbeck, to restrain Margaret's plottings and to restore trade upon terms highly favourable to the English (the Magnus Intercursus; see *Economic Policy*, below). By then, too, Maximilian, now Holy Roman emperor, had handed over the Netherlands to his son Philip, who in 1496 married Joan, elder daughter of Ferdinand and Isabella. This closer linking of Burgundy to Spain meant that Henry could hope for Spanish help if France should turn back from Italy and seek to make another Brittany of Flanders, a hope strengthened in 1501 by the marriage of Arthur, prince of Wales, to the younger daughter of Ferdinand and Isabella, Catherine of Aragon.

By that time Henry had also drawn Ireland to obedience and Scotland to friendship. Sir Edward Poynings had induced the Irish parliament in 1494–95 to acknowledge its subordination to the English and its inability to meet or legislate without the previous consent of the king and council in England. Poynings' other measures for establishing an effective English control over Ireland proved too expensive to pursue. But at least the Irish were too cowed to help Warbeck, who in 1495 failed almost as miserably before Waterford as he had off Deal.

In Scotland James IV's efforts, too, were handicapped by the party which Henry built up among the Scottish nobles. There was some fighting on the border and Henry had to prepare a large army in 1497. The taxation for this provoked the Cornish rebellion, but though the rebels marched right across England to Blackheath, they had been crushed before Warbeck arrived to join them. He did raise another feeble revolt and attack Exeter, but was speedily rounded up. Two years later (1499) he and the unfortunate Warwick were both executed and Yorkist hopes were virtually extinguished. Thereupon James IV, isolated in a world grown eager for Henry's friendship, made peace and in 1503 sealed it by marrying Henry's elder daughter, Margaret.

Henry's system—friendship with France and Scotland, reinforced by alliance with Spain and Burgundy—was now complete. In 1502, however, Arthur's death broke the matrimonial link with Spain. Then in 1504 the death of Isabella of Castile threatened to separate the Spanish kingdoms which her marriage to Ferdinand of Aragon had united. Castile became a source of discord between Ferdinand, father of her heiress Joan, and Philip of Burgundy, Joan's husband. Henry could not side against the ruler of the Netherlands, which lay so close to his southeastern shores and provided the chief market for English cloth. Nor could he allow Ferdinand to take Castile into the French camp, as he took Aragon by marrying Germaine de Foix (1505). He thus had to connive at, even finance, Philip's assertion of Joan's claims, though he used the opportunity to wring new privileges for English merchants in the Netherlands (the Malus Intercursus, 1506). Philip, however, died suddenly in Sept. 1506, leaving the Netherlands to his six-year-old son, Charles, while Ferdinand resumed the government of Castile for the now quite demented Joan. Henry thereupon arranged for his younger daughter, Mary, to marry Charles, with Castile as their marriage portion (Dec. 1507). And in the hope of wresting it more surely and speedily from Ferdinand, he even proposed to marry the mad Joan himself. But before either marriage could be made, Maximilian, Charles, Ferdinand and Louis XII of France sank their differences in the League of Cambrai, designed to partition Venice (Dec. 1508). Henry could thus again be friends with all his neighbours. As long as the league lasted all its members were eager for his good will and he needed no special alliances to insure against a possible return of French ambition toward Flanders. Before it broke up in 1511, Henry VII was dead (April 22, 1509).

Economic Policy.—Henry, as Francis Bacon wrote, "could not endure to see trade sick," and it was his good fortune to rule at a time when English commerce was finding exceptional opportunities for expansion. The Merchant Adventurers had found in Antwerp a most satisfactory channel through which to supply the growing demand for English woolen cloth in central Europe, and thus the astonishing expansion of the Antwerp market was matched by a similar growth in England's one great export industry. This benefited not only the king's revenues and the merchant class but also those landowners whose flocks provided the wool for the expanding cloth manufacturing industry. One of Henry's prime purposes therefore was to confirm and extend the privileges of English merchants in the Low Countries. The Magnus Intercursus guaranteed them against any increase in tolls and dues. The 1506 agreement freed them from many imposts and even allowed them to sell their cloth retail, as well as wholesale, in all the provinces except Flanders. This Malus Intercursus, as the Netherlanders called it, went too far to be fully exploited and in 1507 Henry abandoned some of its gains to further his matrimonial projects. Nevertheless, in the Netherlands, as in Spain after the treaty of Medina del Campo and in France after the treaties of 1492 and 1497, the position of English merchants was much improved.

Henry also tried to get as much English trade as possible carried in English ships. Bounties on shipbuilding were supplemented by the acts of 1486 and 1490, which forbade the export of English goods in foreign ships when English ships were available, or the import of Gascon wines in foreign bottoms. For the same end he consistently interpreted the chartered privileges of the Hanseatic Merchants of the Steelyard in the strictest sense, insisting for example that their right to bring in "their goods" did not cover

goods grown or made outside their own narrow territories. He even began to challenge the Hansards on their home ground. A treaty with Denmark (1490) opened the Iceland trade and fisheries to Englishmen and gave them the chance to regain their recently lost foothold in Scandinavia. A treaty with Riga (1499) for a moment opened the way to the eastern Baltic. In the end the long struggle ended indecisively, for in 1504 Henry confirmed the Hanse's privileges in order to close their towns to two possible Yorkist pretenders, Edmund and Richard de la Pole, brothers of the earl of Lincoln who had supported Lambert Simnel and who had fallen at Stoke in 1487. His encouragement of English merchants to break into the Venetian monopoly of Mediterranean trade was more successful. When the republic put prohibitive duties on English trade to Crete (1488), he made a treaty with Venice's rival Florence, which set up an English wool market at Pisa and limited the Venetians' imports to the bare quantities required for their domestic use. When Venice retaliated he put prohibitive duties upon wines brought in their ships. In the end they relaxed their tariffs, but the English kept theirs, together with their staple at Pisa. Henry also joined, in a modest way, in the attempts to get behind the Venetian monopoly of trade with the orient. John Cabot in his 1497 voyage was, like Columbus in 1492, looking for Cathay when he stumbled upon America; Henry had encouraged him and his Bristol backers by promising them ownership of any heathen lands they might discover.

It is true that across the Atlantic, as in the Baltic and Mediterranean, the king's reach exceeded his subjects' grasp. This, with Wolsey's war taxation, the diversion of capital to land speculation during the Reformation and the excessive attraction of the Antwerp market, retarded these wider developments for another half century. But when English maritime expansion resumed its course under Elizabeth I, the trails that it followed were those which Henry VII had blazed.

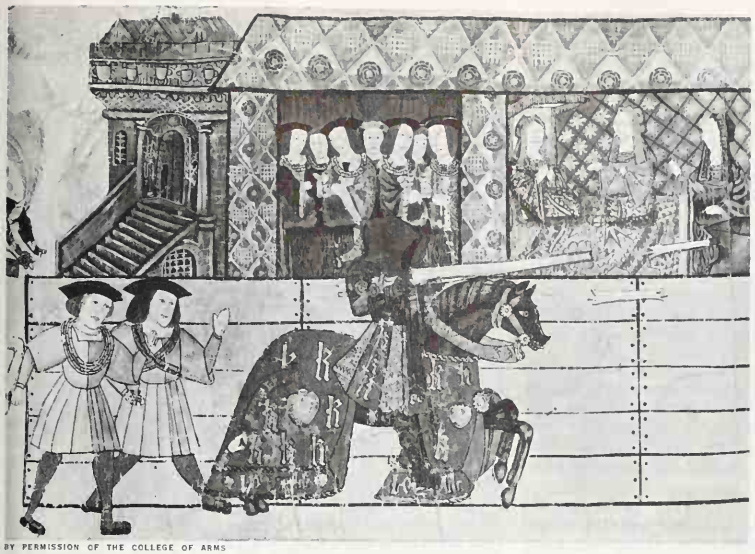
In internal affairs Henry also foreshadowed the more mature economic policy of his granddaughter. He introduced no new forms of economic organization, such as the great Elizabethan monopolist companies; nor did he attempt to create new industries. Yet, though most of his work followed medieval precedents, it considerably extended the field of governmental control over economic activity. Acts forbidding the export of bullion reasserted royal control over the currency. An act of 1504 asserted the supremacy of the common law and central government over guilds and corporations by subjecting their ordinances to review by the chancellor, treasurer and chief justices. Other measures sought to prevent or mitigate the evil effects of the rapid growth of the cloth trade. The wealthier Merchant Adventurers were prevented from squeezing out smaller traders by exacting a £20 entrance fee to their fellowship, though their monopoly of exporting cloth to the Netherlands was confirmed. In 1489 an act was passed to check the enclosures of arable land for sheep pastures, which were leaving a bandful of shepherds on lands which had before provided homes, food and work for a host of husbandmen. This act again extended the field of government interference, though the lack of adequate means to enforce it and the hostility of the rural gentry made the control largely nominal. Finally, acts of 1495 and 1504 did something to discipline the growing numbers of able-bodied unemployed, men whom the great lords could no longer "retain" or whom economic change was driving into roaming predatory bands of "sturdy beggars." The 1504 act, which first imposed upon the justices of the peace the job of dealing with these vagrants, was the first step toward a national system of supervising and relieving the poor and workless which culminated in the acts of 1598 and 1601. Here, no less than in his financial reforms, his development of conciliar government and his foreign policy, Henry VII was indeed the founder of the Tudor monarchy.

Henry VIII (1509–47)

The accession of Henry VIII (April 22, 1509) was hailed by Erasmus and by most Englishmen as the dawn of a golden age. Certainly for a few years a more liberal spirit crept into government. Empson and Dudley were executed for their extortions (1510), the severity of conciliar justice was relaxed and some of its organs perhaps ceased to function (the masters of requests and the councilors in the north). The exchequer was allowed to challenge acquittances given by the general surveyors and to reassert its financial control. Parliament had five sessions in 5 years, as against only one in the preceding 12. Nevertheless, this liberalism was as much the result of the new king's distaste for drudgery as of his benevolence. He was too young, too interested in all the pleasures of mind and body, to toil like his father at the details of government. He was too adept at martial exercises and too avid for glory to be long content with a cautious policy. So when the League of Cambrai broke up and the pope, Ferdinand and Maximilian appealed to him to join their Holy league against France, the temptation to emulate Henry V proved too strong. He had already linked himself to Ferdinand by marrying Catherine of Aragon (1509), who since Arthur's death had been kept in England, together with her dowry, almost as a hostage for her father's good behaviour. Henry was also encouraged to war by his almoner, Thomas Wolsey (*q.v.*), who saw in the Holy league a chance to oust the pacific counselors of Henry VII and to win favour with both king and pope. The war began badly. The marquess of Dorset's expedition to Guienne (1512) ended in mutiny and failure. The Scots soon came in on the French side. But in 1513 Henry routed a French force at the skirmish known as the battle of the Spurs, in northern France, and Surrey destroyed the Scots army at Flodden, killing their king and the flower of their nobility. In 1514 peace with France and Scotland, the marriage of Henry's sister Mary to Louis XII and the regency of his other sister, Margaret, in Scotland restored England to a position outwardly commanding and raised Wolsey to supremacy in Henry VIII's counsels.

The Rise of Wolsey.—Wolsey's rise spelled the end of the brief liberal interlude in Tudor government. A return to chamber finance was the first sign. Once again the chamber became the great spending department, issuing £296,564 in 1512 and £699,714 in 1513. In 1512 the general surveyors were recognized by statute as auditors for its accountants and in 1514 were given the beginnings of judicial authority over them. This recognition of the general surveyors as a court, however imperfect, as a great public department like the exchequer, was the first step toward replacing the direct personal control of the king by bureaucratic, conciliar machinery. A similar change came over the chamber itself. Its treasurer's accounts, which Henry VII had inspected daily, were audited less and less frequently by Henry VIII and Wolsey: an order of 1519 laid down once a month as an ideal. By 1529, when this task was delegated to the general surveyors, the chamber had practically lost direct contact with the king and become in effect the receipt and issue branch of the court of general surveyors.

This tendency for conciliar machinery to replace the king's personal action in the routine affairs of government spread more widely after Wolsey became chancellor (Dec. 1515). Wolsey's control of domestic administration and justice was at first almost as close and personal as Henry VII's. The retirement of William Warham, archbishop of Canterbury, Bishop Foxe and Lovell, and the timidity of the earl of Surrey, the lord treasurer (restored as duke of Norfolk after Flodden), left him without a rival. He controlled the great seal himself; the privy seal through his henchman Bishop Thomas Ruthal; and the signet through a succession of men promoted from his own service to be king's secretaries. He thus controlled all the instruments through which king or council could issue formal orders. This monopoly of authority, however, depended almost entirely upon the king's favour. Wolsey therefore took great pains to prevent possible rivals from gaining the king's ear. Few councilors, and these unimportant, attended Henry when he was away from Westminster, where Wolsey was so much tied by his duties as chancellor. More and more council business was drawn to Westminster, to the Star Chamber meetings over which the chancellor (in the king's absence) presided or to the chancery where he sat as sole judge. And the councilors at Westminster became Wolsey's agents rather than his equals; little groups of civil servants and judges, each tending to specialize in a particular branch of business, rather than makers of policy or influential counselors of the king.



BY PERMISSION OF THE COLLEGE OF ARMS

HENRY VIII JOUSTING BEFORE CATHERINE OF ARAGON; FROM THE WESTMINSTER TOURNAMENT ROLL; 16TH CENTURY. IN THE COLLEGE OF ARMS

The Conciliar Courts.—So when pressure of foreign affairs after 1518, and still more after 1525, caused Wolsey to delegate much of his domestic work, the various groups of councilors to whom he delegated it tended to go out of the council and form separate courts and departments, while the chancellor monopolized the controlling and advisory functions. His servant Christopher Middleton, to whom he delegated piracy cases in 1517 and all admiralty cases in 1525, became virtually judge in a separate high court of admiralty. The committees to whom he delegated poor men's suits in 1518 had by 1529 already taken shape as a separate court of requests. From 1524 a committee of six councilors met regularly to arrange the work and publish the decrees of the council in the Star Chamber, another step toward its differentiation as a court. In 1525 the councils of the north and the Welsh marches reappeared, staffed almost exclusively by Wolsey's servants. In 1529 even the judicial work of chancery was delegated to 20 commissioners. The judicial activity of the king's council was thus losing its extraordinary and intermittent character and developing into a hierarchy of conciliar courts alongside the courts of common law.

Wolsey both multiplied this judicial activity and greatly extended its field. He used it to teach the nobles obedience to the laws against livery and maintenance, riot and the other lawless practices to which they had begun to return during the slack years after Henry VII's death. He interpreted widely the council's right to intervene when the ordinary courts had no remedy or were prevented by force or fraud from applying one. Riot or actions tending to riot, maintenance or perjury, fraud or forgery, provided ample excuse for suitors to appeal to the council or for the chancellor to evoke cases before it or into chancery. And in dealing with, for example, an alleged riot arising out of a dispute over the ownership of a field or manor, it was easy also to decide the title to the property, although the subject's property (like his life and limb) were in strict law outside the council's jurisdiction. The conciliar courts were thus beginning to usurp the functions and draw away the suitors of the ordinary courts. Conciliar law threatened to replace common law as the ordinary law of the land. England seemed about to join its neighbours in a reception of the Roman civil law, upon whose supple procedures and authoritarian principles the judicial action of the king's council was so closely modeled.

At the same time the council's increasing control over local authorities and Wolsey's lavish use of special commissions for administration and inquiry threatened to undermine the independence of local institutions. Also, between his appointment as chancellor in 1515 and his fall in 1529, parliament met but once. Council, local authorities and parliament were ceasing to be checks upon the will of the king or his minister. Moreover, from 1518 Wolsey was, under the pope, as supreme in the English church as he was, under the king, in the English state. Successive commissions as legate *a latere* gave him precedence over Warham, the archbishop of Canterbury, and delegated to him a great part of the papal power in England. This he used with the same vigour, to bring about an unprecedented concentration of authority, arbitrary in form and absolute in character. Indeed, the king was already becoming, through his chancellor-legate, "over all persons and all causes as well temporal as spiritual supreme" well before the Act of Supremacy asserted such sovereignty. The Tudor monarchy was beginning to turn away from medieval ideas of contract and partnership between ruler and ruled, toward the forms of conciliar absolutism which, during the next two centuries, were to triumph almost universally on the continent.

Foreign Policy 1514–25.—That the Tudor monarchy did not continue to develop as Wolsey was shaping it was largely the result of the foreign policy which he and his master pursued. The keys to that policy are probably to be found in two problems which began to appear after 1514—finance and the succession. The king's ordinary revenues would not pay for a warlike policy. The 1512–14 war, costing about £1,300,000, had used up Henry VII's savings and shown again that taxes, even when voted by parliament, would be paid "unwillingly, with extreme complaint." Experiments with new levies, based upon more realistic assessments than that of 1332 which still served for the taxes known as the fifteenth and the tenth, had failed. The 1514 subsidy of £160,000 had yielded only £48,000. For the next few years therefore Henry VIII and Cardinal Wolsey played only a modest part in European politics.

Then the scene changed. Charles of Burgundy inherited the Spanish dominions from one grandfather, Ferdinand of Aragon, in 1516. In 1519 he succeeded the other, Maximilian, as Holy Roman emperor and entered upon a lifelong rivalry with the young Francis I of France. Charles's sprawling lands now encircled and constricted France on almost every side. But Francis, holding Milan, fruit of his earlier victory at Marignano (1515), cut one of Charles's lines of communication between Spain and the other Habsburg lands. He could also hope, if England acquiesced or assisted, to cut the other, up the channel. Hence both sides were ready to bid high for English good will. The temptation was too great for Wolsey and Henry to resist. The emperor's influence over the Spanish and German cardinals (and Italian too, once the French were driven from Italy) might secure Wolsey the papacy. His aid might enable Henry to regain the French lands, even the French crown, which the Lancastrians had won and lost a century before. It might also solve the problem of the English succession which, now that all Catherine's other children had died in early infancy and even her miscarriages had ceased, rested in the infant princess Mary (born 1516). Not since Henry I's day had England

faced the prospect of a queen regnant; and the 19 years' civil war which Matilda had then brought were not a comforting precedent for a generation barely escaped from the Wars of the Roses. It was becoming imperative to find the child Mary a grown-up husband, with enough power and authority to keep the realm in peace. Nobody it seemed was better qualified for this than Catherine's nephew, the emperor Charles V. Indeed, if the reign of the house of Valois in France were destroyed, he and Mary might on Henry's death attain "the whole monarchy of Christendom" and give it the unity to repulse the mounting power of the infidel Turks under Suleiman the Magnificent.

At all events Henry and Wolsey, despite a show of friendship to Francis at the Field of Cloth of Gold (*q.v.*; 1520), joined Charles in war against France in 1521 as soon as he had crushed the revolt of the Spanish *comuneros*. Charles V, however, was too occupied in north Italy to give much help, even in 1523 when the greatest of the French nobles, Charles, constable of Bourbon, turned traitor. English invasions of France achieved nothing in 1522 and 1523; the duke of Albany for a time revived French influence in Scotland; and Charles twice broke his promise to secure the papacy for Wolsey. Moreover, the 1523 parliament would only grant, with slow reluctance, half the £800,000 which Wolsey demanded—only enough to repay the forced loan levied for the 1522 and 1523 campaigns. Another campaign in 1524 was thus impossible and secret peace negotiations began.

Then suddenly Charles's forces destroyed the French army and captured Francis at Pavia (Feb. 1525). "The whole monarchy of Christendom" seemed almost within his grasp. Henry at once proposed a joint offensive to complete the destruction of France and, despite the 1523 rebuff, demanded an "amicable grant" of £500,000 from his lay subjects and a large levy from the clergy. However, popular resistance to the grant soon rose so near to general rebellion that Henry had to abandon it. Charles, too, not only rejected the English proposals but also broke his engagement to Mary and married the rich princess Isabella, daughter of Manuel I of Portugal, leaving Henry to find other means of solving the problem of the English succession.

The Divorce Question.—Henry's immediate answer was to give his one illegitimate son the precedence and offices of an heir to the throne, with the title of duke of Richmond. However, a six-year-old illegitimate son was hardly better than a nine-year-old legitimate daughter and Henry soon turned to a more drastic solution. He convinced himself that his marriage to Catherine, his deceased brother's widow, was null and void. Had not this marriage, within the forbidden degrees of affinity, required an all but unprecedented use of the pope's dispensing power? Did not the death of all but one of Catherine's children show that the pope could not override the curse of Lev. xx, 21? The king therefore resolved—all the more after falling in love with one of the court ladies, Anne Boleyn—that what Pope Julius II had done, Pope Clement VII should undo. Popes had granted like favours on flimsy enough pretexts to his sister Margaret and to Louis XII of France. Surely Clement would not refuse the king of England, who had championed the papal cause both in arms and in his pamphlet against Luther, which had earned him in 1521 the title of defender of the faith?

But Clement was timid and after Pavia, and the imperial army's sack of Rome (1527), he was in the power of Catherine's nephew, the emperor, whose support he needed against the Lutherans in Germany. A project to rush past him a formal verdict from the English archbishops failed because Catherine was able to forewarn him through Charles (1526). The desperate expedient of an offensive alliance with France, to prize the pope out of Charles's clutches, also failed with the destruction of Marshal Odet Lautrec's army in the autumn of 1528. In July 1529 Clement made the treaty of Barcelona with Charles, recalled his legate Lorenzo Campeggio, whom he had sent to England to join Wolsey in hearing Henry's divorce suit, and evoked the cause to Rome. In August the French made their peace with the emperor at Cambrai. Wolsey's desperate diplomacy had failed and on Oct. 17 he was dismissed.

The Break With Rome.—Henry now adopted a new policy, of putting pressure on the pope by making threats against papal jurisdiction over and revenues from the church in England. Parliament was called in November to give the highest national sanction to the royal measures. It was a ready instrument for such a purpose. The bishops and abbots, who formed a majority in the lords, were not the men to stand up for the pope against the king. Nearly all owed their sees, which so rarely saw them, to long service in the royal courts and diplomacy. Few remembered with any pleasure Wolsey's use of papal authority as legate and all felt that the church was too much under fire from the laity to risk a quarrel with the crown. For even noblemen as Roman Catholic as Lord Darcy, later implicated in the Pilgrimage of Grace, were saying that the clergy were "not meet to govern us nor other temporal men." Others were beginning to hint that spiritual and temporal authority might be united in the king's hands just as easily as in a legate's. The commons, composed mostly of the more substantial landowners and wealthier townsmen, professed themselves good Roman Catholics. Lollardry had long lost its hold upon the wealthier classes; Lutheranism had as yet penetrated to only a handful in the universities and seaports. But the commons coveted the church's broad acres and were coming to regard its social and economic teaching as outmoded. The spread of lay education and literacy, too, made them less patient of the church's tutelage and more critical of the clergy's competence. Men bred in the new grammar schools and colleges founded by the practical piety of the last few generations and able to read the books flowing from the new printing presses were already demanding what the Puritans later called a learned and preaching ministry. Such men had little respect for prelates who might, like Foxe, hold four bishoprics in succession and visit only two of them; for the monks' fugitive and cloistered virtue; or for parish priests, of whom barely one in four possessed a degree and some could hardly make sense of the paternoster.

This anticlericalism, shown in the commons' debates about limiting benefit of clergy in 1512 and 1514 and in 1515 over the case of Richard Hunne, who was found hanged while imprisoned by the bishop of London on a charge of heresy, had been further inflamed by the unpopularity of the cardinal's administration. As John Fisher, bishop of Rochester, said in 1529, "with the commons is nothing but 'down with the church.' " They passed bills to limit mortuary dues and probate fees and to abolish (with fairly numerous exceptions) pluralism and nonresidence among the clergy. Under this pressure, and the threat of being held guilty of praemunire for having recognized Wolsey's legatine authority, the clergy in convocation reluctantly acknowledged the king as their "only and supreme lord and, as far as the law of Christ allows, even supreme head" (Feb. 1531).

This, however, and the collection (suggested by Thomas Cranmer) from the universities of Europe of opinions favourable to the king's divorce, failed to move the pope. So in 1532, after the commons' Supplication Against the Ordinaries (March) had attacked not only the church's courts but also its right to legislate without the king's consent, the clergy in convocation was browbeaten into promising to enact no new canons or ordinances without royal licence and to submit those already existing to the approval of a committee, half clerical, half lay, nominated by the king. This Submission of the Clergy (May), which destroyed the legislative independence of the church, was accompanied by measures aimed more directly at the pope. Parliament empowered the king to stop the payment of annates (*q.v.*). Then, early in 1533, another act forbade all appeals to Rome and declared that as England was an empire governed by one supreme head and king, all spiritual causes should be "finally and definitively adjudged and determined within the king's jurisdiction and authority."

These measures still did not move Clement to grant the divorce; but he did consent to Cranmer's consecration as Warham's successor (March 1533). The new archbishop hastened to pronounce Henry's marriage to Catherine null and void (May). Thereupon, Anne Boleyn, whom Henry had already married secretly in January, was crowned queen and on Sept. 7 gave birth to (ironically enough) a daughter, the future Queen Elizabeth I. In the first months of 1534 the breach with Rome was completed by a series

of acts which finally forbade annates, Peter's pence and all other payments to the pope; ended papal jurisdiction over the church in England; gave statutory form to the Submission of the Clergy; and vested the succession to the throne in Henry's children by Anne, a measure which his subjects were to swear to observe.

Henry's Reformation.—In 1534–40, the years of the ascendancy of Thomas Cromwell (*q.v.*) in Henry's counsels, the new order was forced upon the English people by a mixture of terror, propaganda and political bribery. Commissioners enforced the oath to the new succession and against the pope. The Act of Supremacy confirmed the king's headship of the church and another act made it treason to attack his new title even in speech (Nov. 1534). Under this act Sir Thomas More, the greatest of the English humanists and but lately chancellor, was executed (July 1535), as were Bishop Fisher and a number of Carthusians who also refused the new oath (June). Spies and informers flourished. Cromwell even drafted schemes for a standing army, for trusted gentlemen in each county to be sworn of the council and empowered to arrest any speaking for the pope, and for a court of six "conservators of the commonweal" to deal summarily with any statutory offenses committed since 1485.

These schemes came to nothing. But as vicar-general (Jan. 1535) Cromwell did issue injunctions (Aug. 1536) which seemed to promise a thorough reform of abuses among the parish clergy. His visitors swept through Oxford and Cambridge, rooting out scholastic and canon law studies in favour of lectures in Greek, Latin, Hebrew and the civil law. At the same time scholars, writers and printers were set to work to defend and explain the royal supremacy and the clergy were enjoined to preach it. All this, and the inflow of Lutheran ideas, so stimulated theological controversy that Henry, with the assent of the clergy in convocation, issued the Ten Articles (July 1536). These defined the faith on the more disputed points in a generally Catholic sense but with some omissions favourable to Lutheran doctrines. Henry and Cromwell thus played not only on the laity's fears but also upon their jealousy of clerical power and their desire for clerical reform. Moreover, a hurried bullying visitation of the monasteries by Cromwell's agents in July 1535 was followed by an act dissolving all those houses whose annual income was less than £200 (March 1536). The act touched only these smaller houses and it transferred their property to the crown. But the tone of the commons' debate boded ill for the rest and the crowd of suitors besieging Cromwell for grants and leases showed the laity's expectation of a share in the king's spoils. In such circumstances resistance to royal policy was slow to mature. When it did come to a head in the Pilgrimage of Grace (Oct. 1536), it did not spread much beyond the backward and thinly peopled northern counties of Lincoln and Yorkshire. Beginning as a movement of the clergy and "rude commons" under Robert Aske (*q.v.*), it eventually won uneasy support from substantial men like Lord Darcy. But their fundamental loyalty to the crown so enfeebled their actions and divided their counsels that the king was able to crush it by a combination of force and deceit (Feb. 1537).

This conservative protest had perhaps some effect. The Bishops' Book (July 1537) "found again" the four Catholic sacraments omitted from the Ten Articles. Two years later the Act of Six Articles imposed, with stringent penalties, the main tenets of orthodox Roman Catholic dogma and practice. This, however, only checked the rate of progress; it did not reverse what had been done. During 1538–39 the greater monasteries went the way of the lesser and property worth about £140,000 a year, almost half the church's annual income, passed into the king's hands and through them to those of his lay subjects, creating an almost impregnable vested interest in the break from Rome. In addition, in 1538 the order that the Bible in English (in a translation derived from the Protestants William Tyndale and Miles Coverdale) should be kept and read in every parish church gave to the reform movement "a power of permanent appeal to the whole population" (H. A. L. Fisher, *Political History of England*, vol. v).

Foreign Relations.—So far the continued rivalry of Charles and Francis, Lutheran successes in Germany and Scandinavia and the Turkish onrush under Suleiman the Magnificent had kept Catholic Europe too preoccupied to interfere with Henry. In 1536 indeed, just as Catherine's death and Anne's execution removed the original occasions of the emperor's resentment, Charles and Francis went to war again. Thus no foreign aid came to the Pilgrimage of Grace. In 1537 the pope tried to rouse a crusade against Henry, but his legate *a latere* for England, Reginald Pole (*q.v.*), could not even gain admittance to French or imperial territory.

In 1538–40 there was a flurry of alarm when Charles and Francis became outward friends (June 1538). Henry then began a costly program of building castles, mustering the shire levies and preparing ships; and Pole's relations, the marquis of Exeter, Lord Montagu and Sir Edward Neville, were executed for treason. Cromwell even persuaded the king—a widower again since Jane Seymour had died in Oct. 1537 giving birth to his long-sought son, the future Edward VI—to draw closer to the German Protestants by marrying Anne of Cleves. The danger, however, soon passed and Cromwell paid for his misjudgment with his head (July 1540), while Anne placidly accepted an annulment of her marriage. Within another year Charles and Francis were again quarreling and Henry felt safe enough in England to turn to British problems.

He had already tackled their Welsh aspect by obtaining wide statutory powers for the council in the marches (1534) and by dividing the marcher lordships into shires incorporated into the ordinary English system of justice and administration (1536). Now in 1541 he took the title of king, instead of lord, of Ireland. There, too, papal authority was abolished and a policy adopted of turning Irish chieftains into English peers and knights. But Scotland was the biggest problem. While Henry was breaking with Rome, James V was trying to master the Scottish nobility by allying himself with David Cardinal Beaton and the church. Also he had married Francis' daughter Madeleine in 1537 and, on her death a year later, Mary of Lorraine (Mary of Guise). By 1542 England and Scotland were at war and, after the disgraceful Scottish rout at Solway Moss had driven James to his grave, Henry persuaded the Scots to agree to a marriage between his son Edward and James's infant daughter, Mary Stuart (July 1543).

To complete the work and sever the Scots from continental allies, he also joined Charles V in war against France. This secured him Boulogne in 1544. Then after French attempts at invasion had failed (1545), Francis agreed to abandon the Scots, while Henry promised to sell back Boulogne in eight years' time for 800,000 crowns. Meanwhile, however, Mary of Guise and Beaton had been able to rouse the Scots, who were angered by Henry's revival of the English claim to feudal suzerainty. They denounced the 1543 treaty and, despite Lord Hertford's savage invasions of the Lowlands (1544, 1545) and Beaton's assassination (May 1546), were still in arms when, on Jan. 28, 1547, Henry VIII died.

Government—the Council.—Henry's death ended a 62-year period of strong personal rule by adult kings. With his successor, the nine-year-old Edward VI, control of the government passed to the king's council. The past 20 years had, however, done something to prepare that body for this test. Wolsey's fall had brought a revival of that inner council attendant on the king which the cardinal had so jealously subordinated to the general council with himself at Westminster. Henry's Reformation, creating a mass of complicated business too great for one man to supervise personally, had turned the former into a privy council, a score or so of leading ministers to whom Henry delegated an ever-widening range of affairs. By Aug. 1540 it had acquired a clerk of its own, thus marking off its meetings from those in the Star Chamber; and Cromwell's development of the office of king's secretary provided an effective and flexible executive agent for domestic administration and foreign affairs.

A sharp distinction had also begun to emerge between these privy councilors and the other, ordinary, councilors. The masters of requests and the councilors in the north and in the Welsh marches now seldom concerned themselves with any but their own special tasks. The same was true of those who ran the new financial courts of augmentations (created in 1536 to deal with monastic lands and monks' pensions), first fruits and tenths (1536), and

wards and liveries (created in 1540 to manage the feudal revenues). The chancellor, too, no longer aspired to be "prime minister"; he became more entirely a legal official now that the secretary was taking over much of his administrative work. Even the Star Chamber was tending to become just a high court of the leading conciliar and common law judges, reinforced by a few privy councilors and peers. The privy councilors were thus to some extent freed from the routine work of justice and administration and given more time for the general control of government and policy.

Unfortunately, however, factions had developed among them. Henry had generally kept these balanced and under control. His execution of Cromwell and marriage to Catherine Howard (July) did not bring an outright triumph for the conservatives under the duke of Norfolk and Bishop Stephen Gardiner. Nor did Catherine Howard's misconduct and execution (Feb. 1542) and Henry's marriage to Catherine Parr (July 1543) bring victory to the group under Cranmer, Hertford and Lord Lisle, who wanted to push the Reformation further. In Henry's very last days young Surrey's indiscretions did cause the downfall of his father Norfolk and the disgrace of Gardiner. This left the advanced party dominant in the council of 16 which Henry appointed for Edward. But no council could carry the same authority as a grown-up king—certainly not a council controlled by one of the religious factions which were beginning to divide the nation.

Government—Crown and Parliament.—Moreover, the events of the previous 20 years had affected the foundations of royal authority itself, even though this was hardly apparent for another two or three generations. Despite the spoliation of the church, the crown had not regained its financial independence. The defense preparations and wars of 1538–47 had used up the spoils of the monasteries and the profits, almost equally great, from a wholesale debasement of the currency (1544–47). The new government, with a Scots war on its hands and Catholic Europe hostile, had to depend on parliament for money even more than Henry had done in his later years. And those years had given parliaments a new sense of their own importance. The wide field covered by their legislation and its revolutionary character had begun to produce a new idea of law, not as custom to be elucidated but as the untrammeled will of a sovereign. This sovereign, however, was not just the king but the king in parliament. Legislation was not the Roman law's *quod principi placuit* (what the sovereign wills) but statutes enacted by lords and commons and assented to by the king.

Closely connected with this was the revival of the common law, and the check to the encroachments of conciliar law upon it during the chancellorships of Wolsey's less ambitious successors, Sir Thomas More and Lord Audley. Indeed, so much of conciliar development after 1529 had been sanctioned by act of parliament—the new commission to the council in the Welsh marches; the creation of the courts of augmentations, first fruits, wards and liveries; the Statute of Proclamations (1539)—that the prerogative courts themselves were being subjected in some sense to statute law. The Renaissance conception of absolute sovereignty was thus blending with the older English view of the king as beneath a law to which both he and his subjects consented in parliament. The theory of the Tudor monarchy was beginning to diverge sharply from that of continental absolutism.

The Social Revolution.—Such ideas were as yet only struggling to birth among the classes who provided the members of parliament. But those classes had already been made something more than merely theoretical legislative partners of the king. By exploiting their own holdings and the cloth trade's growing demand for wool, they had gained the money to buy up or rent much of the land forced from the feudal nobility's hands by the Wars of the Roses and Tudor attainders. Loyal service to the Tudor kings, at court and locally, had brought not only more lands but also greatly increased responsibilities for, and so influence in, county affairs. And suddenly, on top of all this, the crown's expenses in the last decade had forced it to grant or sell to them about two-thirds of the confiscated monastic lands. The rising middle class was thus fast becoming a new aristocracy. They were as yet far from being an aristocracy of great families able to dominate the crown as the Nevilles had done and the Whig lords were to do. But already they were masters of their counties and spreading their influence over the boroughs, where the great majority of members of parliament were drawn from the county gentry. They were in no sense an organized party; but they were beginning to form an interest that governments could not ignore, an interest widely linked by intermarriage and business connections and by the habit of lesser families looking to a few great court families to secure them offices of profit and social esteem under the crown.

There was, however, a dark side to this social revolution. The growth of larger holdings in fewer hands (whether those holdings were enclosed or still scattered in strips over the open fields) had for a century or more been creating a class of landless rural labourers. The expansion of the cloth manufacture beyond the control of the urban guilds into the countryside and small towns was creating a similar class of industrial labourers. This proletariat of almost propertyless workers was not sharply divided into industrial and agricultural elements; nor was it massed in great cities and factories, though here and there "factories" were appearing, such as William Stumpe's at Malmesbury or John Winchcombe's at Newbury. The weavers mostly worked at piece rates in their village homes on yarn and looms "put out" to them by the clothiers. Much of the spinning was done in similar fashion by farm workers' families. Nevertheless, they were becoming more dependent for their livelihood upon profit-seeking employers. In the last decade of Henry VIII's reign this evolution was greatly accelerated. Almost half the vast lands of the church then passed into the hands of lay owners determined to make the most of their holdings. With the sharp debasement of the currency, prices soared at home and landlords met the rise by raising their tenants' rents and fines. English exports were cheapened abroad and clothiers kept up the real value of their incomes by increasing output (and often lowering quality). In 1547 customs were being paid at London on twice as many shortcloths as in 1509, and on half as many again as in 1537. More workers thus became dependent upon the clothiers and the cloth trade; more landlords and tenants enclosed land for sheep runs, so depriving more villagers of their holdings and rights of common. And in this price revolution, soon to be aggravated by the world-wide inflation which followed upon a vast inflow of precious metals from the new world, the poorer workers had the least power of making their claims heard. Their earnings were to increase, but not at anything like the rate of their employers' profits; and the increase always lagged behind the rise in prices. Moreover, their employment was coming to depend more upon the fortunes of the export trade—a trade now too much concentrated upon a single commodity, cloth, and a single market, Antwerp. This greatly increased the economic ascendancy of London and the Merchant Adventurers, through whom most of this trade went. But it meant large-scale industrial unemployment whenever, as in 1528, war stopped the Netherlands trade or when, as by 1550, the Antwerp market became saturated.

Agrarian and industrial changes were thus by 1547 creating problems which could shake the foundations of law and order, even without the additional upheavals already wrought in traditional religious habits or the ferment of new ideas which this had loosed.

Edward VI (1547–53)

The social and economic changes of the time made the king's minority a most severe test for the council's capacity to govern. The council, indeed, so doubted its own capacity that it at once appointed Edward's uncle, Hertford, as protector with almost kingly powers. He was then created duke of Somerset (*see* Somerset, Edward Seymour, Duke of), while Lisle became earl of Warwick.

Somerset.—The protector laid aside the apparatus of despotism with which Henry and Cromwell had equipped the monarchy. All heresy laws, and almost all treason laws except that of Edward III, were repealed. Free publication and discussion of the Scriptures was allowed, communion in both kinds permitted, the Roman

BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

CORONATION PROCESSION OF EDWARD VI PASSING CHEAPSIDE CROSS, 1547. 19TH-CENTURY PRINT FROM A CONTEMPORARY PAINTING, NOW DESTROYED

Catholic interpretation of the Mass made optional and English introduced into a good deal of the services. Foreign reformers, such as Martin Bucer and Peter Martyr, flocked to England and the tumult of religious debate rose so high that early in 1549 Somerset passed through parliament an Act of Uniformity, imposing upon the church its first Book of Common Prayer in English.

While this religious policy was alarming conservative opinion—the prayer book brought a rebellion in Devon and Cornwall in June 1549—the rest of Somerset's policy was stirring even more dangerous antagonisms. He tried to persuade the Scots to marry their queen to Edward VI and so to unite the two realms in one empire of Great Britain, "having the sea for a wall, mutual love for a defense." Yet there was little love in his methods. After routing the Scots at Pinkie (1547), he occupied most of the Lowlands. Thereupon the Scots sent Mary to France to marry the dauphin, while the French sent troops to Scotland and declared war on England (Aug. 1549). To meet the expenses of all this, Somerset again debased the coinage, sending prices to double what they had been in 1547. He also tried, in the 1548 Subsidy act, to tax sheep, wool and cloth. This was part of a wider policy, based upon the ideas of Bishop Hugh Latimer, John Hales and the so-called commonweal party. It was designed to check the too rapid expansion of cloth manufacture and was reinforced by a great commission to repress enclosures. Such a policy stirred hopes in the peasantry as dangerous as the fears it inspired in the gentry and merchants. Thus the summer of 1549 saw risings all over the country, which in Norfolk under Robert Ket reached the proportions of a major rebellion. Somerset's failure to deal firmly with these disorders ruined his credit, already shaken by the intrigues and execution of his brother, Thomas Seymour. In Oct. 1549 he was swept from power by a coalition between Warwick and the leader of the more Catholic-minded councilors, the earl of Southampton.

Northumberland.—The coalition was short-lived, for Warwick (*see* NORTHUMBERLAND, JOHN DUDLEY, Duke of) meant to monopolize power, not to share it. Once the rebellions had been crushed and the propertied classes reassured by acts making it felony to combine against enclosures or rising rents and corn prices, he drove Southampton from the council (Feb. 1550) and allied himself with the more advanced religious reformers. Stephen Gardiner, Edmund Bonner, Nicholas Heath, George Day and Cuthbert Tunstal were deprived of their sees. The clergy were permitted to marry; altars were replaced by communion tables; organs, painted windows and the old Latin service books were removed from the churches. Somerset, readmitted to the council (April 1550), strove to slow down the pace of change. But Warwick packed the council and won the young king, who made him duke of Northumberland (Oct. 1551). In Jan. 1552 Somerset, charged with a bogus plot, was executed. Thereupon, a second Act of Uniformity and Book of Common Prayer, which could not like the first be interpreted in a Roman Catholic sense, imposed a closer conformity to Protestant doctrine and ceremonial (1552). (*See* also REFORMATION.)

Northumberland also, once he had got peace abroad by ceding Boulogne and abandoning Scotland to the French (March 1550), introduced measures of equal novelty to restore the country's economy. He began to reform the coinage (1551) and sent Sir Thomas Gresham to improve English credit at Antwerp. By abolishing the Hansards' privileges in 1552 he further concentrated the Netherlands trade, as Gresham desired, in the Merchant Adventurers' hands, where the government could more easily regulate it—and, as the 1552 loan of £40,000 showed, raise money from it. He sought to check overproduction and improve quality by an act requiring proper apprenticeship for clothworkers. He also used old Sebastian Cabot's knowledge in a search for new markets to lessen England's excessive dependence on Antwerp. A group of Londoners began to invade the Portuguese monopoly of trade with Morocco and the Gold Coast. Another group, organized on the novel principle of a joint stock company, received by patent a monopoly of discovering northeast and northwest passages to the orient. Their first achievement was Richard Chancellor's opening of trade with Russia through Archangel (1553). Indeed, many of the measures which restored health to England's economy under Elizabeth were planned under Northumberland.

Their earlier effect, however, was nullified by Northumberland's grasping ambition. Even his Protestant allies found him an embarrassing bedfellow, less interested in reform than in despoiling the church to enrich himself and reward his henchmen. Finally, he aroused the hatred of the whole nation when, with Edward's health obviously failing, he plotted to secure the succession for his own family. In May 1553 he married one of his sons, Guildford Dudley, to Lady Jane Grey, eldest granddaughter of Henry VIII's younger sister, Mary. Then he persuaded Edward to leave the crown by will to Jane and her heirs male. To this illegal disregard of the superior claims of Henry's daughters, Mary and Elizabeth, and even of Jane's mother, the duchess of Suffolk, he compelled councilors and judges to subscribe. But when Edward died (July 6, 1553) and Jane was proclaimed queen, the nation soon showed its determination not to exchange Tudors for Dudleys.

Mary escaped and rallied East Anglia. When the duke set out against her, his forces melted away and the councilors in London proclaimed Mary as soon as his back was turned. On July 20 Northumberland surrendered, crowning his ignoble career by professing himself a Roman Catholic upon the scaffold, possibly in the vain hope of being pardoned.

MARY I (1553–58)

Mary, the devoted daughter of Catherine of Aragon, naturally sought to restore not only Roman Catholicism but also papal authority in England. Long out of touch with English politics, she leaned upon her cousin, the Roman emperor Charles V, and welcomed proposals for her marriage to his son, Philip of Spain. Such projects, however, soon revealed the strength of those new forces that Henry VIII had built up and Somerset and Northumberland had released. Mary's first parliament (1553) readily enough annulled Catherine's divorce and repealed all Edward's religious legislation. But it showed no disposition to go further; and the announcement of the Spanish marriage treaty provoked in the rebellion led by Sir Thomas Wyat (*q.v.*; Jan.–Feb. 1554) what was perhaps the most dangerous rising that any Tudor ruler ever faced. Wyat's slowness and Mary's courage saved her throne and she was able to go through with her marriage (July). Her third parliament (Nov. 1554–Jan. 1555) repealed all Henry's antipapal laws and admitted the pope's legate, Reginald Pole, to receive the realm back into the Roman fold. Yet even now it laid down by statute that the holders of former church lands should keep what they held and not be sued about them except at common law.

After this hollow repentance, parliament agreed to revive the older statutes empowering the clergy to take action against heretics. Pole (archbishop of Canterbury, 1556), Edmund Bonner and one or two other bishops at once set to work. Between Feb.

1555 and Nov. 1558 nearly 300 men and women were burned at the stake for heresy. Nearly all of them came from London, Essex, Hertfordshire, East Anglia, Kent and Sussex. The vast majority were humble folk, though Thomas Cranmer, Nicholas Ridley, Hugh Latimer, George Hooper and Robert Ferrar were among them. The nobility and gentry mostly bowed before the storm, though some joined the organized emigration which was to preserve the Protestant cause for better days. The story of this holocaust eventually "stamped on the English mind a hatred, unthinking, ferocious, and almost indelible, of Rome and all its belongings" (A. F. Pollard, *Political History of England*, vol. vi).

The impression was deepened by Mary's failure in almost every other field. The lord treasurer, the marquess of Winchester, did improve the financial administration by bringing all the new financial courts, except the court of wards, under the exchequer's control. Efforts were also made to bridle the cloth industry by increased export duties (1557) and by trying to limit each country weaver to two looms and two apprentices (Weavers' act, 1555). Yet Mary did nothing to prevent her husband from closing the Spanish and Portuguese colonies to English traders, while she joined him in his war against France (1557), a war which, despite Philip's victories at St. Quentin (1557) and Gravelines (1558), brought England only the humiliating loss of Calais (Jan. 1558). The marriage was, indeed, in every way a failure.

Mary was subject to frequent illnesses and, though she several times believed herself to be pregnant, she remained childless. She could not persuade parliament to allow Philip's coronation or to weaken Elizabeth's right to the succession and by the time she died (Nov. 17, 1558) she had turned England not only against Rome but also against Spain.

Elizabeth I (1558–1603)

Elizabeth's accession restored to Tudor government the firm royal direction which it had lacked since the death of Henry VIII in 1547. The privy council, cut down to half the unwieldy numbers it had reached under Mary, became again the crown's efficient servant. Its leading member, the secretary, William Cecil (*see* Burghley, William Cecil, Baron), was soon to prove himself one of the greatest of all Tudor ministers. Thus England was nursed back to unity and strength. The first essential, peace with France and Scotland, was purchased by the reluctant cession of Calais in April 1559. At the same time acts of supremacy and uniformity restored the second prayer book and, with a few exceptions, the religious settlement of 1552. In this, parliament, spurred on by the returning Protestant exiles, probably went further than Elizabeth deemed wise or the nation wanted. The acts were, however, slackly enforced. The queen's chief concern was that the new bishops—for all but one of the bench established under Mary I refused to conform and were replaced—and the increasingly Protestant lower clergy should not stir dissension or prevent the quiet growth of unity and conformity.

A similar moderation marked Elizabeth's foreign policy. While Spain was kept at a friendly arm's length, a daring and well-timed military intervention assisted John Knox and the Protestant faction to seize control of the Scottish government and expel the French finally from that back door into England (1559–60). Moreover, the French were soon fully occupied by their own wars of religion, in the first of which Elizabeth unsuccessfully intervened on the Huguenot side (1562–64).

Economic Policy.—Meanwhile Cecil had been tackling the economic problem. Great companies, some joint stock, some regulated, were given by patent the monopoly and direction of particular trades or manufactures. A new charter (1564) gave the Merchant Adventurers fuller powers over trade not only with the Netherlands but also with Germany, where treaties with Emden (1564) and Hamburg (1567) opened alternative "vents" to Antwerp. Another charter to the Russia company (1566) encouraged it to extend its trade through Muscovy to Persia. Queen and councilors invested in John Hawkins' ventures (1563–68) to west Africa and the West Indies, where he sought to gain Spanish permission to trade in return for policing the Caribbean against Huguenot and other intruders. At home the coinage was at last effectively reformed (1560–61) and Sir Thomas Gresham's royal exchange was founded (1567). New industries, necessary to defense or self-sufficiency, were developed, often again by patents of monopoly such as those for copper and brass, salt, and glass. Masters and skilled workmen, fleeing from Philip II's religious persecution in the Netherlands, were encouraged to bring over their "new draperies" and other crafts. A remarkable expansion of coal mining and the coal trade reflected the beginnings of this early industrial revolution. Government control was further increased by some tightening of the London customs administration, while the Statute of Apprentices (1563) strove to check the excessive expansion and poor quality of clothmaking. It restricted weaving to sons of master weavers or substantial men; required their apprenticeship for seven years; protected journeymen against cheap apprentice labour; and directed other labourers into agriculture. Its enforcement and the regulation of wages according to prices were put into the hands of the justices of the peace, watched over by chancery and the assize justices. A new Poor law (1563) also made the justices and the parish responsible for those unable to work or find work, and for punishing those who would not work. These measures by no means cured the evil effects of rapid economic change; but at least Elizabeth's government neither threw up its hands in helpless *laissez faire* nor shouted, like Wolsey and Somerset, hopelessly against the rising tide.

BY COURTESY OF HIS GRACE THE DUKE OF BEDFORD

ELIZABETH I. "ARMADA" PORTRAIT BY MARCUS GHEERAETS; 16TH CENTURY. IN WOBURN ABBEY

The Revolt of the North.—The stability of the country, however, was not yet assured. Elizabeth, despite her affection for Robert Dudley (Northumberland's son; he was made earl of Leicester in 1564) and her flirtations with the Austrian archduke Charles, remained unmarried. Her nearest legal heir was the Roman Catholic Mary Stuart. Thus, should Elizabeth die (and she nearly did in 1562) another anti-Protestant reaction might be attempted. Mary had returned to Scotland after the death of her husband Francis II of France (Dec. 1560) and for several years her charm and moderation rallied a growing support in both countries for her claim to be recognized as Elizabeth's successor. In 1565 she strengthened her claim and her connection with the English Roman Catholics by marrying Lord Darnley, another of Margaret Tudor's grandchildren. In 1566 the birth of her son James, a potential king of Great Britain, might have given her the whip hand had she not quarreled with her husband. As it was, the next two years brought Darnley's murder (1567), Mary's marriage to the earl of Bothwell and, in 1568, her final defeat in Scotland and flight to England.

This almost coincided with a dangerous quarrel between England and Spain. There had been trouble in 1563–64 over Spanish efforts to curtail English trading privileges in the Netherlands. In 1567 the duque de Alba's arrival with the main Spanish army to restore order in those lands after the 1566 rebellion was stra-

tegically alarming as well as commercially harmful to England. Moreover, Hawkins' efforts to open trade with Spanish America had ended in the destruction of his ships at San Juan de Ulúa (Sept. 1568). Then, in Dec. 1568, Elizabeth "borrowed" a Genoese loan which was on its way by sea to pay Alba's troops.

Philip and Alba replied by seizing all English ships in their ports and stopping all trade with England. This quarrel, following close upon Mary's arrival in the north, gave supporters of the old order a last chance to check the tide of change. It provoked a wide conspiracy to drive Cecil from office, make peace with Spain and settle the succession on Mary after marrying her to the duke of Norfolk, Elizabeth's second cousin. Cecil, however, retained Elizabeth's confidence and soon broke up the motley coalition against him. Only the more desperate men of the far north, led by the earls of Westmorland and Northumberland, went forward into open rebellion. Their attempt (Oct.–Dec. 1569) and a belated rising by Leonard Dacre (Feb. 1570) were soon crushed and about 800 executions finally cowed the north.

This was the turning point of the reign. The failure of their protest doomed the Roman Catholics either to accommodate themselves to the new order or to desperate reliance on plots and foreign aid. The first such plot, that of Roberto Ridolfi to marry Mary to the duke of Norfolk and place her on the throne, brought Norfolk and almost brought Mary herself to the scaffold in 1572. But foreign help was less easy to obtain, even though the pope excommunicated Elizabeth in 1570. Philip II could not risk war. He had no ocean-going navy; his finances were strained by the Lepanto campaign and the Netherlands rebellion; and Alba needed all his army to guard against the growing threat of French intervention on behalf of the rebels. This threat also alarmed Elizabeth. Unable to join Spain to avert it, she sought to control French action by co-operating with them. She made a defensive alliance with them in April 1572 and talked of marrying Henry of Anjou or Francis of Alençon, brothers of Charles IX. English and French volunteers poured into the Netherlands after the Sea-Beggars seized Brielle in April 1572 and Louis of Nassau took Mons in May, and Elizabeth tried to use the English to keep the coastal areas and islands out of French control. However, the St. Bartholomew massacre, with the death of the Huguenot leader Gaspard de Coligny, who had persuaded Charles IX into his anti-Spanish policy, freed both England and Spain from anxiety on this score. Spain, preoccupied with William of Orange's continuing revolt in Holland and Zeeland, thereupon made peace with Elizabeth (1573). The great crisis of 1568–72 thus ended by confirming the stability of Elizabeth's settlement at home and making manifest her independence abroad.

Puritans and Catholics.—Meanwhile the Puritan movement, to purge the church of "Romish dregs" and give it a "learned and preaching ministry," was coming into sharp conflict with the queen's policy of blurring religious differences. In 1566 Matthew Parker, the archbishop of Canterbury, began to discipline Puritan clergy who would not wear surplices, make the cross in baptism and so forth. He and helpers like John Whitgift dealt also with the Puritans at the universities. Laurence Humphrey, president of Magdalen college, Oxford, was called to order in 1564 and Thomas Cartwright lost his fellowship at Trinity college, Cambridge, in 1571. Cartwright by that time was turning to attack episcopacy itself and demanding a Presbyterian organization of synods and elected ministers and elders. In 1571–72 his lay supporters even introduced bills into parliament for reforming the church. This was a novelty of considerable constitutional importance, for it threatened to take the initiative in legislation out of the hands of queen and councilors. It thus challenged the royal supremacy in the state as well as in the church. Elizabeth therefore forbade the introduction of bills on matters of state or religion without government consent. This ruling her parliaments in the main accepted, despite the efforts of such men as Peter Wentworth to persuade them that it infringed their privilege of free speech, and further Puritan attempts in 1576, 1587 and 1593 met with no great success. The Presbyterians therefore began to build up cells within the church on the basis of the classes or meetings of the more zealous clergy and laymen for spiritual "exercises" and "prophesyings." In 1577 the queen suspended Parker's successor, Edmund Grindal, for refusing to suppress this movement, but it was not until John Whitgift succeeded Grindal (1583) that the episcopal courts and the court of high commission really got to work. Their methods then "savoured too much of the Romish Inquisition" even for Cecil, and they probably did more to strengthen than to destroy Puritanism in its wider and vaguer forms. But, helped by the violence of the Marprelate tracts attacking the Anglican clergy in 1588–89 (*see* MARPRELATE CONTROVERSY), the steady improvement in the quality and education of the clergy and the laity's growing attachment to the Anglican services, they do seem to have broken the Presbyterian movement.

Both Anglicanism and Puritanism were stimulated by a revival of Catholic activity. This began in 1574 with the first missionary visits of priests trained overseas in William Allen's seminary at Douai, later reinforced by priests from the English college at Rome and, after Robert Parsons' and Edmund Campion's mission in 1580, by the Jesuits. These brought upon the English Catholics the drastic penal laws of 1581, 1585 and 1593. More than 200 were executed as traitors in the 28 years after 1574; heavy fines were imposed for saying or hearing Mass; and in 1581 the fine for not attending the Anglican service was raised from a shilling a Sunday to £20 a month. This persecution, though spasmodically enforced, goaded the more impatient into plots to assassinate Elizabeth and, with Spanish help, put Mary on the throne. Sir Francis Walsingham (*q.v.*), however, who had succeeded Cecil (now Baron Burghley and lord treasurer) as secretary in 1573, ferreted out the schemes of Anthony Tyrrell (1581), Nicholas Throckmorton (1583), John Somerville and Edward Arden (1583), William Parry (1584) and Anthony Babington (1586). This last plot brought Mary herself to the scaffold in Feb. 1587. After her death and the defeat of the Spanish Armada in 1588 the plots died away, for few would kill Elizabeth to make James of Scotland or Philip of Spain king of England. But already in the minds of patriotic Englishmen Roman Catholicism was fatally associated with treason and Spanish domination. Even among the Catholic priests themselves a growing party began to advocate coming to terms with the English Jezebel.

The Expansion of England.—A remarkable maritime and commercial development followed the crisis of 1568–72. After the closing of Antwerp in 1568, the Merchant Adventurers sought other approaches to their German markets. Hanseatic jealousy drove them from Hamburg to Emden (1580) and then to Stade (1587). Yet probably these ports, with a partial reopening of Antwerp and Bruges (1573–82) and new connections with rebel Holland and Zeeland, absorbed most of the exports which Antwerp alone had formerly handled. Merchants were also encouraged by charters of incorporation and monopoly to open markets further afield. By 1588, 100 English ships were annually visiting the Baltic, most of them belonging to members of the Eastland company. The older Russia company prospered. To the southward, trade with Spain was reopened (1574) and a Spanish company formed. The early 1570s saw a revival of trade into the Mediterranean and the Levant; and the early 1580s saw the incorporation of the Venice and Levant companies. Other merchants, later formed into Barbary and African companies, established regular trade to Morocco and pushed their ventures further down the African coast. All this encouraged industrial development at home and increased England's economic independence and sea power, for such vessels as the 350- or 500-ton Levanters were capable of fighting alongside the queen's ships.

The government also encouraged voyages by Martin Frobisher (1576–78) and John Davis (1585–87) in search of a northwest passage; and abortive schemes by Humphrey Gilbert, Walter Raleigh and Richard Grenville to plant an English colony on the North American coast. Nor could it prevent John Oxenham and others from trying to follow up Francis Drake's 1572 attack upon the Spanish main, though such ventures were discouraged for some time after 1573. English privateering, indeed, stimulated by commissions from the Huguenot and Dutch leaders, was fast spreading from the channel across the Atlantic and into the Caribbean. In

1577–80 it received its greatest fillip from the astonishing success of Drake's voyage around the world. Drake's ostensible purpose was to carry out Grenville's project of discovering, first, the supposed great south Pacific continent of Terra Australis and then a western entrance to the northwest passage. But his most spectacular achievements were to plunder the defenseless Pacific coast of Spanish America and capture a rich Manila galleon. By allowing this voyage and knighting Drake on his return, with a profit of 4,700%, Elizabeth went far toward turning the private warfare of the English sailors into a national conflict with Spain. For, with the Netherlands now a great financial liability instead of a great asset, Philip II was becoming increasingly dependent upon the steady flow of American silver across the Atlantic to Cádiz.

The Spanish Armada.—It was, however, English interference in the Netherlands, provoked again by fear of France, that brought the quarrel to a head. Elizabeth's alarm at the designs of Henry III's brother and heir presumptive, Francis of Anjou (formerly Alençon), led her first to subsidize a German army under John Casimir of the Palatinate (1578) and then to revive the 1572 proposal for her own marriage to Anjou. This Anjou courtship gave her some control over French action but cost her dear in subsidies and naturally aroused Philip's hostility. Nor did it stop his lieutenant, the duke of Parma, from reconquering the southern Netherlands. Moreover, when Anjou's death (May 1584) left the Huguenot Henry of Navarre as heir presumptive, France again collapsed into civil war. The Dutch seemed doomed, with their leader William of Orange assassinated (July 1584) and Parma moving up to their last line of defense along the Maas and Rhine. Elizabeth dared not leave them to a hostile Spain. So, while refusing their sovereignty, she agreed (Aug. 1585) to send Leicester with 6,000 men to aid them. At the same time she loosed Drake against the West Indies and sent money to the Huguenots.

By such pressure Elizabeth probably hoped to bring Spain again to terms. But Philip's long patience now snapped. Also his conquest of Portugal (1580) had given him an ocean-going navy and by 1585 he had built this up enough to think of challenging England. So Parma, whose first attempt to turn the Dutch river defenses was foiled by Leicester in 1586, was kept half idle in Flanders while the Invincible Armada gathered at Cádiz and Lisbon. Delayed by Drake's "singeing of the king of Spain's beard" (his operations against Cádiz and off the Portuguese coast in 1587), the Armada (*q.v.*) eventually sailed in July 1588.

The English fleet under Lord Howard of Effingham and Drake, almost equal in numbers and superior in armament and maneuverability, did not find it an easy prey. After a week's indecisive battling up the English channel, they scared the Armada from Calais roads with fire ships and, but for an unlucky turn of the wind and a shortage of powder after the past week's unprecedented cannonade, might have driven it to disaster on the banks of Flanders. As it was, the badly battered Spaniards fled northward in confusion. Many ships perished on the western shores of Scotland and Ireland and barely half straggled home to Spain.

Foreign Policy After 1588.—The Spanish war lasted another 15 years, partly because England's counterstrokes miscarried, partly because it had not the men or money to attack Spain effectively and at the same time bolster up the French and Dutch. Had Sir John Norris and Drake in 1589 carried out their instructions to destroy the surviving half of the Armada as it lay helpless in Santander and San Sebastian, Hawkins' scheme of keeping squadrons off the Azores to intercept the American silver ships might have brought Spain to its knees. But Norris and Drake spent their strength vainly against La Coruña and Lisbon; and, within a month of their return, the assassination of Henry III (July) concentrated all attention on the struggle in France between the Huguenot Henry IV and the Catholic league.

Twice Parma's intervention saved the league and, with a Spanish army also in Brittany, threatened to bring the whole channel coast under Spain's control (1590 and 1591–92). To avert this Elizabeth sent across 4,000 men under Lord Willoughby in 1589; another 4,000, repeatedly reinforced, to Brittany, 1591–95; 4,000 more, reinforced by nearly as many again, under Essex to Rouen in 1591–92. Not until his conversion to Catholicism (1593) did Henry begin to master his realm.

Meanwhile, Elizabeth still kept 6,000 men, costing about £126,000 a year, in the Netherlands. There fortune had turned since the early setbacks of the earl of Leicester's quarrel with the states, the betrayal by Sir William Stanley and Rowland Yorke of Deventer and the Zutphen forts (1587), and the loss of Sluys and Geertruidenberg. By 1595 Maurice of Nassau and Sir Francis Vere had gained for the Dutch frontiers which almost assured their independence. All this had greatly restricted England's effort at sea. Philip was thus able to rebuild his navy and strengthen the Indies' defenses, as the destruction of Grenville and the "Revenge" (1591) and the failure of the last expedition of Drake and Hawkins (1595) showed. The new armadas, which alarmed England in 1596–97 and 1599, never arrived. But while the English sacked Cádiz, the Spaniards seized Calais (1596) and, after Essex's unlucky voyage to the Azores, where the Spanish treasure ships escaped him, the French made peace (1598) and the Irish rebellion, led by the earl of Tyrone, again diverted most of England's resources to land warfare. Essex failed to put down the rebellion; the Spaniards sent a force to aid it and Elizabeth was dead before Tyrone finally submitted to Lord Mountjoy (March 1603).

Internal Discontent.—Before this, Essex himself had met a traitor's end. After Leicester's death (Sept. 1588), he had succeeded to the position of prime royal favourite. But, unlike Leicester, he aimed at greater things. He became the patron of sailors, soldiers and diplomatic agents and strove to make himself the sole channel to royal favour. In sharp opposition to Burghley and his son Robert Cecil (who became secretary in 1596), he advocated a more vigorous war policy than they or Elizabeth judged England's resources and interests to warrant. The Cecils, in self-defense, built up their already considerable following in the government, country and even parliament. The growing habit of looking to a few leading court families to secure offices and favour from the crown thus narrowed dangerously toward a duel between Essex and the Cecils, especially after Burghley's death (1598) and Essex' failure in Ireland. The queen in the end had to call a halt to Essex' importunity. In desperation he tried a *coup d'état*. It failed and he was executed (1601).

By then the burden of the war had roused a wider and ultimately more dangerous spirit of criticism. According to a 1603 estimate, the wars had cost about £4,000,000 since 1585. Parliamentary taxation rose from a yearly average of less than £35,000 for 1576–84 to more than £112,000 for 1599–1602. The counties and boroughs also spent heavily on musters and coast guard, in ship money and on levying and equipping about 100,000 men for the continent and Ireland. Yet the queen's savings of £300,000 from 1573 to 1584 had gone; more than £800,000 worth of crown lands had been alienated; patents of monopoly were sold lavishly; royal rights of purveyance and pre-emption were exploited to their limits; feudal dues more severely assessed; and the crown was still £440,000 in debt. All this occurred, too, when the wars were closing markets in France, Spain and the Netherlands, and when unemployment was making necessary the Poor laws of 1598 and 1601 with their parish poor rates. Already in the 1597 parliament discontent had appeared and in that of 1601 it broke out in a great monopolies debate. It seems as if only the queen's timely promise to redress abuses then prevented the introduction of a bill to restrict her prerogative in these matters. Certainly the debates showed that parliament had learned from the Puritans the idea of seizing the initiative in legislation, and that tactful management would be needed to keep within bounds its growing interest in "matters of state." At the same time lawyers were beginning to harass the lesser conciliar courts. The Star Chamber had not yet lost its popularity, but Sir Edward Coke and his brethren were challenging the jurisdiction of requests, admiralty and even high commission. The clouds were gathering round the Tudor sunset.

X. THE EARLY STUARTS AND THE COMMONWEALTH (1603–60)

James I (1603–25)

When James VI of Scotland succeeded to the English throne on Elizabeth's death (on March 24, 1603), his problems were still not

insoluble and there were many things in his favour. Scotland became England's partner, completing an insularity which English sea power made almost inviolable. With Ireland subdued and peace soon made with Spain (1604), the crippling war expenses ceased and continental markets reopened. The successful first voyages of the East India company, and the planting of colonies in Virginia (1607) and the Bermudas (1611), gave an added fillip to trade, while the plantation of Ulster (1610) gave another outlet to investors and settlers. But James missed his chances. He alienated the Puritans, whose demands in their Millenary petition were very moderate, by his threats at the Hampton Court conference (1604) and his support of the bishops' measures to enforce conformity (*see* also ENGLAND, CHURCH OF: *Elizabethan Church Settlement and the 17th-Century Conservatives*). His vacillation over the penal laws goaded a few wilder Roman Catholics into the Gunpowder plot (1605), which made all his later efforts at toleration suspect to a nervously Protestant nation. By lecturing parliament on the divine right of kings he provoked the commons into asserting, in their 1604 Apology, that "the voice of the people, in the things of their knowledge, is said to be as the voice of God." By leaving only two or three councilors in the house he let the initiative slip into the hands of critics like Sir Edwin Sandys, who passed bills to abolish purveyance (*q.v.*) and to end the great companies' monopoly of foreign trade, petitioned for concessions to the Puritans and severity against the Roman Catholics and frustrated his sensible plans for a closer union with Scotland.

James also aroused suspicion by using the law courts, as Sir D. L. Keir has said, "to declare, as already existing law, what parliament was unwilling to enact." Among James's proposals had been the naturalization of all persons born under the king's obedience in all places under his dominion at the time of their birth. When parliament rejected this proposition, two suits, one in chancery and another in the king's bench, were brought in the name of a Scots child, Robert Calvin. The judgment in Calvin's case (1608), naturalizing in England all Scots born since the union of the crowns (1603), threatened parliament's legislative prerogatives. Another court judgment, in Bate's case (1606), in which an English merchant was condemned in the court of exchequer for refusing to pay an extra duty, imposed by the crown, on imported currants, was used to justify new impositions, worth £60,000 a year, on overseas trade. (*See* BATE, JOHN.) The commons, knowing their influence, and even their summons, to depend upon the king's need of their grants, took alarm. James's extravagance which, helped by still-rising prices, had in five years of peace doubled Elizabeth's war debt, did not encourage them to open their eyes to his very real need for a larger revenue. In 1610, however, they did offer to accept the new impositions, provided any further increase were forbidden by statute. In the Great Contract they also offered £200,000 a year for the abolition of feudal dues and purveyance (which were worth £115,000). James was probably unwise to refuse. Certainly his conduct after the dissolution of parliament in 1611 made another such chance unlikely. In the next ten years he called only the brief Addled parliament (1614) and after Cecil's death in 1612 he fell under the sway of the incompetent Scots upstart Robert Carr, earl of Somerset. Court scandals culminated in the poisoning of Sir Thomas Overbury by Carr's mistress.

Extravagance, corruption, traffic in titles, sale of wardships, multiplication of monopolies widened discontent. William Cockayne's specious project to pay the king £300,000 a year for a monopoly of exporting finished cloth and the prohibition of its export unfinished (1614) nearly ruined the cloth trade. Exports slumped; clothworkers were unemployed; Dutch and Flemings began to make cloth for themselves; and in 1617 the scheme had to be abandoned. By that time the fall of Carr and the rise of George Villiers, later duke of Buckingham (*q.v.*), had brought some vigour, if no great wisdom or probity, into the administration and Sir Lionel Cranfield had managed by economy and expedients to balance revenue and expenditure. But the episode was no advertisement for James's claim to a kingly understanding of policy beyond the reach of a subject's brain. Nor had his dismissal (1616) of his independent chief justice, Sir Edward Coke (*q.v.*)—"a lesson to be remembered and feared of all that sit in judicial places," the lord chancellor told Coke's successor—done anything to allay the fears which his earlier use of the judges had aroused.

James's Foreign Policy.—The outbreak of the Thirty Years' War brought new discords. Earlier, James had played his part with France and the Dutch in opposing the Habsburgs and the German Catholic league on the question of succession in the duchies of Cleves and Jülich (1609–14) and in thwarting Spanish designs against Venice and Savoy. He had even in 1613 married his daughter Elizabeth to Frederick, the elector palatine, leader of the German Protestant union. He did not approve when Frederick rashly accepted the Bohemian throne, from which Protestant rebels had ejected the Habsburg emperor Ferdinand II in 1618. Yet when the league drove Frederick from Bohemia and the Spaniards invaded the Palatinate in 1620, James felt bound to intervene. Like most Englishmen, he believed that Spain held the key to the situation. But, while most Englishmen wanted to save Frederick by making war on Spain, James turned to his old dream of healing Christendom's divisions by marrying his surviving son, Charles, to a Spanish infanta. By such a match, with some moderation of the penal laws, he hoped to persuade Spain to help in saving Frederick. He did see that a show of force was needed, too, and he planned to send an army to the Palatinate. In 1621 he also sent a squadron to the western Mediterranean, where the Spaniards dreaded the appearance, upon their vital route of communications with Italy, of warships against whose broadsides their Mediterranean galley fleet would be impotent.

There was much to be said for this policy, but James never said it to the parliament which he called to provide the money in 1621. So, though the commons did make a quite inadequate grant quite promptly, they spent most of their time airing the pent-up grievances of the past seven years. They could now expect support among the lords, too, where the older families resented Buckingham's ascendancy and James's doubling of the peerage since 1603. Hence they soon went far beyond the introduction of bills to put down monopolies and protect their freedom of speech. They revived the weapon of impeachment before the lords, forgotten under the Tudors, and used it not only to punish the monopolists Sir Francis Mitchell and Sir Giles Mompesson but also to bring down the lord chancellor, Francis Bacon—a parliamentary lesson for judges and ministers to set against the royal dismissal of Coke. In their second session they refused all but a tiny grant until the king should promise to enforce the penal laws, declare war on Spain and marry Charles to a Protestant. This attempt to dictate foreign policy challenged one of the king's highest and most undeniable prerogatives, however much the commons' later Protestation might claim it as their "ancient and undoubted birthright." It was a long step toward turning the parliamentary campaign against legitimate grievances into a revolutionary challenge for sovereign power. The king could hardly do other than dissolve the parliament. This meant that he could not back his diplomacy with force and Spain showed no haste to help him.

To speed matters Buckingham and Charles went off to Madrid (1623). By reckless concessions they obtained a marriage treaty, but still Philip IV would do nothing for Frederick. In the end even Charles and Buckingham were disillusioned. Returning home determined on war with Spain, they forced James to call another parliament (1624) and make alliances with France, Denmark and the Dutch. But recent events gave the commons little cause to moderate their claims. They granted only half what the king asked, laid down the purposes for which the money was to be used and appointed commissioners to see that it was not spent on other objects. Also they impeached the lord treasurer, Lionel Cranfield (now earl of Middlesex), and passed an act forbidding the grant of monopolies except to boroughs, trading companies and inventors. They thus consolidated the advanced positions their predecessors had seized in 1621.

CHARLES I (1625–49)

Charles, with Buckingham, had so largely shaped James's later policies that his accession (March 27, 1625) offered little chance for a fresh start. Moreover, he soon caused alarm by his favour

to the new Arminian school of divines (*see* ARMINIUS, JACOBUS), whose Protestantism was suspect and whose politics were inclined to absolutism. His relaxation of the penal laws (the price of his marriage to the French princess Henrietta Maria) and the failure of the Cádiz expedition (October), did nothing to restore confidence. Nor did his parliaments show much inclination to moderate courses. The first (June–August) tried to force a settlement on impositions by voting him tonnage and poundage (*q.v.*) for one year, instead of for life as had been usual since the 15th century. From the second (1626) he excluded Sir Thomas Wentworth, Sir Edward Coke and others by making them sheriffs and he also tried to exclude the earls of Bristol and Arundel from the lords. This drove the houses together and emboldened Sir John Eliot (*q.v.*) and the commons to a general attack on royal policy, which culminated in the impeachment of Buckingham. A dissolution saved the duke but plunged Charles into still more desperate courses. He and Buckingham now contrived to quarrel with France as well as with its great enemy Spain. Every device was tried to raise money for expeditions, again quite unsuccessful, in 1627 and 1628 to aid the rebel Huguenots at La Rochelle. A large forced loan, whose legality the judges doubted even after the dismissal of the lord chief justice, Sir Randolph Crewe (1626), roused particular opposition. This was inflamed by the arbitrary imprisonment, sanctioned in *Darnel's Case* (1627; *see* HABEAS CORPUS), of many who refused to pay; by the billeting of soldiers for whom the government could not afford to hire quarters; and by the imposition of martial law on the soldiers and civilians round them.

Also, while the Puritans remained curbed by the 1622 restrictions on preaching and the 1624 proclamation against unlicensed printing, Arminians like Richard Montagu, Roger Manwaring and Robert Sibthorpe were allowed to preach that parliament's consent was not necessary to taxation, and their leader William Laud (*q.v.*) was made a privy councilor. So when at length Charles had to call a third parliament (March 1628), lords and commons combined to prune his prerogative and impeach Manwaring. They passed the Petition of Right, condemning arbitrary taxation and imprisonment, martial law and billeting of soldiers upon private citizens against their will as contrary to the law of the land. Charles reluctantly gave this his assent; but when the session ended he went out of his way to annoy the commons by pardoning Manwaring and making Laud bishop of London and Montagu bishop of Chichester. When the houses reassembled (1629) anger led Eliot and the commons to overreach themselves and the session ended with the speaker held forcibly in his chair while resolutions were acclaimed denouncing as capital enemies of the state any who sought to extend popery or Arminianism or to levy or pay tonnage and poundage. This violence alienated not only the lords, less hostile since Buckingham's assassination (Aug. 1628), but also many moderate men, like Wentworth (*see* STRAFFORD, THOMAS WENTWORTH, Earl of) and William Noye, who had already entered the royal service.

Personal Government.—The reaction gave Charles a chance to do without parliament and during the next 11 years conciliar government reached its zenith. It kept within the letter of the law, using the judges to authorize collection of impositions and tonnage and poundage and (after suspension of the chief baron of the exchequer Sir John Walter) to sanction the imprisonment of Eliot and others. Its efforts to regulate trade in the common interest, to check enclosures, to enforce proper wage assessments and proper administration of poor relief were in the best tradition of conciliar rule. With Wentworth leading the council of the north, until he went to Ireland in 1633, these efforts were felt in the remotest shires. Moreover, once peace was made with France (1629) and Spain (1630), there was a quick revival of trade, further stimulated by colonial expansion. The first New England colony, founded by the Puritan sectaries and others who had sailed in the "Mayflower" (1620), was followed by a second, promoted by the Puritan Massachusetts Bay company (1628), which soon had offshoots in Rhode Island, New Hampshire and Connecticut. Lord Baltimore established Maryland as a Catholic refuge (1634). In the West Indies Barbados, St. Kitts, Nevis, Montserrat, Antigua and Providence were settled (1624–36). By 1640 perhaps 60,000 people had emigrated to these new world colonies and colonial trade had become an important contributor to England's wealth.

Yet the return to purely conciliar government was a dangerous experiment. The tightening of central control encroached upon the power of the gentry and commercial classes in county and borough and demanded more unpaid services from them just when they were being denied means to influence national policy through parliament. Moreover, the government's efforts to enforce the old standards of social justice often cut across their private interests. Most dangerous of all, conciliar machinery was used vigorously to back an unpopular ecclesiastical policy and was maintained by irritating financial expedients. Laud, archbishop of Canterbury from 1633, supported from 1636 by William Juxon, bishop of London, as lord treasurer, acquired a mounting influence over king and council. His metropolitical visitation (1634–37), while it did much to restore dignity and even decency to the Anglican services, enforced Arminian standards of ritual and doctrine which affronted the Puritans and won little lay support. His use of the high commission to back episcopal authority and of the Star Chamber to silence such critics as William Prynne, Henry Burton and John Bastwick (1637) increased the unpopularity of prerogative rule. In addition, Puritan propaganda against the Arminians as virtual agents of Rome, however unjust, was not unplausible—Montagu did favour reunion, Laud was (to his disgust) offered a cardinal's hat and Charles's queen, Henrietta Maria, welcomed a papal nuncio to court (1634). At the same time the nobility and gentry were irritated by traffic in wardships, distraint of knighthood fines and fines for alleged encroachments on the royal forests and for enclosures. Grants of monopolies, often to very dubious companies, annoyed consumers and those who failed to get them. And, as with the London merchant Sir William Courteen and his association, and the East India company, Charles's need of money sometimes led him to infringe in one patent the privileges granted in another. By 1637 the budget was balanced but at the price of a discontent so widespread that revenue could hardly be expanded to meet an emergency.

This became clear after Charles began to levy ship money (*q.v.*) annually from inland as well as coastal counties in 1635. The judges, tuned by the dismissal of the lord chief justice, Sir Robert Heath (1634), pronounced this legal and the money was spent to equip the fleet. Yet by relieving the exchequer of the upkeep of the navy, it in effect gave Charles, without parliamentary sanction, a big addition to his ordinary revenue. The opposition leaders, many of whom were in frequent touch with each other in the colonizing companies, challenged its legality without success in Hampden's case (1637) (*see* HAMPDEN, JOHN). Their challenge showed them the popularity of their cause and brought a sharp fall in the amount of ship money collected.

Scotland.—But this negative check was not enough and Charles's opponents only got their chance when a rebellion in Scotland forced him to call a parliament. By ordering the restoration of church and crown lands (1625) he had already alienated the Scots nobles, whose support had enabled James to impose a modified form of episcopacy on the Scottish Kirk. Now he and Laud recreated the alliance of nobles and kirk by trying to impose a liturgy closely resembling the Anglican (1637). Riots in Edinburgh grew into a national revolt and multitudes signed the covenant to defend the kirk (1638). The revived general assembly organized an army and the first of the Bishops' Wars ended with Charles's capitulation at the treaty of Berwick (1639). A second war soon followed and Charles, on the advice of Wentworth (now earl of Strafford), at last called a parliament to finance it. This Short parliament (April–May 1640) was not of a revolutionary temper and, though under John Pym's leadership it refused supply until grievances were redressed, Charles might perhaps have rallied substantial support even in the commons by patience and reasonable concessions. Instead he hastily dissolved it. Opinion now turned sharply against him. He could raise little money and no effective army. The Scots occupied Northumberland and Durham and a great council of peers produced only a treaty with the Scots which committed Charles to pay them £25,000 a month until a final settlement. This made another parliament unavoidable.

The Long Parliament.—With the Scots at Newcastle to be paid, with London merchants refusing loans to the king and the London mob rioting against bishops, the new parliament held the whip hand. In the next nine months it brought down the king's chief advisers, swept away most of the machinery of conciliar government and made frequent parliaments a statutory necessity. Strafford was condemned by act of attainder, after the evidence against him proved inadequate for impeachment; and Charles, in fear of the mob, consented to his execution (May 1641). Laud was impeached and imprisoned, while the lord keeper, Lord Finch, and the secretary of state, Sir Francis Windebank, fled overseas. The courts of Star Chamber and high commission and the council of the north were abolished. The council in the Welsh marches lost its extraordinary jurisdiction. Persons arrested by the privy council were to have the legality of their commitment decided within three days by the common-law courts. The Triennial act provided for a parliament every three years, even if the king failed to summon one; and another act forbade the dissolution of the present parliament without its own consent. The levying of tonnage and poundage and impositions without parliamentary consent was made illegal, the ship money judgment was reversed, knighthood fines abolished and the boundaries of the royal forests limited.

So far, except over Strafford's attainder, there had been unanimity in the commons and between the commons and the lords. In the summer of 1641, however, divisions began to appear. Some, like Edward Hyde (later the earl of Clarendon) and Viscount Falkland, began to feel that enough had been done; others, like John Pym and John Hampden, could not trust Charles until they had wrested control of the executive from him. Religion widened and embittered the split. The first group, fearing the growing influence and wildness of the more extreme Puritan sects, rallied to the episcopal church. The second, alarmed by army "plots" and the Roman Catholic queen's influence, were driven into alliance with "root and branch" Puritan reformers and alienated the lords by seeking to exclude bishops from the upper house. Then came the Catholic rebellion in Ireland (October). An army was needed to suppress it. But to raise armies and appoint commanders was a royal prerogative which Pym and his friends dared not let Charles exercise. Their attempt to seize sovereign power in its most obvious form, control of the armed forces, turned their episcopalian opponents into a royalist party. Knowing, after the experience of 1629–40, that they would be helpless unless they could identify themselves with parliament, or at least with the house of commons, they forced through the Grand Remonstrance (November). This recited all the alleged misdeeds of Charles's reign and demanded that the king's ministers be appointed with the approval of parliament and the church reformed by parliament in consultation with an assembly of divines. That it was carried by only 11 votes in a house barely two-thirds full shows how precarious Pym's position was. Had Charles been content to sit still, the royalists might soon have gained the upper hand. But Charles, fearing for his queen, tried personally to arrest five members—Pym, Hampden, Denzil Holles, Sir Arthur Hesilrige and William Strode (Jan. 1642). His rash and abortive appeal to force enabled his opponents to keep their majority, to begin to organize themselves as a rival government and to raise troops by ordinance. Charles, after sending the queen abroad, retired to York. There followed an interchange of manifestoes, ending with the Nineteen Propositions (June) which repeated the Grand Remonstrance's demands and claimed also parliamentary control of the armed forces.

Charles naturally rejected this and on Aug. 22, 1642, raised his standard at Nottingham. The Civil War had begun. (*See* CIVIL WAR, ENGLISH.)

THE CIVIL WAR

First Phase.—For 18 months the war went in Charles's favour. After the indecisive battle of Edgehill (Oct. 1642), the main parliamentary army under the earl of Essex withdrew from the midlands. Charles advanced straight on London. His thrust was parried, and next year's three-pronged movement against the capital broke down because Hull, Plymouth and Gloucester held up his northern, western and Welsh forces. Yet by the end of 1643 he was master of almost all England west of a line from the Wash to Southampton. In wealth, trade and even manpower, however, this area was inferior to that still controlled by parliament, which thanks to the earl of Warwick also had the fleet on its side. Hence, as parliament's resources were gradually mobilized, Charles's chance of victory faded. A committee of safety (later the committee of both kingdoms) soon provided a vigorous executive responsible to parliament. New and more efficient and more equitable if heavier forms of taxation were evolved to supplement the customs—first, monthly assessments fixed for each county and borough by parliament but assessed and collected by local commissioners; then (July 1643) an excise. Counties were grouped into associations for mutual defense and in the Eastern association Oliver Cromwell (*q.v.*) began to develop a body of horse which was to offset that superiority in cavalry which the royalist sympathies of a majority of the peers and gentry had hitherto given to the king. But parliament still lacked an effective field army, vigorously led and not tied to local defense. The setbacks of 1643 drove it to call in the Scots, at the price of the Solemn League and Covenant which promised to extirpate prelacy and popery and establish a reformed church in England. Parliament gained by this as much as Charles lost by his Cessation or armistice with Roman Catholic Ireland. The Scots joined Sir Thomas Fairfax's Yorkshiremen and Cromwell's troops to clear the north by routing Prince Rupert and the marquess of Newcastle at Marston Moor (July 1644). But the king saved the southwest by forcing the army of the earl of Essex to capitulate at Lostwithiel (September). So parliament passed the Self-Denying ordinance, which made possible sweeping changes in command and created an effective field force in the New Model army. Under Fairfax and Cromwell this army won the midlands by its victory at Naseby (June 1645). Then, after subduing the southwest, it ended the war by capturing the royal headquarters at Oxford (June 1646) whence Charles had fled to seek refuge with the Scottish army.

Parliament thus achieved victory through the agency of the Scots and the New Model army. Its two agents, however, had their own ideas about a settlement. The Scots were soon satisfied by parliament's Newcastle propositions (August), demanding parliamentary control of the militia, punishment of the king's chief supporters and the establishment of Presbyterianism. When Charles refused these terms and parliament offered £400,000 toward its expenses, the Scottish army handed him over and marched home (Jan. 1647). The English army was less easily satisfied. In it, and especially in the cavalry, a wide variety of Independent sects flourished, bound together by dislike of uniformity whether Presbyterian or Anglican. The whole army soon joined these sects in opposition when parliament decided to send some regiments to Ireland, to disband others and to impose the Covenant on all, without offering either an adequate indemnity for past actions or satisfactory security for their large arrears of pay. The soldiers refused to disband. They seized the king (June) and, strongly influenced by the democratic ideas of John Lilburne and the Levellers (*q.v.*), chose agitators, or agents, to join with the officers in a general council of the army.

There during the summer of 1647 John Wildman, Thomas Rainborough and Edward Sexby expounded ideas hardly heard since Ket's rebellion under Edward VI, maintaining against landed gentlemen like Cromwell and Henry Ireton that the franchise was the birthright of even "the poorest he" and not a privilege confined to men of property. The officers negotiated with Charles for a settlement on the basis of their Heads of the Proposals, with religious toleration, a council of state chosen from the nominees of biennial parliaments and lenient treatment for royalists. The Levellers produced an Agreement of the People (October), calling for manhood suffrage and the establishment of religious toleration, freedom from impressment and equality before the law as inalienable rights of all Englishmen.

Second Phase.—Eventually outside events interrupted these debates. Even after the army's occupation of London (Aug. 1647), the Presbyterian majority in parliament remained intransigent, while Charles laid the foundation of a royalist-Presbyterian-Scottish alliance by offering an Engagement to allow

Presbyterianism for three years and parliamentary control of the armed forces for ten. His flight to the Isle of Wight and conclusion of his Engagement (Dec. 1647) with the Scots were the signals for sporadic risings all over England and Wales and for a Scottish invasion. The army, however, dealt firmly with the risings, Cromwell routed the Scots at Preston (Aug. 1648), and this second phase of the Civil War closed with the army supreme. Parliament, purged by Col. Thomas Pride to a mere remnant, a Rump parliament of about 60 Independents (December), brought Charles to trial and execution (Jan. 30, 1649), abolished the house of lords and declared England a commonwealth.

The Commonwealth

The Commonwealth's first tasks were to establish its authority over the British Isles and colonies and to win recognition from foreign powers. Cromwell's Irish campaign (1649–50), with its brutal massacres at Drogheda and Wexford, broke the back of Catholic Ireland's resistance. But meanwhile the Scots had induced Charles II to take the Covenant (June 1650) and though Cromwell's victory at Dunbar (September) ended the covenanting party's ascendancy (*see* Covenanters) Charles continued the struggle. Next year he invaded England. Few joined him there and Cromwell caught and destroyed his army at Worcester (Sept. 1651), though Charles himself escaped to France.

By May 1652 Ireland was subjugated; the islands of Scilly, Man and Jersey and the royalist colonies of Barbados, Virginia and Maryland had submitted; and Robert Blake and William Penn had chased Rupert's royalist squadron from Kinsale to the Tagus and then into and out of the Mediterranean until it dwindled to a single ship. In full control of the British Isles and empire, the Commonwealth could now take a firm tone with foreign powers. Its Navigation act (1651) and its attempt to maintain English sovereignty in the narrow seas (the seas and straits adjoining the English coast) brought war with the Dutch (May 1652), who had become England's most dangerous commercial and colonial rivals. The war went well for England and, with France and Spain still fighting each other and eager to avoid English hostility, the Commonwealth's prestige stood high in Europe.

At home, however, the Rump parliament and its council of state had outlived their usefulness. With their high court of justice, censorship, heavy taxation and confiscations of royalist property, their rule was quite as arbitrary and much less legal than Charles I's personal government. Their suppression of popular amusements and their attempts to enforce Puritan standards of conduct and morality irritated ordinary people much as the Book of Sports (issued by James I and reissued under Charles I, permitting dancing, archery and other recreations on the Sabbath) and Laud's visitations had offended the Puritans. They represented but a fraction of the nation and owed their continuance to the army. Yet they made little serious effort to produce the lasting settlement and the sweeping reforms in law and church organization that the radical and "godly" party in the army wanted. Nor did they seem ready to dissolve themselves and make way for an assembly more representative, if not of the people of England at least of the "people of God." So at last even the army turned against them and forcibly ejected them (April 1653).

The Protectorate.—The army council now selected an assembly of 140 members (including 5 for Scotland and 6 for Ireland) from nominees of the congregational churches. However, the zeal of this Nominated (or Barebone's) parliament (July–Dec. 1653) for indiscriminate reform soon threatened to overturn rather than to settle the nation. So the army leaders sought a broader basis for their rule, while yet safeguarding such fundamentals as republicanism and toleration for the sects, which were abhorrent to the royalists and Presbyterians who formed the great majority of the nation. In the Instrument of Government they enshrined these fundamentals in a written constitution with a carefully balanced separation of powers. Cromwell as protector was given a fixed revenue but shared executive power with a council of state appointed, like himself, for life; while the legislative power and granting of extraordinary supply were vested in triennial parliaments representing the three kingdoms. The scheme failed because no parliament would accept such limitation of its competence by the army, all the less since the expulsion of the Rump parliament had split the independent republican party and turned such men as Sir Arthur Hesilrige, Sir Henry Vane and Edmund Ludlow into bitter enemies of the military leaders. Hence, despite preliminary purgings and subsequent exclusions, the first Protectorate parliament (1654) persisted in framing a constitution of its own and claiming control of finance, armed forces and religious policy.

Cromwell therefore dissolved it as soon as possible (Jan. 1655). Then, after John Penruddock's royalist rising (March), he divided England into areas governed by major generals, backed by a new militia and financed from a "decimation" of royalist rentals. Their strict administration and their vigorous enforcement of the Puritan moral code turned the country finally against militarism and Puritan rule. When the outbreak of war with Spain brought a new parliament (Sept. 1656), Cromwell realized the need for a fresh settlement and that this was more likely to be accepted if it were devised in parliament.

So, although he allowed the council to exclude more than 100 republican members, he did nothing to prevent the rest from framing a new constitution. This, the Humble Petition and Advice (March 1657), passed despite the opposition of the army members, proposed to make Cromwell king, to create a new house of lords and to give the commons sole control over their own membership. Wisely or unwisely, Cromwell refused the crown but accepted the rest. The army sullenly acquiesced, after the dismissal of John Lambert and a few others; but the excluded republicans were now admitted to the commons and were able to outvote the "court" party, weakened by elevations to the upper house. Once again they drove Cromwell to a dissolution (Feb. 1658). Yet he had taken a considerable step toward placing his authority on a more national basis.

Moreover, Cromwell had given England an influence abroad such as it had not enjoyed since Elizabeth's day. His dream of a great Protestant crusade against the Habsburgs faded because the Germans, Danes and Dutch were now more afraid of Swedish than of Catholic aggression. He did nevertheless improve the British traders' position in the Baltic. Likewise, though the French Huguenots refused his proferred support, he made the French government intervene on behalf of the persecuted Vaudois of Savoy. His Spanish war, too, though neither necessary nor popular, was successful. Jamaica was captured. Blake and Richard Stayner destroyed two Spanish treasure fleets (1656). English troops and their French allies routed the Spaniards at the battle of the Dunes and occupied Dunkerque (June 1658). A victorious peace was within reach and with it the possibility of ending Cromwell's financial difficulties. He had almost tamed the army and his resistance to both the militarists and the extreme republicans attracted moderate men. A new parliament might perhaps have co-operated with him in framing a lasting settlement. But his death (Sept. 3, 1658) dashed all such hopes. (*See* also Cromwell, Oliver.)

Fall of the Republic.—His elder son, Richard, who quietly succeeded him as protector and lord general, was neither a "saint" nor a soldier. He had none of Oliver's hold over the army, which soon resumed its independent role, demanding that Charles Fleetwood (*q.v.*) be made commander in chief with power to appoint and dismiss officers. Richard refused, and the majority of his first parliament (Jan. 1659) supported him with more vigour than tact. Vane, Hesilrige and Ludlow thereupon forgot their quarrel with the soldiers who had expelled the Rump parliament and egged them on to the forcible expulsion of this parliament. Richard abdicated soon afterward and the Rump parliament was restored in May. The old feud between civilian and military republicans, however, broke out afresh and, once Sir George Booth's royalist rising had been suppressed (August), Lambert and Fleetwood expelled the Rump parliament a second time. But now a fatal division appeared in the army itself. George Monck and the troops under him in Scotland, declaring for the civil power and a free parliament, marched on London. Lambert tried to stop them, but his forces deserted him, while in London Fleetwood again restored the Rump parliament. The nation, however, was rapidly turning to a Stuart restoration as the only hope of a settlement. Monck

therefore first restored to their seats the members excluded in 1648, then persuaded them to make way for a newly elected house and lastly opened negotiations with Charles II. The Presbyterians in the new parliament sought to lay down terms of his restoration but the tide of royalism was too strong and Monck too quick for them.

So on May 29, 1660, Charles II entered London, welcomed deliriously by his people and bound by no explicit conditions except for the promises in his own Declaration of Breda to grant, subject to parliament's sanction, a general amnesty, religious toleration and payment of the army's arrears.

XI. THE RESTORATION AND THE REVOLUTION (1660–1714)

Charles II (1660–85)

Though no explicit conditions were imposed on Charles II, he was in no position to dictate terms. Indeed, he and his chief minister, Edward Hyde, earl of Clarendon (*q.v.*), had better reason than Oliver Cromwell to seek "healing and settling." Knowing this, they passed through the Convention parliament (1660) an Act of Indemnity and Oblivion which exempted only the regicides from pardon and, while restoring lands lost by actual confiscation, left to their new owners properties sold by royalists to meet the penal taxes imposed upon them during and after the Civil War. In religion too the Savoy conference attempted, though in vain, to find a settlement which might comprehend both Anglicans and Presbyterians.

Nor could the crown reclaim the coercive power of the Protectorate government or of the old monarchy. The army was paid off, except for about 5,000 men. All acts to which Charles I had assented before the Civil War were regarded as binding on Charles II. There was thus no Star Chamber, no high commission, no local councils. Imprisonment by order of king or council, without trial at common law, was no longer lawful, though until the 1679 Habeas Corpus Amendment act the law might still be evaded and until 1701 the judges held office at the king's pleasure. Justice, apart from the equitable jurisdiction of the chancellor, became the common-law courts' monopoly and the executive's control over local authorities was much loosened.

Moreover, the crown's financial independence was drastically, if in part unintentionally, curtailed. It could not levy impositions, forced loans or ship money by prerogative, and Charles agreed to surrender feudal dues and purveyance in return for a hereditary excise which, with the customs and crown lands, was intended to give him an ordinary annual revenue of £1,200,000. Even the full sum would barely have covered his peacetime expenses; and in fact for many years he received less than three-quarters of it.

The King and Parliament.—So from the first Charles II had to bend his policies to the will of parliament. He and Clarendon were not allowed to redeem their Breda promise of "liberty for tender consciences." The Cavalier parliament (1661–79) drove the Presbyterians out of the church by its Act of Uniformity (1662) and, despite its enthusiastic royalism, would not allow Charles to use his dispensing power to relieve the Nonconformists from the heavy religious and civil disabilities imposed upon them by this and other acts of the so-called Clarendon code (1661–65). The Dutch War of 1665–67 brought another burst of criticism. The war, generated by long commercial and colonial rivalry, was popular. Parliament voted the unprecedented sum of £2,500,000 and then another £1,250,000 (1665), though this time with a clause appropriating it to war purposes only. Yet in 1667 the government, its embarrassments increased by the Great Plague (1665) and the Fire of London (1666), could not put the fleet to sea and so suffered the disaster of the Dutch raid up the Medway. England gained fair terms at the peace (July 1667), retaining New Amsterdam (renamed New York) to close the gap between New England and its other American colonies. But the commons impeached Clarendon (whom Charles had already dismissed) and insisted on appointing commissioners to investigate the royal accounts. Some more positive attempt to work with and manage the crown's parliamentary paymasters than Clarendon had made was clearly called for.

The group of ministers, known as the Cabal, through whom Charles worked for the next few years were, however, hardly the men to win hearty support from a strongly Anglican parliament. Sir Thomas Clifford and the earl of Arlington had Catholic sympathies; the duke of Buckingham's record was ambiguous; Lord Ashley (earl of Shaftesbury from 1672) favoured toleration for all Protestants; the earl of Lauderdale was at heart a Presbyterian. Yet the emergence of the Spanish succession question (*see* Spain: *History*) now made co-operation between crown and parliament more than ever necessary if England was to uphold its commercial and strategic interests. In 1665 Philip IV of Spain had died. His only surviving child, Charles II, was not expected to live long and was unlikely to beget an heir. Among the likelier claimants to his inheritance was Louis XIV of France, and in 1667–68 the short War of Devolution showed that Spain no longer had the strength to defend its vast dominions against a France which, under Louis and Colbert, was already becoming a serious commercial and maritime rival to both England and the Dutch. So for more and more Englishmen fear of France gradually began to outweigh jealousy of the Dutch. In 1668 England joined the United Provinces and Sweden in the triple alliance, which called a halt to Louis's conquests in the Spanish Netherlands.

Soon, however, Charles reversed this policy. In 1670 he agreed by the secret treaties of Dover to help Louis to conquer the United Provinces and even promised with the help of French troops and money to restore England to Roman Catholicism. These secret clauses were revealed only to Clifford and Arlington, but the general drift of Charles's policy soon aroused suspicions which broke up the Cabal, poisoned his relations with parliament and left England powerless to play its part in Europe. In 1672 he joined Louis in declaring war on the Dutch and at the same time issued the Declaration of Indulgence suspending by prerogative all penal laws against Roman Catholic recusants as well as Protestant dissenters.

The Dutch, under the young William of Orange, held off the French armies and the Anglo-French fleets until allies came to their rescue. Meanwhile, parliament forced Charles to withdraw his declaration and accept the Test act, which required all holders of civil or military office to take the Anglican sacraments and a declaration against transubstantiation (1673). In 1674 he had to make peace, and his parliamentary opponents found a brilliant leader in Shaftesbury, whom he had dismissed (Nov. 1673) and who had John Locke as counselor and physician. For another four years Charles was able to hold them off. An improvement of trade and further French subsidies eased his finances, while his new lord treasurer, Thomas Osborne, earl of Danby (*see* Leeds, Thomas Osborne, 1st Duke of), showed considerable skill in building up support in parliament by judicious use of crown patronage and bribery. But all the time the continued Franco-Dutch war and the growing strife between the opposition and government parties organized by Shaftesbury and Danby—the beginnings of the future Whig and Tory parties—were increasing tension until the smallest spark could produce a mighty explosion.

The "Popish Plot."—The explosion came in 1678 with the revelation by the rascally Titus Oates of an alleged "popish plot" to kill the king, put his brother James (now an avowed Roman Catholic) on the throne and restore England to Catholicism. The plot was a fiction, but Shaftesbury and the opposition seized upon it as a means of inflaming opinion against the court, a course in which they were encouraged by Louis XIV, with whom Charles's relations had grown cold since the marriage of James's elder daughter, Mary, to William of Orange (Nov. 1677). So, while a number of innocent men were condemned and executed for treason, parliament passed an act excluding all Roman Catholics from sitting in either house. They went on to demand James's exclusion from the king's counsels and to impeach Danby.

To save his minister, Charles at last dissolved the Cavalier parliament (Jan. 1679). Its successor, however, proved even less amenable, despite Charles's halfhearted adoption of Sir William Temple's scheme for a privy council of 30, half of them parliamentary leaders, half royal servants. The commons passed a bill of attainder against Danby and a bill to exclude James from the succession to the throne. In a third parliament (1680) Shaftesbury and the exclusionists, now becoming known as Whigs, went

BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

THE LAST HORSE RACE HELD BEFORE CHARLES II, AT DORSETT FERRY, 1684; FROM AN ETCHING BY FRANCIS BARLOW, 1687. IN THE BRITISH MUSEUM

even further. Passing over not only James but also his Protestant daughters, Mary and Anne, and Mary's able Protestant husband, William of Orange, they claimed the succession for Charles's illegitimate son, the duke of Monmouth, who was likely to be little better than Shaftesbury's puppet. This was more than the Tories, or indeed moderate men like George Savile, marquess of Halifax, could stomach and they persuaded the lords to reject a new Exclusion bill. Thereupon Charles, again hopeful of French subsidies, dissolved the parliament and summoned a new one to Tory Oxford, away from the pressure of the fanatical London mob (1681). When it met, he dissolved it almost immediately, for he now had Louis's promise of £400,000 over the next three years.

A strong Tory reaction, stimulated by the abortive Rye House plot (1683) to assassinate the king and James, which cost the earl of Essex, Lord Russell and Algernon Sidney their lives, enabled him to do without parliament for the remaining four years of his life. He was able also to deprive the Whigs of the chief strongholds of their influence and parliamentary power, the municipal corporations. On various legal pretexts, London and more than 60 other cities and boroughs were held to have forfeited, or were induced to surrender, their ancient charters. The new ones they were given allowed the crown more control over them and the election of their members. By thus breaking the Whigs, Charles secured the succession for his Catholic brother on his own death (Feb. 6, 1685). With this limited triumph he wisely rested content.

James II (1685–88)

James was not capable of such statesmanlike restraint. The generosity of his first and only parliament, which made him a grant of £400,000 on top of the annual revenues which Charles had had, and the rapid failure of Monmouth's rebellion (June 1685) blinded him to the limitations of that Tory loyalty to which he owed his throne. He bluntly announced that he would neither disband the large army raised during the rebellion nor dismiss those of its officers who refused to comply with the Test act requiring them to take the Anglican sacraments. The Tory parliament, abhorring standing armies and devoted to the Anglican ascendancy, caviled and was hastily prorogued.

James then, in the fatal fashion of his father, sought from the prerogative and the law courts what he could not gain from parliament. He dismissed the judges who were likely to disagree with him, appointing others in their place, and the court upheld in the case of *Godden* v. *Hales* (1686) his right to dispense individuals from the provisions of the Test act (*q.v.*). Thus armed, he gave to many more Roman Catholics commissions in the army, which camped each summer close to London on Hounslow heath. He encouraged the earl of Tyrconnel to raise a more wholly Roman Catholic army in Ireland. He admitted Roman Catholics, even the Jesuit Edward Petre, to the privy council. In defiance of the 1641 act abolishing the court of high commission, he set up a court to exercise the royal supremacy over the Anglican Church and used it, among other purposes, to thrust Roman Catholics into the universities, the stronghold and training ground of the Anglican clergy and the Tory gentry. In 1687 he stretched the prerogative yet further and made a bid for Protestant nonconformist support by the Declaration of Indulgence, which suspended the penal laws against dissenters and recusants. Borough corporations were purged of Anglicans and Tories, and efforts made to win the county authorities to a repeal of the penal laws. In April 1688 he issued a second Declaration of Indulgence, ordering it to be read from every pulpit on two successive Sundays. William Sancroft, the archbishop of Canterbury, and six other bishops petitioned him against this and were prosecuted for seditious libel. Their acquittal almost coincided with the birth of a son to James's Roman Catholic queen, Mary of Modena (June). This event promised an indefinite continuance of his policy and brought discontent to a head. Seven eminent Englishmen, among them the Tory Danby and Henry Compton, bishop of London, wrote inviting William of Orange to come over with an army to redress the nation's grievances.

William's chief concern was to check the overgrowth of French power in Europe. Between 1679 and 1684 England's impotence, and the emperor Leopold's preoccupation with a Turkish advance to Vienna, had allowed Louis to seize Luxembourg, Strasbourg, Casale and other places vital to the defense of the Spanish Netherlands, the German Rhineland and northern Italy. By 1688, however, a great European coalition had begun to form to call a halt to these aggressions. Its prospects depended partly upon England's attitude. William, who had been in close touch with the leading English malcontents for more than a year, accepted their invitation. Landing at Torbay (Nov. 5), he advanced slowly on London and it soon appeared that James had alienated the nation as thoroughly as Charles I in 1640 or the republicans in 1660. Support fell away from him. His daughter Anne and his best general, John Churchill, were among the deserters to William's camp; thereupon James fled. His first attempt to reach France was frustrated by some overzealous fishermen, but William contrived to alarm him into trying again and this time he was not prevented.

William III (1689–1702) and Mary II (1689–94)

The Revolution Settlement.—William was now asked to carry on the government and summon a parliament. When this convention parliament met (Jan. 22, 1689), the Tories were in no position, and many of them, like Danby, in no mood, to resist an exclusionist solution. Some did advocate recalling James upon strict conditions, or appointing a regency to rule in his name. But after some debate the convention agreed to treat James's flight as an abdication and to offer the crown, with an accompanying Declaration of Right, to William and Mary jointly. Both gift and conditions were accepted. Thereupon the convention turned itself into a proper parliament and the Declaration into a Bill of Rights. This bill gave the succession to Anne in default of issue to William and Mary; barred Roman Catholics from the throne; abolished the power to suspend laws; condemned the power of dispensing with laws "as it hath been used of late"; declared a standing army illegal in time of peace; and required frequent parliaments and free elections (*see* Bill of Rights, English). The settlement thus marked a considerable triumph for Whig views. If no Catholic could be king, then no kingship could be unconditional. The adoption of the exclusionist solution lent support to Locke's contention that government was in the nature of a social contract between the king and his people represented in parliament.

These principles were further elaborated in the legislation and practice of William's reign. William had no desire to see the royal power weakened. But his Whig supporters were not supporters of prerogative powers and Tory sentiment could not warm to a cold Calvinist who was king only by act of parliament. Moreover, the war against France, in which England now joined and which cost it about £40,000,000 before it ended in 1697, made him very dependent upon parliament. The commons considerably increased the crown's ordinary revenue, but they now voted it annually instead of for life—a device also adopted in the annual Mutiny act, which alone could legalize military discipline, to assert some parliamentary control over the armed forces. Extraordinary taxation,

raised by a variety of measures of which the land tax is the best known, was of course wholly in their control.

Periodically, too, they appointed commissioners to look into the public accounts and see that the money was spent on the purposes for which it was granted. Even government borrowing came under parliamentary control and the Bank of England, founded in 1694 largely to assist in the raising of loans, was forbidden to lend to the crown without consent of parliament. This, and the creation of a national debt so great that its annual charge equaled the ordinary revenue granted to Charles II in 1660, made yearly meetings of parliament a practical necessity. The prerogative of summoning parliaments thus lost its importance and the 1694 Triennial act deprived the king of the right to keep a complacent house of commons in existence for longer than three years.

The need to support William during the war did, however, set a limit to such encroachments. He was able to keep a close personal control over foreign and military affairs. He chose his own ministers, who were responsible directly and individually to him, not collectively to parliament. In later years he found it easier to select most of them from the same party, but his government never became a party, or even a parliamentary, government. The political division was indeed rather between court and country than between Whig and Tory, for William made some progress in using crown patronage and influence to build up a group of government supporters akin to the later King's Friends. Nevertheless, after the peace of Rijswijk (1697) Tory parliaments cut down his army in England to 7,000 men, called in question his grants of forfeited Irish lands and impeached some of his ministers. They tacked further restrictive clauses to the Act of Settlement, which was passed (1701) to assure the succession after Anne to the next Protestant heir, the electress Sophia of Hanover, Elizabeth of Bohemia's daughter. These clauses insisted that sovereigns after William must join the Anglican communion and must not leave the realm nor engage it in war on behalf of their foreign dominions without parliamentary consent. Further, the act barred foreigners from parliament, privy council and the public service; disqualified anyone holding a pension or office of profit under the crown from sitting in the commons; and made the judges irremovable except upon a request of both houses.

The Struggle Against Louis XIV.—For England the war of 1689–97 had proved an uphill struggle. The first tasks had been to secure control of the British Isles and command of the sea. Scotland was soon assured. Its parliament followed that of England in offering the crown to William and Mary, with a Claim of Right which included the abolition of episcopacy. Graham of Claverhouse, Viscount Dundee, did raise some Highlanders for James, but their effort died away after Dundee's death at Killiecrankie (July 1689). Ireland was a bigger problem, for Tyrconnel had a large Roman Catholic army to which James soon brought French help. William's victory at the Boyne (July 1690) drove James back to France, but Irish resistance continued until the surrender of Limerick (Oct. 1691). That the promises then made to them were soon so shamelessly violated is in part a measure of the Irish menace in 1689–91. Command of the sea came even more slowly. In 1692 the battle of La Hogue, or Barfleur, at last ended French dominance in the English channel and in 1694 an English fleet appeared in the Mediterranean to protect trade against French raiding and to assist Spain and Savoy. By then large English land forces were also engaged in the annual indecisive campaigns in the Netherlands, which went on until the treaty of Rijswijk. By this peace Louis abandoned James.

Soon afterward Louis and William concluded two successive treaties (1698 and 1699) for the partitioning of the Spanish empire, now that Charles II's death was visibly approaching. The Spaniards, however, would not hear of partition, and Charles made a final will leaving his dominions as a whole to Philip of Anjou, second son of the French dauphin. When Charles died (1700) Louis accepted this will for his grandson, knowing that neither the English nor the Dutch would fight for the partition treaties. Then his judgment forsook him. He closed France and the Spanish empire to English imports; seized the frontier fortresses of the Spanish Netherlands and the Milanese; and proclaimed James II's son king of England on his father's death (1701). These actions outraged English and Dutch opinion and brought war.

Anne (1702–14)

War of the Spanish Succession.—William's death (March 8, 1702) and Anne's accession gave the allies in John Churchill, duke of Marlborough (*q.v.*), a commander who equaled William in diplomacy and far outshone him in generalship. For four years the Grand alliance—England, the Dutch, the emperor, with Portugal, several German princes and later Savoy—enjoyed almost unbroken success. They drove the French from Germany by the victory at Blenheim (1704); from north Italy by the victory at Turin (1706); almost out of the Netherlands after Ramillies (1706); while in Spain the English took Gibraltar, and Catalonia rose against Philip V. In 1706 the allies rejected Louis's peace offers and demanded the entire Spanish inheritance for the emperor's second son, the archduke Charles. During the next four years they failed in two invasions of southern France and in Spain suffered a decisive defeat at Almansa. But they conquered Naples, Sicily, Sardinia and Minorca, while Marlborough's victories at Oudenaarde (1708) and Malplaquet (1709) almost opened the way to Paris. Again Louis sued for peace, only to meet a demand that he should help to expel his grandson from Spain. This was too much, not only for Louis but also for the Tories, who from 1710 dominated Anne's counsels. Almansa had made "no peace without Spain" an impracticable war aim and soon the archduke Charles's election as emperor (1711), on the unexpected death of his brother, the emperor Joseph I, made it hardly desirable. So the Tory government began separate peace negotiations, dismissed Marlborough (1711) and issued orders restraining his successor from active operations.

By the treaty eventually signed at Utrecht (April 1713) Philip remained king of Spain, but the French and Spanish crowns were never to be united; and England gained by an *asiento* (*q.v.*) limited trading rights in Spanish America. The Spanish Netherlands were transferred to the emperor and their barrier fortresses were to be garrisoned by Dutch troops. English supremacy in the western Mediterranean was also assured by the acquisition of Gibraltar and Minorca and by the partitioning of Spain's Italian lands between the emperor and Savoy, neither of whom possessed a navy. With the additional acquisitions of Nova Scotia, Hudson bay and Newfoundland, the treaty laid the foundations for 50 years of English preponderance in both the old world and the new. It showed also that the Tories fully shared the Whigs' awareness of England's essential strategic and commercial interests.

Politics and Parties.—Anne, though a strong Anglican and Tory, lacked William's political ability and grasp. Government therefore fell more into the hands of her ministers, whom the financing of the war forced to keep in still closer touch with parliament and to take pains to organize support there. At first the leading figures were Marlborough (whose wife Sarah dominated Anne) and Lord Godolphin, the lord treasurer. In domestic politics their greatest achievement was the completion of the union with Scotland. Tory bills against the English dissenters revived Scots fears for the security of their own Presbyterian Church. The failure of their project to colonize Darién sharpened their desire to be admitted to the freedom of the English commercial and colonial system. Their Act of Security (1704) brought matters to a head by stipulating that, after Anne's death, they would not accept the same successor as England unless full security were given for Scottish freedom, religion and trade. The English parliament replied by prohibiting trade with Scotland and treating Scots as aliens, but at the same time authorizing the queen to appoint commissioners to treat with the Scots. From the negotiations which ensued came the Act of Union, which took effect from May 1, 1707. By this England abandoned its commercial exclusiveness and Scotland its legislative independence. Scotland was to be represented in the Westminster parliament by 16 peers and 45 members of the house of commons; but it was to retain its own laws and law courts, while separate acts incorporated in the Act of Union ensured the independence of the English and Scottish churches. The union of Great Britain was thus at last completed.

Godolphin's chief interest, however, was the war with France, for which as the years passed he found better support from Whigs than from Tories. The ministry accordingly assumed an increasingly Whig complexion. Its uncompromising bellicosity and the high taxation it imposed gradually provoked a growing Tory opposition to the war. Questions of religion and the succession added fuel to the flames. The sermons of Henry Sacheverell suggested, indeed, a revival of the old church-and-king Toryism in their attacks upon the 1688 revolution. Sacheverell's trial and the light sentence passed on him in the lords further stimulated Tory feeling, just when the now largely Whig ministry and the duchess of Marlborough were losing their hold on the queen. In 1710 Anne dismissed them and gave office to Robert Harley (*see* OXFORD, ROBERT HARLEY, 1st Earl of) and Henry St. John (*see* BOLINGBROKE, HENRY ST. JOHN, 1st Viscount).

For the rest of her reign party strife rivaled in bitterness the Exclusion bill days under Charles II. The Whigs savagely criticized the government's peace negotiations and began to fix their hopes, like later opposition groups, on the heir to the throne, George of Hanover. This placed the Tories in a dilemma. They could not endanger Anne's good will by competing for Hanoverian favour and the alternative to George was the uncompromisingly Roman Catholic James, the Old Pretender. They therefore tried, as Bolingbroke put it, to "fill all the offices of the kingdom, down to the meanest, with Tories" and to break the Whigs and Dissenters, so that on Anne's death they would be strong enough to dictate their own terms to her successor, whoever it might be. They passed an act requiring a qualification in landed property for members of parliament (1711); another punishing the occasional conformity by which Dissenters took office in defiance of the Test act (1711); and the Schism act (1714) to destroy the Dissenters' schools and academies. Whether this policy could ever have succeeded is doubtful; too many Englishmen were men of the centre for extremist courses to flourish, and even the moderate Tories began to draw back when Anne dismissed Oxford and put Bolingbroke and the high Tories in power (July 27, 1714). Five days later she died, before Bolingbroke could either make his position impregnable or define his attitude on the succession. Thereupon events once again showed how strongly the Tories' attachment to their church committed them to the revolution settlement. For they now allowed the champions of the Protestant Hanoverian succession to thrust them aside and proclaim George I. The long Whig ascendancy had begun. (R. B. WM.)

XII. THE FIRST HANOVERIANS (1714–60)

The rule of the first Hanoverians, George I (1714–27) and George II (1727–60), saw the establishment of the Whig supremacy, the virtual abolition of the Tory party and the end of the Stuarts' hopes of recovering the throne. The executive secured not only the initiative in parliament but also its virtual control. The interests of the landed gentry dominated its policy but not to the exclusion of the commercial aspirations of London and the great trading cities. It was a period marked by few reforms. The condition of the Dissenters was somewhat eased, but they failed to secure the repeal of the Test act which would have given them complete equality of status with Anglicans. A few minor reforms took place in administration; regulations governing trade were relaxed and government finance so well reorganized that Great Britain avoided the financial difficulties that beset other nations in Europe. Although there was little change in the nature of the constitution, the public interest in politics became more extensive because of the spread of education and the growth of a free, vigorous press. In comparison with the previous century, Great Britain enjoyed remarkably stable government but it was never free from political discontent. The main theme of public criticism was of the growth of corruption in politics. Although modern historians have shown that such criticism was grossly exaggerated, and that no ministry could have survived without public support, nevertheless it cannot be denied that the institutions of government were brought into such disrepute that by the end of the period voices were being raised not only for their purification but also for their fundamental reform.

ESTABLISHMENT OF THE WHIG SUPREMACY

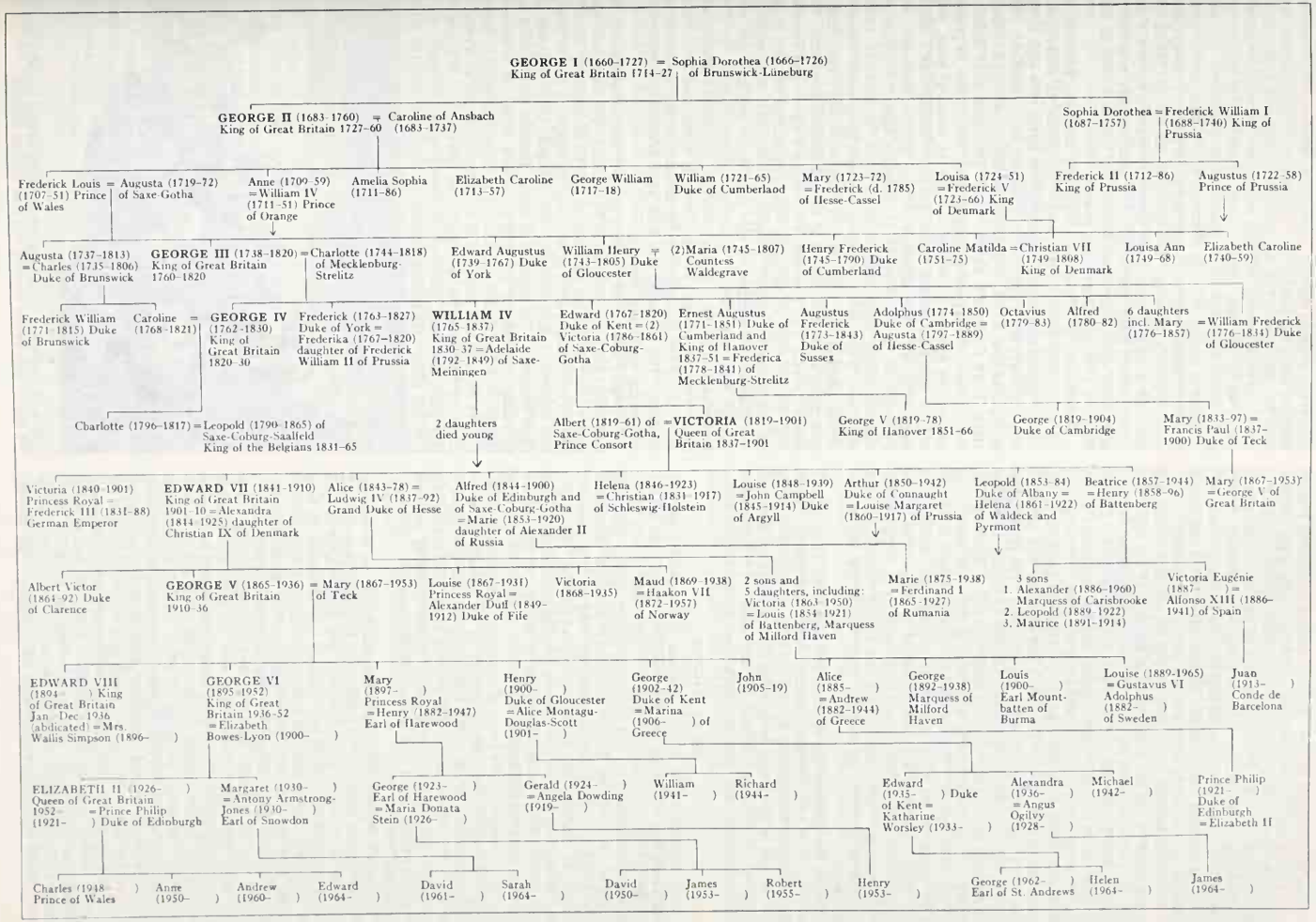
The Jacobite Rebellion of 1715.—In 1714 it was far from certain that the new king, George I, the great-grandson of James I, and elector of Hanover, would secure the succession to the English throne that had been conferred on his family by the Act of Settlement (1701). James II's son James (*q.v.*), the Old Pretender, enjoyed considerable support in Europe, Scotland and northern England. During the last years of Anne's reign, Bolingbroke had pressed James, unsuccessfully, to become a Protestant in order to increase his chances of succeeding the queen. Many Tories, fearful for the safety of the church, had been reluctant to follow Bolingbroke, and they had welcomed George I on his arrival (Sept. 1714) in England. Indeed a few leading Hanoverian Tories had been taken into George I's administration. This ministry, led by the Whigs—the earl of Sunderland, James Stanhope, Lord Townshend and Robert Walpole—after having won overwhelmingly the general election of 1715, set up a committee to investigate the past record of Bolingbroke and his colleague, Oxford. Bolingbroke's nerve failed and he fled to France to become the secretary of state to the pretender who was mounting an expedition to invade Great Britain. This took place in Sept. 1715 and proved a dismal failure. Although the Stuart forces held Scotland for a time, England remained obstinately loyal to the Hanoverians. The Jacobites were defeated at Preston and Sherrifmuir, and James, who had landed in Scotland in Dec. 1715, was soon back in France.

Bolingbroke's alliance with the Stuarts did not last long; after a sharp quarrel in 1716 he quitted their service and began to work for a pardon which he partially achieved between 1723 and 1725.

This Jacobite rebellion stigmatized the Tories as potential traitors, which indeed many of them were, and enabled the ministry to purge itself of their support. Bolingbroke and Oxford were condemned, Oxford being sent to the Tower of London. The insecurity of the times also enabled the ministry to increase the life of parliament from three to seven years (Septennial act, 1716). This certainly made for stability and also increased the wealth derived from those offices granted during pleasure as well as the value of parliamentary property.

Formation of a Whig Opposition.—No sooner had the crown weathered its first difficulties than quarrels broke out between the Whig leaders, partly on personal and partly on public issues. Walpole and Townshend thought that Sunderland and Stanhope were intriguing against them during the king's visit to Hanover in 1716–17, and Sunderland and Stanhope believed that Walpole and Townshend were making a party against them with the prince of Wales and his friends in England. These jealousies were rendered acute by a sharp division of opinion on foreign policy. Townshend disapproved, as contrary to British interests, of Stanhope's desire for friendship with France and his insistence on using British men and money to help George I bring about a settlement in the Baltic. The upshot was that Townshend (who had already been dismissed as secretary of state) and Walpole left the government (April 1717) and went into opposition. Shortly afterward the king quarreled violently with the prince of Wales and for a short time even put him under house arrest. This led to the prince setting up his own court at Leicester house, and his friends and dependents immediately supported Townshend and Walpole in their opposition to the ministry, thereby rendering their opposition more formidable. (*See* also OXFORD, ROBERT WALPOLE, 1st Earl of; STANHOPE, JAMES STANHOPE, 1st Earl; TOWNSHEND, CHARLES TOWNSHEND, 2nd Viscount.)

The government survived, but with difficulty. Stanhope completed Walpole's schemes for putting the country's finances on a sounder basis by consolidating the national debt and instituting a sinking fund to which certain taxes were appropriated in order to pay it off. The ministry was, however, defeated in its attempt to secure greater liberties for Dissenters, to reform the universities and to limit the king's prerogative of making peers. This last measure (the Peerage bill, 1719) Walpole was instrumental in defeating in the house of commons. In foreign affairs, however, in which the king himself took a lively and intelligent interest, the success of this ministry was outstanding. Stanhope brought to a



(1) (2) Next to Name Indicates Order of Marriage

successful conclusion his negotiations, through Guillaume Dubois, with France and the Netherlands, and formed the triple alliance, which was converted to the quadruple alliance by the adherence of the emperor in 1718. This combination of powers was successful in thwarting the ambitions of Spain in the Mediterranean, largely through the action of the British fleet at Cape Passaro. Stanhope's policy in northern Europe neatly blended Hanoverian and British interests and helped to bring about a general settlement.

South Sea Bubble.—The insecurity of the ministry in the house of commons on the one hand, and the inability of the opposition to bring about its defeat on the other, led to negotiations between Walpole and Townshend and Stanhope and Sunderland. The ostensible reason for their reconciliation was the settlement of the quarrel between the prince of Wales and the king. No sooner had Whig unity been achieved than the country was involved in the South Sea Bubble (*q.v.*). The South Sea company had agreed to take over a large proportion of the national debt. The extension of public credit in order to stimulate trade was an excellent idea but the methods used were financially dangerous and at times corrupt. An orgy of speculation led to a collapse of the stock market and the ruin of thousands. The king, his German mistresses and many of his ministers had been involved in dealings that bordered on corruption. Walpole, who had speculated as intensely as any of the public, had not, however, been in the ministry at the time when the scheme had been adopted. He now fought vigorously in the commons to preserve the ministry and, with it, the Hanoverian succession.

During the acrimonious debates, Stanhope died (Feb. 1721), some ministers were disgraced, another committed suicide. These deaths enabled Walpole and Townshend to improve their standing at court and in the cabinet by securing further offices for members of their party. Finally, the death of Sunderland in 1722 gave them a clear supremacy that remained unchanged for the next 20 years.

Age of Walpole

Tory Decline.—The effective defense of the ministry by Walpole meant that the Tories had been unable to exploit the difficulties created by the South Sea Bubble. They fared poorly in the general election of 1722, and the impeachment shortly afterward of one of their leaders, Francis Atterbury, bishop of Rochester, on a charge of treason, further weakened their cause and enabled Walpole to represent the Tories as crypto-Jacobites. From this time onward the support for Toryism steadily diminished; family after family turned to the Whigs and for the rest of the century the Tory party played little part in formal political life and mustered few adherents in the house of commons. Its principles, however, still secured considerable support in the country at large. The unassailable position of the Whig party did not strengthen it but led to divisions within it. The political climate did not become any gentler, and opposition was carried on effectively by dissident groups of Whigs, allied with Tories and independents. (*See* also Whig and Tory.)

Walpole's Domestic Policy.—The stability of the throne was Walpole's prime concern. To achieve this he sought peace abroad, low taxation at home and a powerful, yet loyal, party in both houses of parliament. The first essential was, of course, to keep the support of the crown. A brief crisis followed the death of George I (June 11, 1727), but Walpole quickly gained the support of the new king, George II. Next to royal support, Walpole realized that his greatest need was to control parliament. The basis of his strength in both houses was the court and treasury party which he carefully built up by using appointments to offices at court, in the administration, in the church, army and navy with positive ends in view. In return for office, Walpole expected not only personal loyalty but also assiduous attendance in parliament. By such means he created a powerful group of efficient politicians. He provided these with a policy which also drew support for the ministry from the uncommitted members of the house of commons. Walpole wished to secure the Hanoverian succession and to avoid all situations which might give opportunities for Jacobites (*q.v.*) or their Tory supporters to upset the settlement. Prosperity, he believed, would create satisfaction with the monarchy. He encouraged trade, reduced taxation, stimulated confidence in the national debt and made the cumbersome machinery of administration work with fair efficiency. This he achieved by rigorous application to detail, for he reformed little and neither his economic or financial policy was in any way novel. When his scheme for excising wine and tobacco met with bitter opposition in parliament and acute divisions at court (1733), he discarded it. Political considerations, that is, the security of the throne and of his ministry, always dominated Walpole's decisions. For these reasons, too, in 1724 he withdrew the flood of copper coins known as Wood's halfpence from Ireland where their introduction had been strongly resented; he acted with similar discretion toward the Porteous riots in Edinburgh in 1736 (*see* Porteous, John).

Opposition to Walpole.—Walpole's methods, however, did not meet with universal approval. His choice as colleagues of men of second-rate ability embittered the gifted men such as William Pulteney, Lord Carteret and the earl of Chesterfield, whom he rejected (*see* Bath, William Pulteney, 1st Earl of; Granville, John Carteret, Earl). These politicians were drawn to Bolingbroke, whose pardon Walpole had been unable to prevent, although he secured his exclusion from the house of lords, which at least kept him outside formal politics. Bolingbroke's alliance from 1726 with the dissident Whigs, and in particular with Pulteney, led to a brilliant press campaign against Walpole who was mercilessly lampooned in newspapers, broadsheets, ballads and plays. Pope, Swift, Fielding, Gay and many other writers of distinguished ability took part in Walpole-baiting from time to time, much to the enrichment of English satirical literature. *Gulliver's Travels, The Dunciad, The Beggar's Opera* and *Jonathan Wild* are but a few of the remarkable productions inspired in part by this campaign against Walpole. After 1733, this opposition was joined by a number of able politicians who had lost office through their criticism of Walpole's excise scheme. In 1737, Frederick Louis (*q.v.*), prince of Wales, quarreled furiously with his parents, set up his court at Leicester house, and so increased the ranks of the opposition by the addition of his supporters and servants. This happened at a time when Walpole's foreign policy was facing increasing criticism.

Foreign Policy.—In the early years of Walpole's ministry, foreign affairs had been directed by Townshend. He continued Stanhope's policy of alliance with France and by the treaty of Hanover (1725) succeeded in more or less isolating Spain. A period of near war followed, indeed Gibraltar was besieged, which Walpole disliked because of the expense of hiring German mercenaries and providing subsidies for the king's German allies. A settlement, largely of Walpole's making, was achieved by the treaty of Seville (1729) and this was extended in 1731 by the treaty of Vienna to cover the difficulties between Spain and the empire. Walpole felt that peace had been achieved by relying upon the good will of Cardinal Fleury, the leading minister of France; indeed the Anglo-

BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

"EXCISE IN TRIUMPH," A CARTOON DEPICTING SIR ROBERT WALPOLE DURING THE EXCISE BILL CRISIS OF 1733. IN THE BRITISH MUSEUM

French alliance, strongly pressed by Walpole's brother, Horatio, the ambassador at Paris, was the core of Walpole's foreign policy. The pursuit of this general settlement, almost regardless of consequences, led to a quarrel between Walpole and Townshend who in the end resigned (May 1730). Nor was Townshend the only critic of Walpole's foreign policy; many, including the majority of London's leading merchants, thought that too many concessions were being made to Spain and that, in the end, this would lead to Great Britain's losing the rich Spanish colonial trade to France. As William Pitt told the house of commons: "if trade is at stake, you must defend it or perish." This "patriotic" opposition so worried many of Walpole's colleagues that they inclined to a war against Spain. They also realized that Walpole was an aging man, and that the future would bring some of the opposition to power.

In the end Walpole reluctantly yielded to their persuasion and brought the protracted negotiations with Spain to a close in 1739. Although the so-called War of Jenkins' Ear (*see* JENKINS, ROBERT) marked the defeat of Walpole's foreign policy, he remained in office for another three years.

The Pelhams

Political Maneuvers.—At the election of 1741 Walpole's majority was much reduced; his friends, although doubtful of the future, remained loyal but he was finally forced to resign (Feb. 1742). The king created him earl of Orford; the house of commons' attempt to impeach him, although it failed, established the important precedent that the house of commons itself is, in a sense, responsible for the policy pursued by a ministry. Walpole's bitterest opponents entered office, along with some of his old colleagues. A short period of very confused political maneuvering followed, caused primarily by the highhanded conduct in foreign affairs of Carteret who enjoyed the wholehearted confidence of George II. Skilfully advised by Walpole, the duke of Newcastle and his brother Henry Pelham (*qq.v.*) forced the reluctant king to dismiss Carteret (now Earl Granville) in Nov. 1744. In Feb. 1746, by means of the threat of mass resignation, they and their fellow ministers forced the king to employ the elder William Pitt, whom he loathed for his attacks on Hanover. This, as Walpole forecast, drew the sting of the opposition and gave the ministry a secure majority in the commons. Under the leadership of these two brothers, Walpole's old corps of Whigs, supported by William Pitt and other former dissident groups, maintained themselves in power until the death of Henry Pelham in March 1754. Pelham, Walpole's choice as his successor, proved himself to be a skilful if somewhat humdrum leader of the house of commons; his brother, the duke of Newcastle, busied himself with foreign affairs with greater competence than any of his contemporaries allowed him. Their ministry was made difficult by war and invasion. The war against Spain had developed into a general European conflict, the War of the Austrian Succession, on the death of the emperor Charles VI. In spite of the pragmatic sanction by which the major powers had guaranteed the inheritance of his daughter, Maria Theresa, Prussia under Frederick II (the Great) seized Silesia. Anglo-French rivalry in North America and India added an extra dimension to this conflict, in which Great Britain at first enjoyed modest successes and George II distinguished himself at the battle of Dettingen (1743).

The '45.—In 1745 Prince Charles Edward (*q.v.*), the Young Pretender, landed at Moidart in Scotland, rallied the Highlands to him and took Edinburgh. A swift invasion of England carried his forces to Derby (Dec. 4), but, fearful of his extended communications, he decided to retreat and it was only a matter of time before the stronger Hanoverian forces crushed the Jacobites. George's son, the duke of Cumberland, won an absolute and bloody victory at Culloden on April 16, 1746. After weeks of pursuit Prince Charles Edward escaped to France. The Highlands were savagely pacified; the clan system was destroyed, the power of the chieftains broken and even the tartan forbidden, roads were driven through the mountains and forts constructed. In consequence, emigration to America sharply increased. The outcome of "the '45" had never been in doubt. Few of the Lowland Scots rallied to Charles Edward, and even fewer English. This danger passed, the European war dragged on without decisive actions and an inconclusive peace treaty was signed at Aix-la-Chapelle in 1748.

Domestic Policy.—By this time the national debt had grown to what Pelham regarded as dangerous proportions. He was worried, too, by the high rate of taxation. For the next six years he pursued a policy of peace and retrenchment. He reduced the army and navy establishments, cut subsidies to foreign prices, lowered taxation and also brought about the final consolidation of the national debt by the issue of 3% Consolidated Annuities or consols (*q.v.*). These measures stimulated a quick recovery of the country's economy and attracted a great deal of Dutch capital which induced further prosperity. Pelham also supported the reform of the calendar in 1751 and attempted, without success, to liberalize public policy toward Dissenters and Jews. By 1754, when he suddenly died, Britain's internal position was stronger than it had been since the days of Walpole; in foreign affairs, however, the situation was growing ever more grave.

Newcastle and Foreign Affairs.—The duke of Newcastle had steadily pursued the traditional foreign policy of Great Britain, the key to which was the alliance with Austria and the German princelings as a counterpoise to the ambitions of France. Newcastle, too, was willing to subsidize Russia heavily in order to check France's principal ally, Prussia. Indeed, he did this so effectively that Frederick the Great questioned the wisdom of a continued French alliance. Newcastle exploited these doubts with alacrity and brought about a firm Anglo-Prussian alliance at the convention of Westminster (Jan. 1756). The Austrian chief minister, W. A. von Kaunitz, had for some time doubted the value of the British alliance and the wisdom of enmity with France. Russia immediately opened negotiations with Austria. The upshot was the famous Diplomatic revolution by which France, Austria and Russia entered into league against Prussia, England's sole European ally of any importance. England's position was one of weakness when the Seven Years' War (1756–63) began.

The Seven Years' War

The European war, however, was already complicated by Anglo-French rivalry in North America, where hostilities had broken out on the frontier of Virginia as early as 1754. The disastrous defeat of Gen. Edward Braddock in 1755 helped to precipitate a full-scale war, the French and Indian War (*q.v.*), between France and England in North America. The position in India was similar to that in North America; the truce of 1748 had already broken down by the time the war in Europe began. Furthermore, the political situation in Great Britain had rapidly deteriorated after the death of Henry Pelham.

Pitt's Rise to Power.—The most gifted member of Pelham's ministry had undoubtedly been William Pitt (*see* CHATHAM, WILLIAM PITT, Earl of). From his first entry into parliament he had deeply impressed his contemporaries, particularly the merchants of London, by his magnificent oratory and his deep conviction that Great Britain's destiny was to be the greatest trading nation in the world. His bitter attacks on the government had led Walpole to advise Henry Pelham to buy him off and bring him into the government, which Pelham did in 1746. After a short while as vice-treasurer of Ireland, Pitt became paymaster general of the forces and showed great skill as an administrator, much originality as a strategist and outstanding powers as a spokesman for the government in the house of commons. During his time as paymaster general Pitt had absolutely refused to avail himself of the perquisites of his office which had been the foundation of the fortunes of many of his predecessors, including Walpole. The greatest profits of this office were the right to invest the army funds at a personal profit and to extract considerable sums from each subsidy paid to foreign princes. Pitt's refusal, on moral grounds, to avail himself of these rights immensely strengthened the regard with which the public held him. Unfortunately, however, George II had never forgotten Pitt's attacks on Hanover, and so he remained only paymaster throughout Pelham's ministry. At the death of Henry Pelham, Pitt confidently expected to become at least leader of the house of commons, but after a year of complicated negotiation he was passed over for his rival, Henry Fox, a

capable, hard-working but undistinguished politician who was a close friend of the duke of Cumberland.

Pitt then began to frequent Leicester house, still the home of the opposition despite the death of the prince of Wales in 1751, and his break with the ministry became only a matter of time. He attacked Newcastle's foreign policy, particularly the subsidy treaties with German princelings. "These traitors," he said, "will hang like a millstone about the neck of any minister, and sink him along with the nation." He also called for stronger support for British colonists in North America, those "long-injured, long-neglected, long-forgotten people." The burden of his argument was to sacrifice Hanover if need be, but support true British interests to the death.

Although Pitt was immediately dismissed, his attitude secured his great popularity in London and the country at large. It also brought him a considerable following among the independent members of the commons.

Outbreak of War.—The Seven Years' War (*q.v.*) played into Pitt's hands. Beset by three formidable armies—the French, Austrian and Russian—Frederick the Great was hard pressed to defend his own kingdom. The French, fully mobilized, had ample troops to spare: they massed in the Pas-de-Calais, threatening invasion of England. They took the initiative in the French and Indian War in North America and gobbled up the island of Minorca, Britain's chief naval base in the Mediterranean. The arrest, trial and execution of Adm. John Byng proved an insufficient sop to public resentment at the loss of Minorca and did not save Newcastle's ministry, which fell in 1756. The initial disasters of the war destroyed confidence in the duke of Newcastle's ability to conduct the war to a successful conclusion without Pitt.

Pitt took office (Nov. 1756) with his erstwhile rival Henry Fox, but, without the support of the old corps of Whigs who followed Newcastle, the ministry had little hope of continuing long in office. The end came when the duke of Cumberland, the commander in chief of British forces in Europe, refused to accept orders given by Pitt. This led to Pitt's dismissal (April 1757) but it proved easier to dismiss Pitt than to replace him. Cumberland, known as "Butcher" for the excesses of his troops at Culloden, was a hated man and his part in bringing about Pitt's fall merely added to the public's support for the fallen minister. Led by London, city after city conferred their freedom on Pitt. At last, stimulated by the fear of their rivals, both Pitt and Newcastle awoke to their need for each other. By the end of June 1757 negotiations had been completed and Pitt took office for what was to prove one of the most critical ministries in modern English history.

Pitt's Great Ministry.—The direction of the war became Pitt's charge; appointments were left, much to his joy, to the duke of Newcastle who, as first lord of the treasury, also had the more grievous burden of finding the money to pay for the war. As early as Oct. 1756 Pitt told one of his contemporaries, probably the duke of Devonshire, "my lord, I know that I can save this country, and that no one else can." His intentions, however, were not merely to save his country from defeat but, if possible, to crush France and thereby seize the trade of the world.

Pitt had formulated many of his strategic plans with his friends among the financiers and merchants of London. His appetite for factual information was immense, and all his moves were strictly calculated on a basis of profit and loss. He believed that his strategy could be made to pay for itself by capturing France's most profitable colonial ventures. Trade, however, was not Pitt's sole concern. He wished to deny France strategically valuable ports, naval supplies and training grounds for her seamen. His desire for victory was absolute: it was to be for all time and France would lack the means of becoming formidable again. There were four major areas of conflict, outside Europe—Africa, North America, the West Indies and India: in each Pitt intended to destroy French power.

Pitt's ministry, however, began inauspiciously and 1757 was marked by a series of defeats. Cumberland in Germany, overwhelmed by the French, negotiated the convention of Klosterseven (Sept. 1757) which was tantamount to abandoning Frederick the Great. Although this settlement was immediately repudiated, and Cumberland disgraced (even George II was incensed by his son's actions), nevertheless Frederick the Great's western flank was dangerously exposed. In Canada, the British failed to take Louisburg, the gateway to the St. Lawrence. In India, the nawab of Bengal demonstrated what he thought of British chances by seizing Calcutta and throwing his captives into a tiny prison, the notorious Black Hole; the majority died. The prospects for Britain did not look encouraging.

Pitt quickly concluded a new treaty with Frederick the Great, sending him large subsidies, numerous mercenaries and a few English regiments. Then he devoted himself to his major strategy. To help his continental allies, he mounted a number of raids on the French coast: these were expensive in men and materials, unpopular alike with the public and the armed forces, and scarcely troubled the French. The Prussians, however, believed in their efficacy and they kept Frederick the Great reasonably content. Much more serious for the French was the unfolding of Pitt's grand strategy. Trading posts on the west African coast were brilliantly captured; Louisburg in Canada fell and so did the most important of all French forts—Duquesne, the citadel that lay to the west of the colonies on the Ohio river in North America. Frederick the Great, also, during 1758 fought a series of brilliantly defensive battles and stopped his enemies from overrunning his country.

These victories of 1758, however, paled into insignificance before the triumphs of 1759, the *annus mirabilis*. During this year, Quebec was taken by a brilliant and daring maneuver by Gen. James Wolfe. The capture of Canada denied to France the valuable fur trade, important naval stores and the splendid Newfoundland fishing grounds which were a nursery for French seamen. This victory also promised security for the future of the North American colonists. The fall of Gorée Island off Cape Verde (Dec. 1758) gave to Britain the only deepwater harbour on the west African coast; the rich gum trade, and the even more profitable traffic of the French in slaves, fell into British hands. In the West Indies the important sugar island of Guadeloupe was taken by the British navy and, within a few months, its trade had so prospered that increased customs paid the cost of the expedition. As Pitt had forecast, trade could be made to flourish by war. In India, too, victory had followed upon victory—Robert Clive, an East India company clerk of military genius, had rapidly retaken Calcutta, annihilated the nawab of Bengal at Plassey (1757) and gone on to conquer the whole of Bengal. Farther south, Eyre Coote, with fresh troops sent by Pitt, had begun to drive the French from the Carnatic region along the eastern coast of India. In Europe the triumphs were equally spectacular. British troops won a victory at Minden and the British fleet crushed the French in two daring battles—off Lagos, Port., and at Quiberon bay (*see* QUIBERON BAY, BATTLE OF).

The year 1760 saw no diminution in the number of victories. Pitt, who had always shown reverence for his own abilities, began to treat his contemporaries with contempt and to plan ever grander adventures, including the conquest of the eastern seas. Even if George II had not died (Oct. 25, 1760), it is unlikely that Pitt would have continued much longer in office, in spite of the brilliance of his victories. Many feared that these extravagant successes would turn all Europe against Great Britain, while most members of parliament found the ever-increasing burden of taxation irksome. Pitt's concept of empire was still unattractive to many who desired a more modest future. During these years, however, and in spite of the vast concessions made to France in the treaty of Paris (1763), Great Britain became a world power for the first time in its history. (J. H. Pl.)

XIII. REIGN OF GEORGE III TO 1789

George III (1760–1820) inherited from his grandfather a great war which was almost won and a political system which seemed stable. Beneath the surface, however, were explosive issues which speedily destroyed the calm of the past 40 years. The fixity of the social system, the decentralization of power to local magistrates, had left parliament little controversial work over the past decades. The old labels of Whig and Tory were still worn but their meaning was worn out. The king's ministry was made up by using

royal patronage to assist a selected number of those great aristocrats who had the brains and the personal following of dependent politicians to aspire to office; it had sufficient votes in parliament to carry on routine business. The independent members of parliament, who were in a majority, were suspicious of this ministry but they were, in normal times, inert. The conflicts that arose after 1760 involved issues of principle. The personal quarrels of leading politicians became mixed up with these issues. A disgruntled leader naturally appealed to the majority of independent country members against his successful rivals. In this way new party alignments began to form and the king's authority as a patriotic head of the government to be questioned.

War, not the king, was responsible for this change in politics. The burden of taxation (with the land tax at 4*s.* on the £) provoked grumbles from the smaller squires. The gains made overseas by war presented the government with a need to evolve a policy of empire for which it had no precedent. The ending of the war (in 1763) left England isolated in Europe and with no clear ideas of how such isolation could be ended. At the same time economic dislocation, in changing from war to peace, aggravated the perpetual problem of maintaining order and suppressing riot.

Political Instability, 1760–70.—The ministry in power in 1760 was that of Pitt and Newcastle. It was a powerful combination because Pitt was the organizer of war and the author of magniloquent phrases while Newcastle was the master of managing a parliamentary team by the distribution of favours. But Pitt and Newcastle disliked one another. The king trusted neither and put his friend the earl of Bute (*see* BUTE, JOHN STUART) into office to control them. Pitt resigned in Oct. 1761 because he was not allowed to enlarge the war by an attack upon Spain. Newcastle followed into uneasy retirement because he could brook no suspicion of interference with his control of the treasury. Bute's administration concluded the peace of Paris in 1763 at the cost of antagonizing England's ally Prussia. The peace settlement gained Canada at the cost of restoring to France her West Indian sugar islands. France had her posts, but not her influence, restored to her in India. The peace was welcome to the majority of Englishmen but was attacked by Pitt as unworthy of such brilliant efforts in war. The ministry also had to face a storm of protest against the introduction of a tax on cider in 1763.

Bute resigned in April 1763 and the king was forced to arrange a ministry under George Grenville (*q.v.*), who was efficient but bullied both the king and the house of commons. Opposition was given new ferocity by John Wilkes (*q.v.*), member of parliament for Aylesbury. In No. 45 of his paper the *North Briton* Wilkes not only assailed Bute and Grenville but accused the king himself of untruthfulness. An attempt to stifle him by the seizure of his papers under the authority of a "general warrant" from the secretary of state, Lord Halifax, enabled Wilkes to pose as a champion of freedom. His answer to a charge of seditious libel (soon followed by one of obscene libel arising from his *Essay on Woman*) was that general warrants of the type used against him were illegal and that members of parliament must have complete freedom of expression. Wilkes fled to France but had roused the suspicion that the king's ministry wished to increase its powers at the expense of traditional liberties.

George Grenville stirred up even worse trouble in his attempts to lighten the burden of taxation once war was over. The North American colonists had been relieved from a French menace by the conquest of Canada. Should they not therefore bear part of the cost of war? The cost of their protection, Grenville argued amid applause, should not be subsidized by English squires. So he increased the efficiency of the collection of the sugar duties and imposed a stamp duty upon the American colonies in 1765. This provoked the colonists to protest, in their turn, against an increase in the power of the government. They rioted against direct taxation without representation (*see* UNITED STATES [OF AMERICA]: *History*).

The king rid himself of the autocratic Grenville in 1765. The only possible ministerial team was that of the marquess of Rockingham, who had inherited Newcastle's influence. Pitt refused to join such a combination; his demand for complete control made it impossible to build up a satisfactory ministry which might settle the problem with the colonies. Rockingham compromised by passing a Declaratory act to affirm England's right to tax the colonies and then repealing the Stamp act, the obnoxious example of the use of such power. From this political compromise, to meet a particular situation, Edmund Burke (*q.v.*), Rockingham's publicist, was to evolve a general theory of empire. The final sovereign power must reside in Westminster; but it was inexpedient to use this power except in an emergency such as war. The bond of empire was not in law but in affections so "magnanimity in politics is not seldom the truest wisdom and a great empire and little minds go ill together."

Pregnant for the future, such aphorisms were powerless to solve the problems of the moment. Rockingham's government, divided internally and suspected by the king, resigned in 1766. Pitt, who now became earl of Chatham, next made an attempt to defy the factious combinations of leading politicians. He attempted to base his administration upon "measures and not men." He attacked three problems: the East India company's acquisition of Bengal by Robert Clive's treaty of 1765, Great Britain's isolation in Europe and the situation in America. Unable to control the intrigues of his ministers in the East India company's affairs, he failed to assert the crown's control over the company's new Indian territories. He also failed to persuade Prussia to negotiate with a view to alliance. More important still he failed to persuade the American colonies to contribute voluntarily to imperial costs or even to provide for troops used for their own security. Chatham became insane; he retired in 1767 and eventually resigned in 1768. His admirer the duke of Grafton carried on the government for two more years. In these years it became clear that an American minority wished to provoke a quarrel with England while the majority, loyal as they thought themselves, were too remote from London to suffer any effectual interference with their destinies. This was made clear by their rejection of the disastrous attempt of Charles Townshend (*q.v.*) to raise money by indirect taxation of glass, paper, tea and other imports and by their anger at attempts to limit their expansion westward into Indian territories.

Grafton allowed the quarrel to drift. His only positive achievement was domestic. He did battle with Wilkes. Wilkes had returned to England to meet the charges against him in 1768. He was returned to parliament as member for Middlesex. He was then expelled, as an unfit person, by the house of commons on the advice of the ministers. Wilkes was elected for Middlesex three times: on the last occasion the house of commons declared that the defeated candidate Henry Luttrell had been elected. The opposition could now claim that a house of commons, in which government influence was strong because of patronage, had set aside the popular will. New force was given to the legend, which had existed since Bute's resignation, that George III was attempting to subvert the constitution and rule through his personal friends. Burke expressed the opposition line and indicated future trends in his *Thoughts on the Cause of the Present Discontents* (1770) when he denounced the influence of the "king's friends" and advocated government by political parties united on grounds of principle. When in 1771 the house of commons attempted to exclude reporters and to enforce the ban on publishing reports of their debates the irresponsible cry of "freedom in danger" was again raised. The Middlesex election crisis had revealed new and articulate forces of opposition outside parliament and it marked the beginnings of a radical program for parliamentary reform.

Lord North and the American Revolution.—The immediate significance of these events was slight, however, for George III found (1770) in Lord North (*see* NORTH, FREDERICK NORTH; GUILFORD, BARONS AND EARLS OF) a more tactful minister who was able for a time to rebuild the old noncontentious system. He met the fiscal problem by economies. Foreign policy was to be conducted on pacific lines so as to allow for economy in the armed forces. The problem of India was solved for the time being by a Regulating act (1773) which left the East India company most of its power but allowed the king to nominate a governor general in Bengal. Wilkes was quietly permitted to take his seat in 1774. North excluded no one from his ministry on lines of principle. His

parliamentary good humour lulled the independent members back into their former indolence. He wished no quarrel with the American colonies and repealed the Townshend duties on goods imported into America, with the exception of that upon tea: this was left as a symbol to content the English backbenchers. It was used also to help the East India company by controlling sales of tea in the new world through licensed dealers. North's policy was successful until 1774. From that time onward he was caught between colonial defiance of the tea legislation and English unwillingness to allow the Americans to escape all control. A shipment of tea was thrown into Boston harbour in 1773 and North, obliged to restore respect for the law, suspended the constitution of Massachusetts. When blood was shed at Lexington and Bunker Hill in 1775 by British troops facing the colonists, a war was inevitable.

Lord North fought the war halfheartedly, aware that his strength lay in conciliation, not in conquest. The nature of his support at home, however, made concessions to the Americans politically difficult. With the meeting of the second continental congress which, on Dec. 6, 1775, disavowed allegiance to parliament and the signing of the Declaration of Independence in 1776, colonial feeling had a centre for resistance. The last hope of British success was to cut off the New England rebels from those in the rest of the colonies. Gen. John Burgoyne's plan to do this failed when, advancing from Canada toward New York, he was surrounded and forced to surrender at Saratoga (1777). His failure had been due, in part, to the vagueness of the instructions sent to Gen. Sir William Howe in New York by the secretary of state Lord George Germain. But to fight this war at all was difficult. It depended upon the co-ordination of the movements of armies and fleets across 3,000 mi. of water by a secretary of state handicapped by slow communications and imperfect intelligence (*see* AMERICAN REVOLUTION).

After Saratoga the British commanders tried in vain to bring George Washington to fight a pitched and decisive battle. They were reduced to holding strong points in the north while roaming commanders, such as Lord Cornwallis, marched through the southern states. Cornwallis had brilliant success but nowhere did he permanently occupy territory. The rebels resumed control as his army marched on. Such operations might have continued as long as the British treasury could find armies. But from 1778 France intervened on America's side. The temporary loss of control of the sea to the French fleet left Cornwallis facing American and French troops in Virginia. In 1781 he surrendered at Yorktown and the war in America was lost.

North's Fall and the Reform Movement.—The king wanted to fight on. But the politicians saw that they might not only lose America but be defeated in the war which had opened nearer home. A hostile fleet was in the English channel; Gibraltar was besieged by Spain. North resigned in 1782 in favour of a ministry under Rockingham which was pro-American and included Chatham's followers now led by the earl of Shelburne (*see* LANSDOWNE, MARQUESSES OF). Peace was made as quickly as possible, first with America (by granting independence) and then with the European powers.

Lord North's collapse had not only been caused by defeats in America. In Ireland too the situation was menacing. The "Volunteers," though their patriotic motive had originally been to defend the country and allow troops to be sent to America, had become an armed pressure group demanding legislative independence for Ireland (*see* IRELAND: *History*). This too Rockingham conceded. In England itself a new reform movement had been organized. It was popularly supposed that North was responsible for the failure of the war, for the continued high rates of taxation and for the dislocation of trade. Yet he continued, until 1782, to command parliamentary majorities. This proved, it was argued, that parliament was corrupt. The Petitioning movement, strongest in Yorkshire where Christopher Wyvill led the Yorkshire association, therefore pressed for reform of parliament to diminish the influence of the crown and to increase county representation. Parliamentary opposition had been given new life, both by the failure of the war and by adopting the cause of "economical reform." In 1780 a resolution introduced by John Dunning, later Baron Ashburton (*q.v.*), was carried that "the influence of the crown has increased, is increasing and ought to be diminished."

This agitation inside and outside parliament was temporarily checked in June 1780. In that month the eccentric Lord George Gordon led petitioners of his own to Westminster to protest against the greater tolerance the government had recently shown toward Roman Catholics. In the resulting riots prisons were opened, houses burned, distilleries stormed; the whole of London was in a state of revolution for a week until the troops restored the civil authority. This alarmed some opposition leaders about the dangers of using popular passion against North. Nevertheless they continued to press for a reduction of government influence in elections and to attack, by implication, the power of George III. On North's fall, therefore, the Rockingham administration passed acts disfranchising some government employees and reforming the king's expense account or Civil List (*q.v.*).

The Fox-North Coalition.—Rockingham's death in July 1782 (before peace had been concluded) revealed a quarrel between Charles James Fox (*q.v.*), a leader of the old Rockingham faction, and Shelburne, now chief of the Chathamites. Shelburne was a radical reformer, but he extolled royal power and was chosen by the king to succeed Rockingham. Fox therefore planned, after resigning, simultaneously to destroy his rival and to give the king a hard knock. This was achieved by an alliance between Fox and North which brought down Shelburne's government by an attack on the terms of the treaty of Versailles (1783). The new Fox-North ministry (under the nominal control of the duke of Portland) was condemned as an "unholy alliance." Fox was a spokesman for reformers; North, even after his humiliation, was still trusted by the conservatives who had been alarmed by the American Declaration of Independence and by radical pressures in England. The new ministry was unanimous on few things. But it felt it was safe to act on India.

Under Warren Hastings as governor general British power in India had been maintained, in spite of French intrigue, through wars against the Marathas and against Haidar Ali of Mysore. The East India company was near bankruptcy owing to the cost of the war, although its servants had grown rich. Corruption was known to be rife. For the crown to take over the control of India would have meant increasing royal influence by control of Indian patronage. Fox's India bill (1783) proposed a control by a board of commissioners who he would appoint and who would have a fixed tenure. Popular indignation against this was aroused by the suggestion that Fox's plan was to increase his personal power through the use of patronage in India. George III had chafed under the control of the coalition. In Dec. 1783 he persuaded William Pitt (*q.v.*), Chatham's younger son, aged 23, to support him. The king then brought about the defeat of the Fox-North ministry by ensuring the defeat of the India bill in the house of lords and appointed Pitt his chief minister. Pitt withstood hostile majorities in the house of commons until, in March 1784, a general election returned a majority favourable to him.

Pitt's Ministry: Peace, Retrenchment and Reform.—The successful replacement of Fox by Pitt was the highest point of George III's personal power. It was reached because Pitt had the support of country gentlemen, tired of political faction, as well as treasury influence with which to win the election. Pitt was a reformer but was supported by many who hated reform. The Foxites opposed him, not because of his ideas but because he had assisted the king. Pitt had no party behind him: he had to rely upon the support of those who feared party as disruptive of patriotic unity. He was therefore unable to effect many reforms he supported. He proposed parliamentary reform but did not push it as a government measure. He maintained his friendship with those pressing to abolish the slave trade, but gave them no government help. He proposed, but was forced to abandon, measures to liberalize the trade of Ireland. He did pass an India act (1784) which imposed state control upon the East India company. The act set up a board of control for India to oversee the political but not the trading activities of the company: it was not to exercise patronage: its members were appointed and dismissed by the crown like any officeholders.

Pitt maintained himself by his hold upon the king (who had no alternative to him) and upon both the uncommitted and the officially minded members of parliament. As time went on he increased his strength as conservatives of North's old following came over to his side. He remained, all the same, a reformer. He gradually increased the efficiency of the civil service, cutting out sinecures and so reducing corrupting patronage. In the fiscal sphere he established the consolidated fund into which all customs and excise returns were to be paid. By his Commutation act of 1784 he reduced indirect duties (but enforced collections more stringently) and increased the taxation of wealth by an increased window tax. His principle was to tap the increasing commercial wealth of the country. He was so successful that by 1786 he was in a position to set up a sinking fund of £1,000,000 a year for the repayment of the national debt (*see* GREAT BRITAIN: *National Finance*). Meanwhile, to expand trade, in 1786 he negotiated with France a commercial treaty which was planned to benefit England's growing industries.

Pitt's prestige from such achievements gave to him as first minister a dominance hitherto unknown; his cabinets were subservient to him. His continued program depended upon continuing peace and retrenchment. Eventually he hoped to carry, in a more tranquil country, the whole of his youthful program of reform. His foreign policy was limited to finding, in the triple alliance with Prussia and the Netherlands, friends to help in keeping the Netherlands out of French hands. In face of such success Fox's party was nonplused. The outbreak of the Revolution in France in 1789 was radically to change the situation, but Pitt did not at first perceive that it could affect his smooth progress. (J. S. WA.)

XIV. ECONOMIC AND SOCIAL DEVELOPMENT, 1603–1789

It is not difficult to present the beginning and the end of this period as belonging to widely different economic worlds. James I came to rule an England which, for all that had happened in the 16th century, was little changed in the essentials of its economic life from that to which Henry VII had succeeded over a century earlier. By contrast, the year of the French Revolution—an event which, like the accession of James I, had little direct but some indirect influence on English economic history—was a witness to the dawn of the Industrial Revolution (*q.v.*). This picture of change has a ready appeal and some truth. But the more prosaic truth of continuity must not pass unheeded. Agriculture still fashioned most people's lives, wood and sail still carried all sea-borne trade, men could still travel on land no faster than the horse could take them, most cloth was still woven on looms that had altered little over many centuries. Such testimonies to the slow pace of change must be set against the spectacular achievements by which England had gained—and nearly lost—an empire and had passed into the van of economic advance.

During the 180 or so years of this period the population of England and Wales perhaps about doubled, reaching 9,000,000 at the first census of 1801. In the first two decades of the 17th century total numbers were probably still rising but the upsurge of the previous century petered out in the series of Europe-wide epidemics which attacked England from the 1630s to the 1660s, culminating in the famous plague of 1665–66. Reasonably accurate estimates put the total numbers at about 5,500,000 at the end of the 17th century, and it was probably not until about the middle of the 18th century that notable increase set in again, this time to be long sustained. The reasons for that increase still remain the subject of controversy. It seems probable that initially it depended little upon early developments in medical skills, for what advances there were at that time could have had only a trifling effect upon the high mortality rates, especially among infants; that the virtual disappearance of the plague allowed the effects of a moderate but sustained rise in the birth rate to make themselves felt; that this rise itself owed something to earlier marriages and increased fertility; and that the increased production and variety of food had some effect.

By 1789 more people lived in towns than they had in 1603. London, despite its growth from 420,000 in 1690 to not far short of 1,000,000 at the end of the 18th century, did not house any significantly different proportion of the total; the change came with the rise of such towns as Manchester, which, from a mere village in 1603, had become a city of over 50,000 in 1790.

AGRICULTURE

Both at the beginning and at the end of this period many of the people worked on the land. In some areas in the 17th century the total was as much as 70% or more; elsewhere, in such counties as Gloucester, with its rural cloth industry, or Kent, with its diversified economy, the proportion was nearer 40%–50%. These proportions had fallen by the later 18th century, especially in the growing industrial areas of Lancashire and the west Midlands. But estimates are hazardous. Multiple occupations were far from uncommon: farmers were sometimes also millers or maltsters; and many a rural labourer and his family augmented their agrarian earnings by spinning and weaving.

Improved Techniques and Reorganization.—Agriculture itself was slowly changing, both in organization and in techniques. Important new methods of cultivation gradually made their appearance in English farming; they owed much to techniques practised in parts of the Netherlands. The vital features of the improved agriculture were the replacement of the traditional fallow by root crops such as turnips, which served as winter fodder for livestock, and the growing of "artificial" grasses, such as clover and lucerne (alfalfa), capable of restoring nitrogen to the soil. Combined in rotation with the customary arable crops, the new methods were especially suitable on light soils. The improved supply of foodstuffs for cattle also encouraged a new interest in systematic stock breeding. It is known that on some estates in Norfolk, Suffolk and Kent the new methods were being practised in the later 17th century, but their adoption by the farming community as a whole was slow and erratic, and certainly incomplete by 1789.

In the course of the 18th century, however, "improvement" became increasingly fashionable among the gentry. It was advocated in sundry books, ranging from those of Jethro Tull (*q.v.*), early in the century, whose ideas offered a typically mixed bag of usefulness and absurdity, to those of Arthur Young (*q.v.*), whose diverse and indefatigable activities led to his becoming the first secretary of the new board of agriculture in 1793. Improvement was stimulated by the efforts of a few exceptionally active, and sometimes aristocratic, landlords like Thomas Coke, earl of Leicester (*q.v.*), and carried out gradually and unspectacularly by substantial farmers. (*See* also AGRICULTURE: *British Commonwealth.*)

The adoption of new methods was aided by the continuing enclosure of land—commons and open fields, as well as waste. In the 17th century this was often achieved by agreement between landlords and tenants; in the 18th the private act of parliament was the increasingly favoured device, large numbers being passed especially from the 1760s onward. Enclosure itself was not synonymous with improvement; it facilitated but did not guarantee better farming. Nor are its social effects susceptible to simple generalizations. Sometimes it brought greater employment on larger and more efficient farms; sometimes it bore harshly upon the poorer tenants who found themselves with inferior holdings and without their common rights. But it certainly speeded the already decaying manorial system to its final demise, and helped to consolidate in its place the characteristic English rural social structure: the landowner with his estate, the rent-paying tenant with his farm and the wage labourer with his cottage. Though here and there the structure was modified by freeholding owner-occupiers, "yeomen" of varying status, the general pattern was clear and it was to remain so for years to come.

In the course of this period the output of agriculture produce was increased and diversified. To the cereals—which made not only the Englishman's bread but also the growing quantities of beer and spirits which he drank—were added more meat, butter and cheese as well as a widening range of vegetables; English wool still provided much of the raw material for the cloth industry; more and more skins and hides were made into that important and

versatile material, leather; and the meagre forest resources of the country were much in demand, for ships, houses, firewood and charcoal. From about the 1670s an export trade in cereals developed. Encouraged by bounties it averaged about 500,000 quarters annually. In the 1770s, however, when rising population at home turned the tide of demand, the country became a net importer of grains. It is uncertain how far increased output resulted from new methods and how far it was simply a product of bringing waste lands under cultivation. It seems likely, however, that for most of the period and despite the known and often-publicized improvements, it was the extension of the cultivated area that was the more important. The drainage of large areas of the Fens during the 17th century was only the most striking example of a process by which moor and marsh, heath and woodland were being turned into enclosed farms.

Industry

Textiles.—The country's industry, although undergoing a number of vital changes, continued to be dominated by textiles throughout the period. During the 17th century the manufacture of the various finer types of woolen cloth, earlier introduced from the continent and known generally as the "new draperies," made great headway, supplementing and sometimes replacing the older types of broadcloth. The silk, hosiery and linen industry all developed in the course of this period; and toward the end of the 17th century the cotton-linen or fustian industry, using cotton yarn for wefts and linen for warps, made rapid strides in Lancashire and Cheshire. Taken as a whole, textiles—worsteds and woolens, dyed and undyed, cheap kerseys from the West Riding of Yorkshire, expensive coloured cloths from Gloucestershire, Colchester "bays" and Nottingham stockings—provided the largest single source of industrial employment in the country.

For most of the period the work was almost entirely carried on, as it had been earlier, within the framework of the "domestic system." The very spread of this system, however, created problems for the clothiers. The extension of the area over which workers were scattered produced increasing delays and militated against effective control of production; the low wages which the domestic worker earned encouraged theft of raw materials. This dependence on labour encouraged efforts to develop machinery which would replace the ancient hand processes, but it was not until 1770, when James Hargreaves of Blackburn patented the spinning jenny, that an effective substitute for the old spinning wheel appeared. This relatively simple machine did not, however, necessitate any change from the domestic system. By contrast, the invention of the water frame, patented by Richard Arkwright (*q.v.*) in 1769, was revolutionary for it was designed to be driven by water power and thus took the time-honoured domestic process of spinning out of the home and put it into a power-driven mill. In 1771 Arkwright set up a large water-powered mill at Cromford on the Derwent. This was the first of the big cotton-spinning mills, a herald of the industrialization of the future.

For various reasons, partly technical, partly social and economic, mechanization made its mark on the new cotton industry far more rapidly than on the older woolen industry. The former spread from Derbyshire into Cheshire and Lancashire; and more mechanization came with the invention by Samuel Crompton between 1774 and 1779 of the spinning "mule" which spun a finer thread than the water frame and for the first time allowed machine-spun yarn to be woven into all-cotton fabrics that could compete successfully with the handmade imported Indian cottons. Other processes of textile manufacture were meanwhile being similarly improved. Weaving of both woolens and cottons had been speeded up, though in no way revolutionized, by the introduction of John Kay's "flying shuttle," patented in 1733; in 1784 Edmund Cartwright patented a power loom but practical success in power weaving had to await later developments; in the silk industry the introduction from the continent of new methods of silk throwing had resulted in the erection by Thomas Lombe of a large water-powered factory, also on the Derwent, as early as 1717.

Mining and Metals.—The first steps in modern industrialization were not confined to textiles. The English mining and metal industries—iron, coal, tin, copper, lead—had continued to expand during the 17th century though at varying and fluctuating rates; coal output especially rose, reaching more than 2,500,000 tons annually by 1700. But two problems became increasingly urgent: that of drainage in mining; and that of finding a substitute for charcoal in smelting. The solutions that were found to these problems—the development of steam power for pumping and the successful use of coal for fuel and for smelting—went far to prepare the ground for the English Industrial Revolution.

As mines became deeper, alike in the tin mines of Cornwall and the coal workings of Northumberland and Durham, labour was needed to keep them clear of water. Some progress was made by Thomas Savery's invention in 1698 of a pumping engine which used the twin forces of atmospheric and steam pressure; and by Thomas Newcomen's much more successful atmospheric engine in 1708, many of which were set up during the 18th century at mines and collieries. In the 1760s James Watt (*q.v.*), a mathematical instrument maker of Glasgow, began to experiment with steam pressure and to study the defects of the Newcomen engine. During the following decades, and with the financial support of his partner after 1774, the Birmingham businessman Matthew Boulton, he successfully developed the true steam engine and made it capable of imparting rotary motion instead of the to-and-fro motion of the Newcomen pumping engine.

The scarcity and rising price of charcoal retarded the expansion of the metal industry. In the smelting of some metals, such as copper, coal had been successfully used in special furnaces. But in the vital iron industry no such success was achieved until 1709 when a Shropshire ironmaster, Abraham Darby, managed to smelt iron with coked coal. Even then the new methods were not widely adopted and much of the iron used in England in the 18th century was imported from Sweden and Russia. In the second half of the century, however, more and bigger blast furnaces in south Wales, south Yorkshire and the Midlands began to produce coke-smelted pig iron. The turning of cast pig iron with wrought iron had still to be done by an old process which used great quantities of charcoal. The search for another replacement of this expensive fuel by the relatively cheap coal was rewarded in 1783–84 with the patenting by Henry Cort of the "puddling" process. With steam power and cheap iron the stage was set for dramatic industrial advance. (*See* also Coal and Coal Mining; Iron and Steel.)

The interactions between these developments and the textile factories were manifold. Steam engines, set up to replace water power, drew the mills from the dales to the towns; the iron industry was encouraged by the demand for iron used in making steam engines and steam engines were used to drive rolling mills and hammers in the ironworks; all stimulated the output of the coal mines. But few of the effects were of great magnitude within this period. The same is true of the changes in industrial organization. Some big factories, spinning mills or coal mines, were to be found: in 1785 the Darby's Coalbrookdale works was said to be worth more than £100,000, some of the big new spinning mills employed from 150 to 500 persons and represented substantial concentrations of capital, just as did a few of the big collieries, breweries or paper mills. But all these were exceptional in 1789. The average enterprise was still small and ancient methods were far from wholly displaced by the new techniques. Just as many thousands still wove in their homes, so were many others busy at their ill-paid labours in the "Black country" of the west Midlands, making nails, chains and other small metalwares at home or in a huddle of tiny workshops.

Though the pottery industry was just beginning to make great strides as a result of the enterprise of Josiah Wedgwood whose Etruria works near Burslem were opened in 1769, in many industries crucial change lay in the future. Steel, for instance, remained expensive and scarce despite the invention of the important "crucible" process by Benjamin Huntsman in the 1740s.

Commerce

Communications.—England's industrial growth depended much upon its trade, internal no less than external. A growing volume of coasting trade testified both to the country's vigorous eco-

nomic life and to the difficulties of overland travel. Water transport, by sea or river, came into its own in the carriage of bulky goods: coal arrived in a fleet of colliers from Newcastle, to give London its smoky pall against which John Evelyn wrote (in vain) as early as 1661; grain came in hundreds of little ships from East Anglia or Kent to help feed the sprawling city. For much of this period the roads grew worse rather than better. Efforts to improve them mainly rested in the hands of the turnpike trusts—private enterprises designed to maintain roads from tolls authorized by statute. In the course of the 18th century the number of trusts rose but their effectiveness remained variable despite the remarkable work of John Metcalf, a blind Yorkshire road builder under whose direction several good turnpike roads were constructed in the latter years of the century. The problems of land travel led logically first to improving the navigability of rivers and then to building canals. Many stretches of English rivers, remote and weed choked today, were opened up by locks, and "cuts," and in the mid-18th century were busy with local traffic. In canal construction England lagged behind continental Europe until an era of rapid building began after 1750. The biggest and most famous, though not the first, of the new canals was the Worsley canal (opened 1761), financed by the duke of Bridgwater and built by James Brindley. Designed to link the duke's collieries with the Manchester market about ten miles away, it was a great success, lowering coal prices and earning a profitable revenue. Those that immediately followed were mainly in the north and in the west Midlands, a link to London being completed in 1790; this, however, was only the beginning of the most rapid and vigorous phase of canal building which was to come with the speeding up of the Industrial Revolution.

Overseas Trade.—In the export trade, as in industry, textiles retained their dominance, though there were changes in the composition and direction of the trade. The old concentration on woolen cloth exports to northern and central Europe was split up as increased amounts of woolens, worsteds and, later, fustians and cottons found their way farther afield, to the Mediterranean as well as to Amsterdam and Hamburg, to Africa as well as to North America. Although the bulk of the trade continued to be with Europe, some striking changes appeared as the fruit of overseas commercial expansion and colonial settlement: from the growth of trade with India and the east, from the conquests and settlements during the first half of the 17th century in the West Indies and North America, from England's successful entry into the slave trade through the establishment of forts and trading stations or factories on the west African coast. Imports and re-exports of colonial produce grew rapidly in the second half of the 17th century. Sugar from Barbados and other West Indian islands, tobacco and some cotton from Virginia, Maryland and other American colonies, spices and Indian textiles from the east—all came to London thence to be re-exported to European countries. Conversely, the colonies took increasing quantities of English wares—pots and pans, cloth and nails, paper and guns—as well as foreign goods shipped by English merchants in English ships; and the sugar and cotton plantations depended on English merchants alike for capital, for slaves and for sales.

This outward expansion of commerce meant longer voyages and larger ships; it encouraged investment in trade and shipping and shipbuilding. Much of it was still concentrated in London, though with the growth of Atlantic trade, Bristol and Liverpool grew in size and importance. As industry and population expanded at home so did imports of raw materials and, after about 1770, of food: timber and iron from the Baltic, cotton from America, cereals from continental Europe, fish from Newfoundland. England's role as an exporter of manufactured wares, an importer of food and raw materials and a great trading and shipping centre was apparent well before the Industrial Revolution; it was to make the country, in the 19th century, the workshop and entrepôt of the world.

Expansion was not achieved without conflict. The Anglo-Spanish clash of the 16th century was succeeded by the Anglo-Dutch rivalry of the 17th, when the Dutch Wars resulted in part from economic jealousies. The commercial and financial dominance of the Netherlands in 17th-century Europe was resented by French and English alike; and the latter were further incensed by the Dutch mastery of the North sea herring fisheries. Rivalry also brought conflict farther afield: in the east between the English and Dutch East India companies; in Africa over the slave trade; across the Atlantic from the Caribbean to the promising town of New Amsterdam which passed into English hands and became New York in 1664. Later, England's main economic rival was France, probably the richest country and greatest producer of manufactured goods in 18th-century Europe. Their home markets shielded from each other by stiff tariff walls, English and French merchants clashed elsewhere, as in the Mediterranean; farther afield, rivalries flared into wars which, in India and North America, left the English in position of striking economic superiority.

Commercial Policy and Finance.—International economic rivalry was fomented by government action, in England as elsewhere. In a world in which there seemed little scope for peaceful economic expansion, it was not perhaps surprising that a policy of economic aggression should be followed. The state had also to grapple with the problem of financing its wars and to contend with the social and economic problems of a society in which a quarter or more of the population lived below the bare margin of subsistence, in which the spectre of starvation lurked around many a corner, and in which social and economic discontent caused by heavy taxes and empty bellies readily found expression in riots or rebellion.

Typical of the commercial policy of the time were the Navigation acts of 1651 and 1661. The main aims of these acts were to restrict English trade with the colonies to English or colonial ships and merchants and to require most of the trade to pass through England. At one and the same time the acts struck at the Dutch, sought to stimulate the English merchant marine and tried to bind the colonies to the mother country. Efforts to make the colonies into markets for English manufactures and suppliers of its raw materials were reinforced by later enactments, limiting the growth of colonial manufactures. Such legislation contributed to the discontent of the American colonists. Attempts to push the export of manufactures were not confined to colonial trade. They were a commonplace of the time; it was hoped that thereby employment at home would be increased and a favourable balance of trade more readily obtained. So customs legislation tended to favour the export and tax the import of manufactured wares. This protective tendency, encouraged also by the need for revenue, became especially apparent from the later 17th century onward. Consequently a substantial tariff wall was erected around much of English industry. The very prolixity of this legislation, together with the vision of an ever-widening commerce, encouraged men to think of a freer system; the first systematic advocacy of a freely working economy was made by Adam Smith (*q.v.*) in his *Wealth of Nations* (1776), and its influence was soon evident, for example, in William Pitt's efforts to simplify customs duties in the 1780s.

During the 17th century the excise, the unpopular hearth tax (briefly) and the land tax were all introduced, but government borrowing remained haphazard and inefficient. Charles I's attempts to raise extraparliamentary revenue helped to precipitate the Civil War. Charles II depended heavily on London bankers and financiers, just as Elizabeth I had on their Antwerp counterparts. It was in the usual circumstances of financial stringency brought by war that a private scheme of a group of financiers led by William Paterson was accepted by the government in 1694 and the Bank of England (*q.v.*) was founded. The bank took over the management of the national debt—which was put on a funded basis during the early years of the 18th century—and undoubtedly brought a new element of stability to English government finance. But the speculative boom of the South Sea Bubble in 1720 gave a shock to credit and the resulting Bubble act restricted the promotion of joint-stock companies. The Bank of England had a virtual monopoly of joint-stock banking; in the provinces banks were increasing in number though they tended to develop with inadequate resources and to suffer accordingly.

Society and Social Policy

The Social Structure.—In Jacobean England contemporaries

assessed a man's place in society largely in terms of hereditary rank and tried as far as possible to fit occupations into a structure of which the main members were nobility, landed gentry, yeomanry, and that vast mass which counted for nothing, the poor. In fact, of course, English social structure was already far more complex than this. The *bourgeoisie* of market towns and ports, the prosperous rural clothiers, the craftsmen and industrial wageworkers, the rich merchants and flourishing lawyers of London—all combine to make nonsense of too simple a picture of early 17th-century England. To a remarkable extent the outlines of the structure remained the same in 1789. The great peers and the landed gentry were still the rulers of the land, their social sway unchallenged; and the continued recruitment into the ranks of the aristocracy and upper gentry of successful lawyers and merchants helped to preserve the structure. But the forces of change were growing stronger. The merchants and financiers of London were being joined by the big businessmen of Liverpool and Manchester; the middle classes, varying in composition from professional men to lesser shopkeepers, were growing in size and diversity, in outlook and ambitions. Above all, the expansion of industry was making its mark as ironmasters, cotton-mill owners and factory workers were beginning, but only beginning, to form part of the social structure.

Attitudes to economic endeavour and to the social order similarly changed as a more rational ethos slowly spread, helping to destroy the ready acceptance of traditional values. Some writers have seen in Puritan, Nonconformist Christianity, with its stress upon work and frugality and its conception of a "calling," a spiritual driving force urging men toward worldly success. The links of causation between religious ideas and business practice are not, however, easy to isolate or to assess.

Despite the presence of many Quakers among, for instance, the Midland ironmasters and the undoubted importance of the inquiring, rationalistic education provided by the Dissenting academies (as opposed to the dormant Anglican conservatism of much of 18th-century Oxford and Cambridge), it should not be too readily assumed that capitalism and Puritanism have some special causal relationship.

Social and Economic Policy.—To some extent these changing attitudes affected state policy. In the field of industrial regulation and social policy, the provisions of the Statute of Artificers and the Elizabethan Poor law remained on the statute book. After the Civil War, however, and particularly after the revolution of 1688, industry tended in practice to shake off the regulations of Tudor days; and neither state policy nor the law did much to stop it. Guilds continued to decay, apprenticeship tended to decline and wage assessment to lose its force. In business organization a more individualistic note was heard. Many of the big companies were either swept away in 1688 or died of inanition. Some remained, notably the East India company. But the trend was evident; the private firm, one-man enterprise or partnership, be it trading venture, banking house or ironworks, operating in an economy in which the government usually sought to control not the processes of production, as it often did in France, but the structure of overseas trade. In contrast to this trend toward less stringent industrial regulation, the policy toward the poor retained its old ingredients and enjoyed little success. The Settlement act of 1662 empowered justices of the peace to send paupers back to the village whence they came. The burden of poor relief mounted with more parish relief, charitable bequests and workhouses founded from about the 1690s onward in London, Bristol and provincial cities. Despite the expansion of trade and the dawning Industrial Revolution, despite the first signs of the application of the new science to industry, for all the great country houses of the Augustan age, the "improving" landlords and the freedom which Voltaire praised and which set England apart from much of continental Europe, there remained at the base of society the great mass of the poor. And the appropriate contemporary attitude to them was expressed in the classic statement of Arthur Young: "Everyone but an idiot knows that the lower classes must be kept poor or they will never be industrious."

See also Social Welfare; Poor Law. (D. C. Cn.)

XV. REVOLUTION AND REFORM, 1789–1837

Reaction to the French Revolution

The news of the French Revolution was received with enthusiasm by the vocal minority of radicals and reformers. Prominent among them were many of the leaders of Nonconformity. The London Revolution society which had just celebrated the centenary of the revolution of 1688 as a festival of political liberty, congratulated the French National Assembly in Nov. 1789 on the victory of liberty and justice over arbitrary power. This enthusiasm was shared by poets and writers. Wordsworth, representing the new generation, saw the Revolution as the triumph of truth and light: Cowper joined with Burns and Blake in singing the praises of this "wonderful period in the history of mankind." Charles James Fox, speaking for the Whigs, saluted the capture of the Bastille: "How much the greatest event it is that ever happened in the world, and how much the best."

Praise of the Revolution was not, however, unanimous even before the course of events in France had shocked many liberal-minded Englishmen. In particular Edmund Burke, Fox's friend, wrote to a French correspondent in Oct. 1789 that true liberty which rested on security of life, property and opinion, had not been guaranteed in France; and when he read the chief address delivered (by Richard Price) at the London Revolution society's meeting a month later Burke was impelled to write his powerful rejoinder, *Reflections on the Revolution in France*. The publication of this work in Nov. 1790 marked the first stage in an English counter-offensive against revolutionary ideas. At the time of its publication the attitude of William Pitt to the Revolution was unemotional and correct—as late as Jan. 1792 he talked of European tranquility and anticipated 15 years of peace—and the chief effect of Burke's eloquence was to provoke a breach within the ranks of the Whigs between himself and Fox. In 1790 they publicly asserted their differences; in May 1791, in a tense scene in the house of commons, they publicly announced the breaking-up of their friendship, which had been so rich and fertile before 1789. Burke went on to establish his reputation as the prophet of conservatism: Fox, in face of increasing unpopularity both in parliament and outside, remained faithful to the cause of reform.

The division in the Whig ranks was followed by the strengthening of Pitt's forces. As alarming news arrived from Paris of the "progress" of the Revolution there was strong pressure to unite Englishmen "behind the constitution." In Dec. 1792 the duke of Portland, the official leader of the Whigs, reluctantly allowed his friends to support Pitt's ministry to "maintain the constitution and save the country." The earl of Loughborough (afterward earl of Rosslyn) became lord chancellor in Jan. 1793. This was the first of a number of ministerial changes which culminated in July 1794 with the formal entry of the duke of Portland and most of the Whigs into a coalition government with Pitt as leader. In the meantime France had declared war on Great Britain on Feb. 1, 1793. The news of the execution of Louis XVI followed by the declaration of war led Portland to remark that "we were embarked upon a war on the issue of which the very existence of our government depended." In both the house of commons and the house of lords those Whigs who continued to side with Fox were in a tiny minority, and it became increasingly difficult for them to mobilize support for reforms which had been well supported before 1789. In 1789 itself a bill to repeal the Test and Corporation acts had mustered 102 votes in the commons; when in May 1793 Fox proposed a less important concession to the Nonconformists, leaving the Test and Corporation acts unrepealed, only 13 members of his party would support him. Even the cause of emancipation of the slaves, led as it was by such friends of Pitt as William Wilberforce, was treated with suspicion at this time as a revolutionary cause. "You will see Clarkson," Wilberforce wrote to a supporter in 1791, "caution him against talking of the French Revolution; it will be ruin to our cause." When Wilberforce was elected a French citizen in Sept. 1792 along with Thomas Clarkson, Joseph Priestley and Tom Paine, he hastily joined the committee for the relief of the French emigrant clergy to show unmistakably where his sympathies lay.

Radicalism.—In the country the views of Pitt and Wilber-

force commanded greater support than those of Fox or Paine. There was a vigorous radical movement in the early 1790s, nourished by Paine and other writers, but it generated hostility not only among men of property but in the growing cities, where the cry of "church and king" rallied widespread support. Paine, in England and France from 1787 to 1802, was the most able of the 38 authors who rushed to criticize Burke's *Reflections*. His brilliant statement of radical principles, *The Rights of Man* (1791–92), was so widely read by working men in all parts of the country that the government had to take steps to support counter-propaganda. But before Paine's book appeared a Birmingham mob had destroyed Priestley's manuscripts, house and laboratory, and a year later there was a similar but smaller disturbance in Manchester. An attempt to disseminate *The Rights of Man* in Leeds led to Paine's effigy being paraded and burned.

Both the energy and the courage of the radical reformers were well displayed in this period of turbulence and excitement. Old radical societies, like the Society for Constitutional Information, sprang into renewed life in 1791–92; new societies appeared both in London and in the provinces. The advanced Whigs formed the Society for the Friends of the People in April 1792; one of its leaders was Charles Grey, the great Whig aristocrat who later carried the Reform bill of 1832. The subscription to the society was one guinea and a half. At the other end of the social scale the London Corresponding society was founded in Jan. 1792 by a number of journeymen and artisans, the élite of the common people. Its members paid a subscription of one penny a week, and its program included social as well as political reform. The title of this organization, which in many ways heralded 19th-century developments in working-class political association, reflects the concern for correspondence with radical groups in other places. Its founder Thomas Hardy was on close terms not only with experienced radicals like Horne Tooke but with new provincial radical leaders in Sheffield, Manchester, Warwick, Stockport and in Scotland.

Repression.—Relations between the different radical organizations were not always cordial, but they were all subject to official suspicion and popular attack. In Nov. 1792 John Reeves, a lawyer just back from Canada, formed the Association for Protecting Liberty and Property Against Republicans and Levellers. The avowed purpose of this association was to suppress seditious publications and to promote the cause of "church and king." A few months earlier the government had issued a proclamation warning the king's loyal subjects against seditious meetings and publications, and Paine was summoned to appear before the court of king's bench for a seditious libel. More prosecutions followed in 1793 and 1794, and there was a flood of loyal addresses from the provinces. News from France was just as important in stirring patriotic feelings as it was in shifting the balance of forces in parliament, and after war had been declared it was easy to identify reform and treason. The most sensational trial in 1793 was in Scotland where Thomas Muir, a young liberal-minded lawyer, was sentenced to transportation for 14 years. Lord Braxfield, the lord justice clerk, who presided, took advantage of the occasion not only to attack the French as "monsters of human nature" but to defend privilege as the basis of British government. "A government in every country should be just like a corporation; and in this country, it is made up of the landed interest, which alone has a right to be represented; as for the rabble, who have nothing but personal property, what hold has the nation of them?"

This extreme statement concerning the privileged nature of political authority was the dominating orthodoxy of the long period of war and disturbance which lasted from 1793 to the 1820s. The orthodoxy was buttressed by patriotism and when need arose supported by legislation. In 1794 there was a wave of state trials in London—of men like Hardy, Horne Tooke and John Thelwall, the most effective of the radical political lecturers. All were acquitted, but the cause of parliamentary reform, which had once attracted Pitt, was broken. The Constitutional society ceased to exist, the Society of the Friends of the People languished, and although the Corresponding society held together for some time, many of the provincial bodies disappeared. During the winter of 1795–96, when harvest failure created a favourable situation for radicals to exploit, the government passed two acts, the first extending the definition of treason to include a number of political offenses (the Treasonable Practices act) and the second banning large public meetings (and political discussions) unless held by special authority (the Seditious Meetings act). A year later the government raised stamp duties on newspapers and introduced a system of registration of printing presses. By a series of further acts in 1799 all secret associations, federations and trade unions were placed under government ban (the Combination acts). The coercive powers of the government were complete. Radicals were conscious of their inability either to act themselves or to influence other people. As early as 1793 an English Jacobin had declared that he found it prudent to say as little as possible upon political subjects in order to keep himself out of Newgate: in 1797 Horne Tooke wrote to one of the most consistent of the reformers, Major John Cartwright, who like Tooke spanned the 18th and 19th centuries, that the cause of reform was "dead and buried." Four years later in 1801 Fox told Grey that until the public showed some dislike ("indignation I do not hope for") for absolute power, "I see no use in stating in the house of commons the principles of liberty and justice."

It was not until 1806 and 1807 that radicalism revived with Francis Burdett and William Cobbett (*qq.v.*) in the forefront.

War With France

Pitt and the War.—With little continuous radical opposition to hamper him and with Fox, his great political rival, reduced to parliamentary impotence, Pitt was in a strong position to lead the nation in war. Certainly he established himself in war and peace as the "man of and for the time"; yet he had neither the experience nor the talents to run the protracted war with great success. Moreover after the rise to power of Napoleon he was confronting an adversary of genius. For the military aspects of the wars *see* French Revolutionary Wars; Napoleonic Wars; Peninsular War; Waterloo Campaign.

Pitt had no "grand strategy" to defeat the French. At first he relied on scattered British actions outside Europe coupled with guaranteed loans and direct gifts to allies on the continent. Some of the former actions were successful, but they contributed little directly to the winning of the war in Europe. The latter policy did not guarantee the success of the four coalitions which were formed against France but only held together precariously and temporarily.

Fortunately for Pitt the country's economic power, greatly increased as a result of the industrial advances of the 1770s and 1780s, gave Great Britain a long-term advantage over the French, while British naval power and leadership kept the ocean lanes open and prevented the French from striking directly at England. Adm. Lord Howe, a hero of the Seven Years' War, commanding the channel fleet won the "glorious first of June" in 1794 and the Spanish fleet was beaten off Cape St. Vincent by Sir John Jervis (afterward Lord St. Vincent) and Horatio Nelson in 1797. After Napoleon had gained control of Egypt in 1798 Nelson blockaded and annihilated the French fleet at the battle of the Nile. These naval victories improved Britain's position at the end of the first period of the war, and Pitt also strengthened the national economy by a series of measures designed to combat financial difficulties, which reached a climax in 1797. Cash payments by the Bank of England were suspended in 1797, income tax was introduced in 1799, and indirect taxes were increased. A continued rise in national income was accompanied, however, by a great increase in the national debt which over the whole period of the wars against France rose from £228,000,000 to £876,000,000.

Pitt was prime minister continuously from 1783 to Feb. 1801 and from April 1804 until his death in Jan. 1806. The break in his office had nothing to do with his conduct of the war; it was brought about by a disagreement with the king on what was and remained the vexed question of Roman Catholic emancipation. In 1800 Pitt carried through an act of legislative union with Ireland, and tried to follow it with a measure of Catholic emancipation. Roman Catholics—and Protestant Dissenters—would be admitted both to parliament and to office on taking an amended

oath of allegiance and Catholic priests in Ireland would receive state assistance. George III, who was on the verge of another fit of insanity, resisted these proposals, and Pitt, who showed no desire to convert his sovereign, resigned from office. He helped his successor Henry Addington (later Viscount Sidmouth) to form and win support for a new ministry, and in addition promised George III not to revive the question of Catholic emancipation during the king's lifetime.

Addington's "Protestant" cabinet was close to the king and not unpopular in the country, although Addington himself, who had been speaker of the house of commons for 11 years, was hardly an outstanding personality (*see* SIDMOUTH, HENRY ADDINGTON, 1st Viscount). His colleagues included two future prime ministers—Robert Jenkinson (later Lord Hawkesbury and earl of Liverpool) as foreign secretary and Spencer Perceval as leader of the house of commons. The ministry was ushered in with outstanding British victories. Nelson won a great victory over the Danes at Copenhagen in April 1801 and far away in Egypt British forces captured Cairo and Alexandria. At the same time the new ministers were anxious, if possible, to bring the war against Napoleon to a close. The loss of allies, the intensity of French preparations to invade England, the existence of widespread discontent, particularly with the high price of bread, and relevant grievances centring on the dislocation of trade, predisposed Addington and his colleagues to enter into negotiations for a peace. Pitt and his foreign secretary Lord Grenville had resisted Napoleon's peace overtures in Dec. 1799, largely on the grounds that Napoleon's political power would not last. By 1801 it was clear that Napoleon's power, far from waning, was being consolidated each week. Hawkesbury made the first new peace overtures in March 1801 and stated clearly Great Britain's demands in April. A preliminary treaty was signed on Oct. 1, the day before the news of the French capitulation in Egypt reached Britain. It was followed, after the British had made several concessions in an atmosphere of mutual distrust, by the signing of the peace of Amiens in March 1802. It was a peace that did little to afford confidence: as Sheridan put it, "this is a peace which all men are glad of, but no man can be proud of." Within 14 months Great Britain and France were once more at war. There was to be no further break in hostilities until the defeat of Napoleon.

Pitt's Last Ministry.—When it became clear that the peace of Amiens could not last Pitt declared his willingness to resume office, making it clear that he would wish to have associated with him Lord Grenville and William Windham, two of his former ministers who had strongly attacked Addington's conduct of affairs. After months of wrangling Addington at last resigned in April 1804. Pitt wished to include Fox in his ministry—they had recently joined together in supporting motions against Addington's government—but the king, who insisted on a renewed pledge against Roman Catholic emancipation, refused to allow him a place in the cabinet. Angered by the veto on Fox, Grenville also refused to join the new ministry, and Pitt had to depend on his loyal supporters along with most of Addington's old colleagues. The discussions showed the complexity of group politics at this date. They also showed the degree of Pitt's dependence on the house of lords. Of Pitt's 11 colleagues in the cabinet, Lord Castlereagh alone (as president of the board of control) had a seat in the house of commons, and he was an Irish peer. By a coincidence Pitt took his seat in parliament again (after accepting office) on May 18, 1804, the same day Napoleon was declared emperor by the French senate.

After the resumption of hostilities Napoleon concentrated most of his efforts on the invasion of Great Britain. Pitt retaliated on traditional lines not only by taking defensive measures in the south of England but by seeking to build up a third coalition in Europe. The fate of the coalition was sealed by Napoleon's victory over the Austrians at Ulm (Oct. 1805) and by his subsequent entry into Vienna, but in the meantime Nelson had relieved the pressure on Britain itself by his great victory at Trafalgar on Oct. 21. Cheered as he was by the battle of Trafalgar (*q.v.*), Pitt was disheartened at the failure of the coalition. His last ministry had not been a happy one. One of his closest friends, Viscount Melville (formerly Henry Dundas), was charged with having misapplied public money as treasurer of the navy in Pitt's former ministry. He was impeached before the house of lords and, although acquitted, could no longer hold high office. Pitt took the blow hard. He died on Jan. 23, 1806, tired and exhausted with the end of the war not yet in sight. His memory continued to inspire his followers, but he had been a victim of the circumstances of his time, called upon to serve as the "Atlas of our reeling globe." His early policies had crumbled during the course of the struggle against France. Yet, if he did not win the war, he had established a permanent reputation for disinterestedness and patriotism and had succeeded in drawing deep on the national reserves of strength.

The Way to Victory.—Pitt's death was followed by political upheaval. Grenville formed a new ministry, the so-called "ministry of all the talents," which included Fox as foreign secretary. The Whigs were restored to office after years in the wilderness in a coalition which also included Sidmouth. The ministry proved as incapable of waging war successfully as Pitt, and it suffered a heavy blow when Fox followed his old rival to the grave on Sept. 13, 1806. The main achievement of this period was the abolition of the English slave trade which was carried on a motion not of the government but of Wilberforce (*see* SLAVERY). The bills for the abolition of the trade had been carried in the house of commons before the close of the century but defeated in the house of lords. In 1807 the determined leadership of Grenville in the upper house made possible the achievement of this long-contested reform. Grenville proposed to follow this reform with a measure of relief to Roman Catholics. The Catholic question split the cabinet, however, and brought about its downfall in March 1807.

The two succeeding ministries—the first lasting until 1809 and led (nominally) by the duke of Portland and the second led by Spencer Perceval from 1809 until May 1812—saw great changes in the balance of international power which were to influence the subsequent fate of Napoleon. Napoleon and the emperor Alexander of Russia reached agreement at Tilsit in July 1807; from this date onward Napoleon sought to seal off Great Britain from the continent in order to break Britain's economic power by enforcing the continental system (*q.v.*), the origins of which were to be found in the Berlin decrees of Nov. 1806 declaring Great Britain to be in a state of blockade and forbidding neutrals as well as allies to trade with that country or its colonies. It was not so much the policy of counter-blockade of the British government or the vigour of George Canning, who had taken over the foreign office, which saved Britain as the superior strength of the British economy. Moreover the attempt to enforce the continental system against Britain led Napoleon into serious difficulties. The invasion of Portugal in 1807 and of Spain in 1808 did not produce the results he anticipated, and once the situation there had been clarified, the Iberian peninsula provided an effective bridgehead for the troops of Wellington's army. Yet it was not until the beginning of the second decade of the century that the full measure of Napoleon's dangerous commitment in Spain was revealed: in 1809 there was a repetition of the familiar sequence of military disasters. The Austrians were defeated at Wagram in July, a British expedition to the island of Walcheren proved a fiasco and in December British troops had to be withdrawn. Furthermore there were fierce internal discussions in the Portland ministry. Lord Castlereagh and George Canning resigned from the ministry and fought a duel in Sept. 1809, and after Portland's resignation and death in Oct. 1809 Spencer Perceval, Portland's chancellor of the exchequer and the ablest commoner in what was left of the cabinet, became prime minister. His ministry, however, was a weak one. Canning wanted a strong government and a strong prime minister who would bring method to the fighting of the war, and stood out on one side. So too did Castlereagh and the two main sections of the Whigs, one led by Grenville, the other by Lord Grey. Hawkesbury, who soon afterward succeeded his father as earl of Liverpool, was given charge of the war department.

The Regency.—It was Liverpool who succeeded Perceval as prime minister after the latter had been assassinated by a madman in May 1812. He was to remain prime minister until forced to resign through ill-health in 1827, seeing the country through to victory and from victory through the extremely difficult period of

domestic disturbance that followed the end of the war. This long period of rule could not have been predicted in May 1812. Indeed, Liverpool would not have come into power at all had the prince of Wales satisfied the expectations of the Whigs when he took over full power as prince regent in Feb. 1811. His decisions to retain Perceval as prime minister and in 1812 to choose Liverpool were important retrospectively not only in the history of the monarchy but of 19th-century British politics.

George III had celebrated his jubilee in Oct. 1809, when his subjects provided a striking demonstration of his genuine popularity. In Nov. 1810, however, he finally lapsed into permanent mental derangement. His sight had been failing for more than six years, but he had suffered no return of insanity since 1804. When it became clear in 1810 that he was incapable of exercising his powers the precedent of 1788 was followed (*see* GEORGE III) and by the Regency bill of Feb. 1811 the prince of Wales was made regent subject to certain temporary restrictions. The prince regent stated that he still hoped for his father's recovery, flinched from making dismissals and retained Perceval in office. He strongly disliked Grey and was beginning openly to take the same line as his father on Catholic emancipation. When he acquired full powers in Feb. 1812 he did not disturb existing political arrangements. When, after the assassination of Perceval, attempts to form a new government under the leadership of the Marquess Wellesley or of the Whigs had failed, Liverpool was confirmed in office. His position was further strengthened after the general election of the same year. The failure of the Whigs in 1811 and 1812 kept most of their leaders out of power until 1830. They had proved themselves not only divided but bankrupt in ideas, and Liverpool, despite his initial weakness, was able to continue to attract support.

The End of the War.—Domestic problems were quickly overshadowed by events abroad. Strained relations between Napoleon and the Russian emperor pointed to a new conflict, and in June 1812 Napoleon's great army crossed the Niemen en route for Moscow. In September the remnants of his cold and hungry army arrived in Moscow only to find the city in flames. In October Napoleon evacuated the city and turned west again through the icy Russian winter. Had Russia been able to carry the war into French territory the war might have ended. As it was the British armies at the other end of Europe in the Iberian peninsula were passing from the defensive to the offensive. Victory at Salamanca in July 1812 and at Vitoria in June 1813 forced the French to evacuate all Spain south of the Ebro river. Soon they were pushed back to the Pyrenees, and the duke of Wellington entered Toulouse in March 1814.

While Wellington was winning these victories a new coalition was forming against Napoleon. Castlereagh (*see* LONDONDERRY, ROBERT STEWART, 2nd Marquess of), who was far more effective as foreign secretary than he had been as minister of war, played a leading part in the forging of the last of the great coalitions which by June 1813 included Great Britain, Russia, Prussia and Austria. In a three-day battle at Leipzig in Oct. 1813—"the battle of the nations"—Napoleon was defeated. Germany was abandoned by the French and the vast Napoleonic empire began to disintegrate. On March 1, 1814, the allies signed the treaty of Chaumont which bound them to make no separate peace and to continue the war, if necessary, for another 20 years. Less than a month after Chaumont the allies, led by the emperor Alexander, entered Paris. The Bourbons, in the person of Louis XVIII, returned to the throne and Napoleon was exiled to Elba. He escaped in Feb. 1815 but when Wellington won the decisive victory against him at Waterloo on June 18, Britain had at last emerged successful after the long struggle, and Napoleon was exiled to the remote island of St. Helena.

Great Britain secured its main aims by the first peace of Paris, signed on May 30, 1814, before Napoleon's escape. French boundaries were re-drawn to coincide generally with those that had existed in 1792. At the same time barriers were erected in Europe in an attempt to check any further encroachments by France on its neighbours; in particular Belgium was joined to the Netherlands. Outside Europe Great Britain's command of the world ocean routes was strengthened. The Cape of Good Hope, Heligoland, Malta, Ceylon, Mauritius and islands in the West Indies were secured in British hands. Britain also obtained from the other great powers agreements for opening certain rivers to navigation and (less effectively) for abolishing the slave trade. These territorial and "moral" gains were not the only gains of the war. The differential advantage which the British economy enjoyed over the other economies of Europe had been widened. In industrial development and shipping in particular Britain was well ahead of any potential competitors. At the congress of Vienna, Castlereagh took the initiative in securing agreement to the final peace, signed on Nov. 20, 1815.

COST OF VICTORY

Economic and Social Consequences of the War.—Nonetheless the war had imposed burdens on the British people, and although the economy had continued to expand, it had expanded less rapidly and more jerkily than otherwise would have been the case. The cotton industry, for example, had gone through years of great difficulty, when both supplies and markets were in jeopardy, while the iron industry, tied largely to war requirements, grew in fits and starts and was over-capitalized when the war ended. Businessmen had been very critical of high taxation and of the restrictions on trade with the continent which were set out in the series of orders in council, first introduced in 1807 as Great Britain's reply to the continental system. Many businessmen were strongly opposed to the war with the United States which began in June 1812 and did not end until Dec. 1814. It was largely as a result of business pressure that the orders in council were withdrawn in 1812.

The years 1811 and 1812 were troubled years in English social history, being marked not only by the increasing articulation of business grievances but by a resurgence of popular radicalism and an epidemic of machine-breaking and industrial disturbances in the textile districts. The Luddites (*q.v.*), so-called after a legendary leader, Ned Ludd, set about destroying machinery and intimidating employers. In Nottingham, the main centre of the agitation, it was claimed that local disturbances had "no parallel in history since the troubled days of Charles I." Social disturbances were aggravated with the end of the war when there was dislocation of war industry, demobilization and a considerable amount of unemployment. Indeed, the domestic conflict of the period from 1815 to 1820, reaching its peak in years when bad harvests and a high level of unemployment coincided, tested the Liverpool government more thoroughly than the struggle against Napoleon. Sidmouth, Liverpool's home secretary, acquired a reputation during this period for crude repression, and it is certainly true to say that the policy of the government was to stifle discontent rather than to introduce reforms. It was terrified of the militant radicalism which gained in momentum largely as the result of the growth of a popular radical press and the activities of William Cobbett (*q.v.*). There were radical riots and disturbances in 1816, 1817 and particularly in 1819. The government continued to follow a policy of repression culminating in the Peterloo massacre at Manchester (*see* PETERLOO) and the six acts of 1819 which limited freedom of speech and of the press. Fear and suspicion threatened the stability of the whole social order.

During the war agriculture had been in a thriving condition and landlords in particular had benefited from high rents and food prices. Strongly represented in parliament, the landed interest attempted to safeguard its wartime position after the end of the struggle. In 1815 it secured a corn law which was designed to keep up rent and grain prices (*see* CORN LAWS). Yet despite the passing of the law the agriculturalists suffered after the war, particularly after 1819, from a serious depression in agricultural prices. Debts contracted during the war rose in real value as prices fell. There were many complaints of agricultural distress to add to the chaos of woes in the early 1820s.

Postwar Economy.—All these signs of social strain were aggravated by fiscal policy in the postwar period. Of the annual public revenue more than half had to be used to pay the interest on the national debt. The abolition of the income tax in 1816 meant

that much of the taxation necessary to pay the yearly interest charges to a small class of bondholders was levied on consumers and on industrialists. The archaic and regressive nature of the central taxation system and the local burden of increasing Poor-law rates provoked widespread anxiety and criticism. In the agricultural districts and the industrial midlands there were bitter complaints about the government's decision to return to payments in gold in 1819. Yet this decision was implemented in 1821.

Age of Reform

Beginnings of Change.—When George IV came to the throne on the death of his father (Jan. 29, 1820) political and social strains were at their most intense. George himself contributed to them or at least gave them a convenient outlet. He was anxious to divorce his wife, Queen Caroline, who immediately became a heroine of the radical crowds. George's personal unpopularity added to the unpopularity of his ministers. The decision of the cabinet that it could not meet the new king's wishes and carry a divorce bill was greeted as a popular victory. George IV, who ostentatiously displayed his deficiencies of character in public, continued to remain unpopular for the rest of his reign, but there was a revival of national confidence and economic prosperity in the early 1820s, and significant changes took place in the outlook of Liverpool's government. The spirit of "improvement" was beginning to influence politics and legislation. Between 1821 and 1825 duties on raw materials were very much reduced; tariff schedules were simplified; and in 1828 the fixed corn law duty was replaced by a sliding scale which rose and fell as the price of corn varied. During this same period the repeal of the Combination laws of 1799–1800, which had prevented associations of workmen, allowed for the eventual development of collective bargaining between workers and employers within the new labour market. Although a boom in 1824 and 1825 was only temporary, and hard times prevailed during the later 1820s, no attempt was made to return to the negative policy of repression which had been followed by Lord Sidmouth at the home office between 1815 and 1822. Sidmouth's successor, Robert Peel, was prepared to make concessions, while other members of the cabinet, notably William Huskisson (*q.v.*), were of a liberal frame of mind.

Foreign Policy.—There was some shift in foreign policy, as in home affairs, during and after 1822. Castlereagh, who had been responsible for the direction of British foreign policy since 1812, committed suicide in 1822 and was replaced by George Canning (*q.v.*) who had been kept in the background of politics during the previous ten years because of personal unpopularity, first with his colleagues and then with the king. Already before 1822 there had been signs of an increasing cleavage between Britain and the other victorious allies. The world-wide interests of Great Britain and its dependence on sea power differentiated it from the land powers of Europe, which were controlled by conservative regimes anxious to control or repress the development of liberalism and nationalism. Castlereagh himself withdrew from active interference on the continent to a position of relative neutrality, and objected to basing British policy on "abstract and speculative principles of precaution." He refused to follow up the peace settlement by a continuous policy directed against the forces of constitutional change in Europe, or to allow the periodic congresses of the great powers (1818–22), provided for in the settlement of 1815, to be turned into instruments of interference in the internal affairs of states. Canning expressed the British viewpoint in more colourful language than his predecessor and emphasized differences between Britain and the other great powers rather than common interests. He protested strongly in 1823 against the French invasion of Spain and made it clear that if European countries attempted to send troops across the Atlantic they would find their way barred by the British fleet.

Canning recognized the independence of the Spanish American colonies in 1824, for commercial as well as diplomatic reasons, and declared that he had "called the New World into existence to redress the balance of the Old." He used British power again to defend constitutional government in Portugal in 1826, and he supported the cause of Greek independence. By working in co-operation with Russia and France he drove a wedge into the European coalition of monarchical powers which was defending the *status quo*. Two months before the English, Russian and French fleets destroyed the Turkish fleet at Navarino, Canning died. Although his policy was temporarily reversed by the duke of Wellington, a new Greek state came into existence and after 1830 the principles of Canning's policy were reasserted by Lord Palmerston.

The Movement for Constitutional Reform.—The personality of Canning exercised an important influence on English domestic policy as well as on foreign affairs. Many of the ruling groups looked upon him as a "charlatan parvenu" and were suspicious not only of his foreign policy but also of his interest in Catholic emancipation. When Lord Liverpool resigned and Canning became prime minister in March 1827, hidden conflicts came to the surface. The government bloc, which had first been brought together by Pitt, was finally shattered and constitutional reform re-emerged as a practical rather than as a theoretical question. Without the political confusion of 1827–30 there could have been no Whig Reform bill in 1832. Lord Liverpool had held together governments which included men of many conflicting opinions. Although he lacked the full resources of patronage which had been considered indispensable in the 18th century, and although he frequently faced the opposition of the king, he managed his cabinets and parliament with considerable success. Nonetheless, after 1822, he was poised between two hostile sections of his supporters, the one favouring a program of liberal reforms, the other anxious to stand still or to look back to the past. When Canning formed his ministry seven former ministers, including Wellington and Peel, refused to join him, and although Wellington himself became prime minister in 1828 existing differences were widened after the new government, in the face of Irish disorders (*see* O'Connell, Daniel), was compelled to pass Catholic emancipation, against the resistance of a large group of "ultra-Tories." The opposition of both Tories and Whigs to Wellington's administration placed him in a weak position in 1830.

The death of George IV on June 26, 1830, speeded up political events. After the accession of William IV and an inconclusive general election, Wellington was defeated in the house of commons over a vote on the civil list. He resigned and the king sent for the Whig leader Lord Grey (*see* Grey, Charles Grey, 2nd Earl). At last there was a break in the continuity of regime which had been inaugurated by Pitt. The new government was the parent of most of the Whig-Liberal administrations of the next 50 years.

The year 1830 was one of economic and social grievances, in the southern countryside as well as in the northern towns. Corn laws, Poor laws and Currency laws were all attacked and political reform movements in different localities converged. It was in such an atmosphere that the new Whig government prepared a Reform bill.

Whig interest in parliamentary reform went back to the 18th century and Grey himself provided a link between two periods of agitation. In the country as a whole there were two distinct approaches to the reform question. Middle-class reformers were anxious to secure representation for industrial and commercial communities and interests; "popular radicals" were more concerned with a protest against distress, a theory of rights and a dream of "a different system after the Reform bill is disposed of." Neither middle-class nor working-class reformers had much in common with the aristocratic temper of Grey. It was the agitation in the country which kept the reform question alive between 1830 and 1832 while Grey was facing constitutional difficulties with the king and with parliament.

The Reform Bill of 1832.—The first Reform bill was introduced in March 1831 and was carried by one vote on its second reading. When an opposition amendment was carried Grey induced the king to dissolve parliament and won a substantial majority at a new general election. A second Reform bill passed the commons with no difficulty but was defeated in October in the house of lords. Popular agitation in the country reached its height and Grey, afraid of civil disorder, first pursued paths of compromise and then, when the lords still proved obstreperous in May

1832, asked the king to create not less than 50 new peers. The king refused but after Wellington failed to form an alternative ministry William offered Grey a full pledge and the Whig cabinet returned to office. The lords passed the bill without the creation of new peers being necessary.

The Reform bill was not a democratic measure; it was concerned with giving the new middle classes a share in responsible government rather than with changing the whole basis of government. It included a redistribution of seats, which gave greater representation to the new industrial areas and disfranchised entirely 56 boroughs, and made a change in the conditions of the franchise. The distinction was retained between counties and boroughs: in the former, 40*s.* freeholders provided the base of the electorate; in the latter, £10 occupiers. The total electorate of England and Wales was increased by 217,000, but the artisans and some of the lower middle classes still remained outside what Gladstone called in his famous speech of 1864 "the pale of the constitution." The bill thus failed to meet radical demands. The manner of its passing had demonstrated the force of organized opinion in the country, but the new house of commons differed little from the old. In composition it continued to reflect property rather than population, and the landed interests remained by far the largest property interest represented there. (*See* PARLIAMENT: *The Parliament of the United Kingdom.*)

Although many of the Whigs thought of the bill as a final measure, they carried out additional reforms after their sweeping victory at the general election of 1832, the first under the new electoral system. There was a thoroughgoing reform of the admiralty in 1832; in 1834 the new Poor law was passed which grouped traditional parishes into new units called unions and put them under the control of boards of guardians with a national Poor law board in London. In 1835 the Municipal Corporations act was passed, which swept away old oligarchies in local life. Town government was committed to elected councils which were required to appoint town clerks and treasurers (*see* BOROUGH). Many of the new industrial towns, which lacked any effective local government, were incorporated during the following ten years.

The chief impetus to reform came from middle-class pressure and the intellectual and administrative drive of Benthamite reformers. (*See* BENTHAM, JEREMY.) Inspired less by philanthropy than by economy and efficiency, the supporters of the Benthamite or Utilitarian doctrine of "the greatest happiness of the greatest number" were the makers of the modern state, interesting themselves in the Poor law, public health and education. (*See* CHADWICK, SIR EDWIN.)

While middle-class groups were demanding freer trade and attacking the obsolete apparatus of state interference, the preoccupations of government were being gradually extended. Benthamite ideas were popular neither with working-class radicals nor with Tories, but the Whig government was favoured by good harvests until 1836, and despite a political crisis in 1834 which led to a temporary return of Peel, there were few signs until the financial crash of 1836 and the economic depression of 1837 and 1838 that serious domestic difficulties lay ahead.

XVI. THE VICTORIAN AGE, 1837–1901

Early Years of the Reign

Corn Law Agitation.—The first years of the reign of Victoria, who came to the throne on June 20, 1837, were characterized by bitter social cleavages and the rise of extraparliamentary organizations. The background of discontent was a sharp economic depression and a series of bad harvests. Behind the crisis in the economic system there was, however, a more fundamental social crisis. The old obligations of an agrarian society were beginning to break down in many parts of the country and the new social institutions of town and factory were creating problems which baffled contemporaries and excited prophets. The "condition of England" question, as Benjamin Disraeli called it, might be ignored by parliament but it could not be shelved.

The two most important extraparliamentary bodies were the Chartists and the Anti-Corn-Law league. The Chartists struggled for a political formula first presented to the public in 1837—the six points of the "People's Charter"—but behind the formula there was social distress. Chartism (*q.v.*) was a "knife and fork" question as well as a working-class reflex of middle-class political radicalism. The demand for annual parliaments, universal male suffrage, equal electoral districts, the ballot, payment of members and no property qualification for members was not new; what was new was the mass enthusiasm of workers in all the industrial and many of the semi-industrial parts of the country. The Chartists failed to secure any of the six points during their agitation. Local differences, disagreements about tactics and improvement in economic conditions, first between 1844 and 1846 and then after 1848, led to the withering away of the movement.

The Anti-Corn-Law league, founded in March 1839, was essentially a middle-class organization which collected funds from the new manufacturing groups, particularly in the north of England, and organized a vigorous campaign which involved the use of every available device of propaganda, including new media of communication such as the penny post (started in 1840). The formula of the league was a simple one. Repeal of the corn laws would settle the two great issues that faced Britain in the "hungry '40s"—the prosperity of Britain's manufacturers and the condition of the people. The only barrier to the progress of the nation, it seemed, was the landlord. In a landed parliament a few leaguers led by Richard Cobden and John Bright (*qq.v.*) told Peel, who became prime minister in 1841, that "he would be a criminal and a poltroon" if he hesitated to repeal what was considered an immoral as well as an economically restrictive measure. When Peel repealed the corn laws in 1846 he gave Cobden much of the credit for enlightening both himself and the people.

The social bitterness of the 1840s made contemporaries talk openly of class war and revolution. That there was no revolution in 1848 as there was on the continent was a result of the character of Britain's social structure, order and "deference" in the countryside, and links between the classes even in the crowded cities. It also resulted, however, from the willingness of statesmen in the last resort to yield and to accept reform—Grey and the Whigs in 1832, Peel and the Tories from 1841 to 1846.

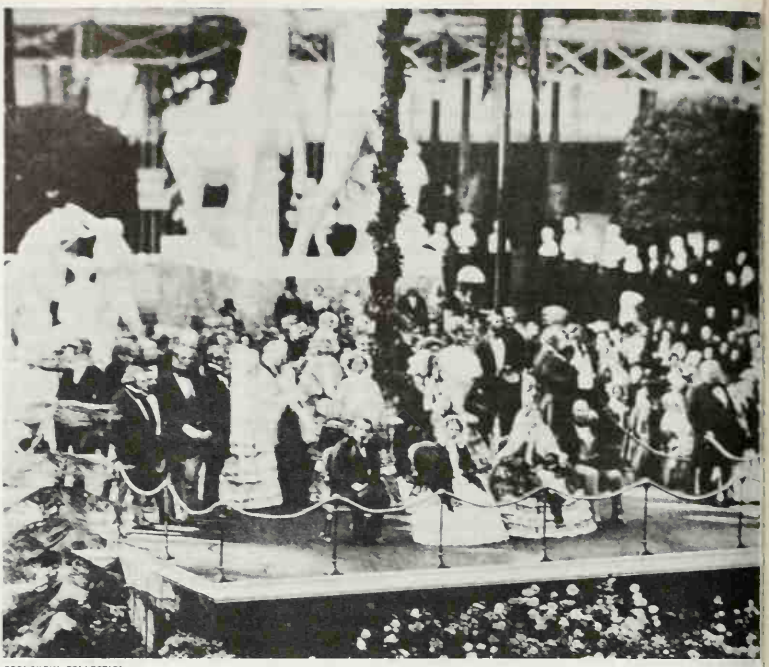
Peel and the Peelite Heritage.—Peel took over from the Whigs in 1841 and turned the parliamentary party which he had built up after 1832 into an active instrument of social and economic change. The Whig ministers under Lord Melbourne (1834–41) had shown little grasp of social and economic questions; Peel, although brought up as a Tory, was a child of the Lancashire cotton industry. Afraid of violence, he realized the need for sound administrative solutions to the "public necessities" of an increasingly industrialized society. He was the presiding genius of a powerful administration—in Rosebery's phrase, "the model of all prime ministers"—strictly supervising the business of each branch of government. Beginning with the budget of 1842 he set about simplifying and reducing tariff restrictions on trade; he reintroduced the income tax in 1842; in 1844 his Bank Charter act laid the foundations of a sound banking system; and finally in 1846 he repealed the corn laws. (*See* PEEL, SIR ROBERT.)

Peel's policy and the strict parliamentary discipline that he imposed upon his followers produced political differences within his own party. Peel himself said he was more surprised that the union of his supporters was "so long maintained than that it was ultimately severed." The repeal of the corn laws brought conflicts to a head. The squirearchy revolted and the parliamentary party was shattered. Peel put his sense of duty to his conscience, to his sovereign and to posterity first; his obligations to his party came second.

The result of the dramatic split of 1846 was a general blurring of party boundaries until 1859. The Peelites remained as a small group of disciples and administrators even after Peel's death in 1850. Some of them, particularly W. E. Gladstone, eventually became leaders of the Liberal party which emerged from the mid-century confusion. The protectionists, most of whom abandoned protection after 1852, formed the nucleus round Benjamin Disraeli, but they were unable to secure a majority at a general election until 1874. Leadership in the mid-century rested with powerful personalities of whom Lord Palmerston was the most prominent.

The repeal of the corn laws did more, however, than change the party situation: it ushered in an age of equipoise, a period of balance between agriculture and industry. The landlord ceased to be considered in the cities as the villain of society, while in the country sharp suspicion of urban industrialism gave way to tacit acceptance and even to understanding. Industry continued to expand and large quantities of British capital were exported overseas. At the same time agriculture profited from industrial prosperity and improved transport facilities within the country. From 1850 to 1873 many of the social cleavages of the 1840s disappeared. As prices, profits and national income rose, all classes of society secured substantial gains in wealth. A new mood of national pride and unity was demonstrated at the Great Exhibition of 1851 when Britain could boast that it had become "the workshop of the world."

The exhibition was a triumph for Victoria and Albert (whom she had married in 1840) as much as for the nation. Neither the queen nor Albert had previously been popular, but during the exhibition they both shared in the delights of the national carnival. "In England loyalty is a passion," wrote a continental observer. Despite outbursts of opposition to Albert, particularly in 1855, the family life of the court began to be considered increasingly as a model for the whole country.



GERNSHEIM COLLECTION

MEMBERS OF THE ROYAL FAMILIES OF GREAT BRITAIN AND FRANCE VISITING THE CRYSTAL PALACE AT SYDENHAM, APRIL 20, 1855. SEATED LEFT TO RIGHT: EMPEROR NAPOLEON III, EMPRESS EUGÉNIE, QUEEN VICTORIA AND PRINCE ALBERT

The death of Albert in 1861 and the withdrawal of the queen from public life led to a decline in the popularity of the court, but in time the queen's subjects came once again to consider both her developed virtues and her obvious limitations as the very essence of Victorianism itself.

Mid-Victorian England

It was during the 1850s that the word Victorian was coined to describe the new national self-consciousness which took political form during the Crimean War (1854–56). In social life, from the 1850s to the 1870s a balance was struck between the busy industrial north and midlands and the sleepy countryside described by Anthony Trollope. New forces such as the railway system, the telegraph and the postal service helped to bind provincial England closer together. "The march of improvement," heralded by the Whigs more than 20 years before, was by now unmistakable. Although writers like Henry Mayhew described the grim underworld of London and the Social Science association directed public attention to the complex problems of change, there was more talk of self-help and of individual advancement than there was of barriers between the two nations of rich and poor. The poor were in any case becoming more "respectable" and the trade unions of the period, particularly the central junta, aimed at establishing their position by proving their "respectability."

Despite obvious progress the Victorians were not, however, uncritical. Writers like Matthew Arnold, John Ruskin and later William Morris dissected the weakness of industrialization. Although there was less cataclysmic prophecy and more analysis of individual belief and conduct, the presuppositions of materialism were not accepted without question. Even during the year of the Great Exhibition, Alfred, Lord Tennyson's *In Memoriam* (1850) considered the triumphs of 50 years of economic advance against the background of the whole long story of man and of the earth. Geology and biology continued to challenge not only the Victorian belief in themselves but also accepted views of religion, handed down from the past.

The late 1850s were frontier years of Victorian thought. In the same year as Samuel Smiles's *Self-Help* (1859) there appeared John Stuart Mill's *Essay on Liberty* and Charles Darwin's *Origin of Species*. A year later came *Essays and Reviews* by Benjamin Jowett, Frederick Temple and others, a contemporary appraisal of fundamental religious questions, which provoked violent controversy.

Despite sharp differences of opinion, most mid-Victorians shared the same ideal of conduct. Even those writers who were most agnostic sought the "religion of humanity" or tried to be good "for good's sake, not God's." There was a dutiful acceptance of the importance of standards in institutional as well as private life; these were years when the extension of the civil service produced its own code of institutional morality.

Religion and Society.—The mid-Victorian emphasis on religion and conduct distinguished this from other periods of English history. The religious spectrum was of many shades. The Church of England was flanked on one side by Rome and on the other by dissent. Both were active forces to be reckoned with. The Roman Catholic Church was growing rapidly in numbers and influence not only in the industrial towns but also among the intellectuals. Dissent was the stronghold of many of the new industrial middle classes although for Matthew Arnold its leaders were "mutilated and incomplete men." In the cities Dissenters often secured a position of dominance in local life. In smaller towns, where the battle between Anglicans and Nonconformists was often bitterly contested, they acted as a militant group, opposing church rates, challenging closed corporations, leading movements of temperance reform, educational improvement and civil advance. The National

Education league of 1869 was a characteristic product.

The Church of England itself was a divided family with different groups contending for positions of influence. The high church movement was given a distinctive character first by the Oxford movement or Tractarianism, which had grown up in the 1830s as a reaction against "liberal Whiggery," and then by the provocative and controversial ritualist agitation of the 1850s and 1860s (*see* ENGLAND, CHURCH OF).

The fact that many members of the church flirted with Romanism and even went over to the Roman Catholic Church raised the cry of the "church in danger." Peel's conversion to free trade in 1846 scarcely created more excitement than J. H. Newman's conversion to Roman Catholicism the previous year, while in 1850 Lord John Russell, the prime minister, tried to make political capital out of the violent antipapal demonstrations stimulated by the pope's creation of Roman Catholic bishoprics. The leaders of the low church movement, particularly the evangelicals, joined with many of the Nonconformists in opposing a common danger. Later they joined with the high church Puseyites (*see* PUSEY, EDWARD BOUVERIE) in opposing such "liberal" approaches to Christianity as were expressed in *Essays and Reviews*.

The evangelicals emphasized the importance of individual conduct and the guidance of the Bible in public as well as in private life. They had acted as the spearhead of the agitation for the abolition of the slave trade which was finally successful in 1833, and in the 1840s their chief parliamentary spokesman, Lord Ashley (later earl of Shaftesbury), was zealous in the cause of the protection of child and female labour, factory reform, mines legislation and public health. It was as a result of his leadership that the important Factory act of 1847, a landmark in welfare legislation, was passed. Along with Christian Socialists, like F. D. Maurice, men like Ashley showed that the church was not completely insensitive to the problems of the new social order.

Outside the influence of both church and chapel there were thousands of people in mid-Victorian England who were ignorant of or indifferent toward the message of Christianity. Although movements such as the Salvation Army (*q.v.*) founded by William Booth in 1878 attempted to rally the poor of the great cities, there was much religious apathy. Toward the end of the century it was accompanied by an increase of skepticism and agnosticism among intellectuals. The great religious controversies of mid-Victorian England were not so much settled as shelved.

Politics.—Religious questions helped to divide the electorate, the Nonconformists leaning toward the Liberal party and the Anglicans often supporting the Conservatives. Party divisions were based on customary allegiance as much as on careful scrutiny of issues, and there was considerable scope for bribery at election times. The Corrupt Practices act of 1854 provided a more exact and complete definition of bribery, but it was not until a further act had been passed in 1883 that election expenses were subjected to a really careful scrutiny. In the middle years of the century the way to parliament often led through a noisy and expensive electoral contest.

The prestige of the individual member of parliament was high and the fragmentation of parties allowed him considerable independence. Groups of members supporting particular economic interests, particularly the railways, could play a prominent part in parliamentary strategy. Contemporaries feared such vested interests less than they feared "the most dangerous of all sinister interests," executive government. Overpowerful government and large-scale reform were considered dangerous to social peace and to economic advance. Politics was the art of wise leadership rather than the formulation of desirable objectives.

In his interesting analysis, *The English Constitution* (1867), Walter Bagehot considered the cabinet as "a board of control chosen by the legislature, out of persons whom it trusts and knows, to rule the nation." Its primary task was to administer and not to legislate. Between the two great Public Health acts of 1848 and 1875 there was no comparable national public health legislation although there was intermittent and limited local action. (*See* PUBLIC HEALTH.) There was no national legislation concerning trade-union problems; the Irish agrarian problem was left almost untouched between 1849 and 1870. The most important statute passed was the Companies act of 1862, which consolidated the limited liability legislation of the 1850s but set out to encourage enterprise rather than to control it. Typical of the trend of legislation was the Bankruptcy act of 1869, which assumed that the enlightened self-interest of the creditor provided the best means for dealing with the affairs of the bankrupt. "Everybody knows," a judge, Lord Bramwell, told a number of men who had been convicted of illegal activities in a strike of 1867, "that the total aggregate happiness of mankind is increased by every man's being left to the unbiased, unfettered determination of his own will and judgment as to how he will employ his industry and other means of getting on in the world."

Palmerston.—Lord Palmerston stands out as the dominant political personality of mid-Victorian England precisely because he was opposed to dramatic change and because he knew how to maneuver gracefully and effectively within the halfway-house constitution of his day. Although this constitution was still based on oligarchical rather than democratic leadership, government was becoming increasingly sensitive to outside opinion working through pressure groups and directed by the local and national press.

In a period when it was difficult to collect parliamentary majorities, Palmerston often, as at the election of 1857, forced decisions on the simple question, "Are you for or against me?" He was very skilful in using press and public opinion to increase his power. At a time of party confusion, when the queen might have increased her influence in English politics, he found the answer to royal opposition in popular prestige, carefully stage-managed. The royal influence, although still brought to bear on questions of personality and policy, was becoming persuasive rather than compulsive.

Palmerston's chief preoccupation was with foreign affairs and his approach was diametrically opposed to that of the court, which saw European questions from a German angle. The supremacy of British sea power, the economic ascendancy of the middle classes and the fervent belief in "liberal institutions" gave Palmerston his opportunity to follow an active foreign policy of his own. He reacted against what he called somewhat unfairly "the sliding-scale policy" of Lord Aberdeen, foreign secretary from 1841 to 1846, a policy of caution in international dealings and reconciliation with France. Palmerston, by contrast, opposed Louis Philippe in France, encouraged liberal constitutionalists in Italy and refused to identify Britain with the forces of counterrevolution after the revolutions of 1848. Ideas never ran away with him, however, and in 1848–49 he preferred a powerful Austrian empire as a bulwark against Russia to a liberal-nationalist Hungary.

Palmerston was always assertive and controversial. He sent a blockading squadron to Greece in 1850 to enforce payment of debts due to Don Pacifico, a British subject from Gibraltar, and restated the doctrine of *civis Romanus sum* in a Victorian setting (*see* PALMERSTON, HENRY JOHN TEMPLE; PACIFICO, DAVID). Although dismissed by Russell in Dec. 1851 for welcoming Louis Napoleon's *coup d'état* without consulting the cabinet, he had his revenge a year later by taking a leading part in the tit-for-tat overthrow of Russell's shaky government. In Feb. 1855, after Aberdeen had shown his incompetence as a leader in the Crimean War (*q.v.*) and the war machine as a whole had revealed alarming inadequacies, Palmerston was recalled as prime minister and was successful at the peace treaty of Paris (1856) in having the Black sea closed to battleships and opened to commerce.

Palmerston's interventions were not confined to Europe. In 1840–41 he forced open the China ports to foreign trade and by the Nanking treaty (1842) acquired Hong Kong for Great Britain. In 1857, in a further dispute with China, he went to war and, after being defeated in the house of commons, appealed triumphantly over the heads of the "factions" to the constituencies outside. Although his government was defeated in 1858, he was back again as prime minister a year later.

In his ministry of 1859–65 it was impossible for Britain to dominate the international scene as effectively as in previous periods of Palmerstonian power. His policy had been founded on the maxim, "we have no eternal allies and we have no perpetual enemies; our interests are eternal," but its success depended on a

forceful opportunism. When in 1863 Palmerston encouraged the Danes to resist the Germans, hoping that Prussian power (now directed by Bismarck) would thereby be restrained, the maneuver was this time ineffective. Prussian troops occupied Schleswig and Kiel.

High-sounding phrases could not cover "the meddle and muddle" of the ministry. The union of modern Italy, the American Civil War and the rise of Bismarck's Germany were reshaping the world in which Palmerston had secured his greatest triumphs. When he died in Oct. 1865, it was clear that in foreign relations as well as in home politics there would have to be what Gladstone described as "a new commencement."

In home politics only the continued influence of Palmerston had delayed overdue reforms. The 1859–65 government, which included Whigs, Peelites—now shorn of their separate organization—and Radicals, was the cradle of the Liberal party, and in the large urban constituencies the demand for a new and active liberalism was gaining ground throughout the 1860s. The former Peelite W. E. Gladstone (*q.v.*) was beginning to identify himself not only with the continued advance of free trade, but with the growing demand for parliamentary reform. In 1864 he forecast a new turn in politics when he claimed that the burden of proof concerning the cause of reform rested not with the reformers but with their Conservative opponents. A year later he lost his seat at Oxford university and appeared "unmuzzled" as representative of a populous Lancashire constituency. The death of Palmerston was followed by the reopening of the reform question and eventually by the passing of the second Reform bill in 1867.

"Leap in the Dark."—It was Benjamin Disraeli and not Gladstone who claimed the credit for the act of 1867. On Palmerston's death Russell and Gladstone introduced a modest and colourless Reform bill (1866) which was severely mauled by Conservatives and by a group of antireform Liberals, led by Robert Lowe, who formed themselves into what John Bright christened "the Cave of Adullam" (*see* ADULLAMITES). The government resigned and Lord Derby and Disraeli took over for the third of their minority governments. It was impossible to shelve the demand for parliamentary reform and Disraeli, who was willing to "dish the Whigs" and "take a leap in the dark," was compelled by Radical opinion to carry an extreme measure (*see* BEACONSFIELD, BENJAMIN DISRAELI, Earl of).

The county franchise was not substantially changed, but in the boroughs a household franchise, conditional on payment of rates, and a lodger franchise, based on lodgings of £10 annual value, were introduced. Householders who paid rates through their landlords were given the vote. The electorate was almost doubled, 938,000 new names being added to the register, and in the towns the artisan masses now formed a clear majority. Although there was no thorough redistribution of seats, 45 new seats were created by taking one member from each borough with a population of less than 10,000. Disraeli hoped that in return for passing this daring measure the urban artisans would vote for his party at the first election held under the new conditions; instead in 1868 they voted for Gladstone and provided him with a large majority.

Gladstone and Disraeli.—The choice between Gladstone and Disraeli was the first of many similar choices subsequently offered to the electors. The two men became bitter rivals, and by the intensity of their conflict captured the imagination of contemporaries and accustomed them to the working of a two-party system. The parties took root in the country as much as in parliament. With the emergence of clearly defined party leaders and the creation of national party systems the role of the crown was reduced to that of merely registering the result of a general election. Although the queen greatly preferred Disraeli to Gladstone, she could not keep Gladstone out. Her obvious partisanship made some of her acts seem "unconstitutional," but they would not have been considered so in any previous period of English history.

Gladstone and Disraeli, although they both sharpened party differences and enhanced the prestige of parliament, differed greatly in their methods. Gladstone made his mark first and lived on far longer than his rival. Feeling his way forward in politics and seeing issues in terms of principles, Gladstone was always suspicious of what he considered Disraeli's cleverness and cynicism. Disraeli, in turn, was unimpressed by Gladstone's political righteousness and for the most part unmoved by the "exuberance" of Gladstone's "verbosity." The clash between the two men brought out the genius in each and they became the centres of every political storm inside and outside parliament.

Gladstone's ministry of 1868–74 was complemented by Disraeli's ministry of 1874–80. Each administration had striking results to its credit. Gladstone's ministry, though it disintegrated in its last phases, carried through the greatest program of specifically liberal reforms of the century; Disraeli's ministry carried through an impressive program of social reform and marked an important stage in the building up of the British empire.

The main achievements of Gladstone's first administration included the disestablishment and partial disendowment of the Church of Ireland, accomplished in the face of the opposition of the house of lords in 1869; the Irish Land act conceding the principles of secure tenure and compensation for improvements (1870); W. E. Forster's Elementary Education act establishing dual church and state control of schools (1870); and the Ballot act introducing the secret ballot for all parliamentary and municipal elections (1872). There were many other important reforms aimed at equalizing individual opportunities or reforming cumbrous administrative machinery. In 1871 the older universities were thrown open to Dissenters (University Tests act); in the same year, as the last step in a chain of reforms that went back to 1855, the civil service, with the exception of the foreign office, was thrown open to competitive examination.

In 1873 the Judicature act, amended in 1876, simplified the forest of legal institutions and procedures, while between 1868 and 1873 the cumbrous military machine was renovated by Edward (afterward Viscount) Cardwell (*q.v.*). The system of dual responsibility of commander in chief and secretary at war, and the system of purchase of commissions and promotions were abolished.

Many of these great reforms did not satisfy the interests affected. The Irish Disestablishment act failed to placate the Irish and alarmed English churchmen. The Education act was passed only in the face of bitter Nonconformist opposition; Dissenters objected that Forster's system provided neither compulsory nor free education and that the monopoly of the church was barely broken in many parts of the country. Although in 1880 education was made compulsory and in 1891 the small fees charged in many elementary schools were abolished, Nonconformists continued to oppose church influence in education and struggled hard to turn the new "school boards," set up in 1870, into strongholds of secular control. If Gladstone alienated the Dissenters he also alienated their most strenuous enemies, the brewers. The Licensing bills of 1871 and 1872 so roused the brewers and distillers that they threw all their influence behind the Conservative party. "We have been borne down in a torrent of gin and beer," wrote Gladstone, as he studied the results of the 1874 election. Gin and beer were not the only cause, however: more than a year before the election, Gladstone had tried to resign, and the Liberal ministers had been described by Disraeli as a "range of exhausted volcanoes."

Disraeli's ministry directed its attention to matters of social reform rather than to the political reforms which Gladstone favoured. Gladstone preferred a free but cheap government to one which concerned itself actively with social interference. He was anxious indeed to abolish the income tax on which the public finance of the future was to depend. Disraeli had always been interested in the "condition of England" question and, with the assistance of men like R. A. Cross, the home secretary, was anxious to justify his reputation as a social reformer. By the Employers and Workmen act of 1875 masters and men were placed on an equal footing as regards breaches of contract, while as a result of new legislation trade unions were allowed to engage in peaceful picketing and to do whatever would not be criminal if done by an individual. The Public Health act of 1875 created a public health authority in every area; the Artisans' Dwellings act of the same year enabled local authorities to make a start on slum clearance; the consolidating Factory act of 1878 set out an extensive code of regulation; while further legislation dealt with friendly so-

cieties, land improvement carried out by tenants and the adulteration of food.

Foreign Policy.—If there was a significant difference between the character of Disraeli's domestic policy and that of his predecessor, there was an even bigger difference in their approach to foreign affairs. Gladstone had never believed in Palmerstonian foreign policy. He considered the best guarantees of international justice to be "firm but moderate language" to deter the strong and the molding of an enlightened public opinion in Europe. British power and credit, he maintained, "form a fund, which in order that they may be made the most of, should be thriftily used." Holding such views, Gladstone often created the uneasy impression of lack of vigour in action. This was the case when he allowed Russia to abrogate the treaty of Paris in 1870 and to rebuild Sevastopol as a naval base, and when he referred the "Alabama" dispute with the United States to arbitration. Gladstone was always anxious to avoid war, while Disraeli was willing to take risks in foreign policy in order to enhance British prestige, and to profit from rather than to moralize about foreign cleavages and dissensions.

Disraeli's first ventures—the purchase of the Suez canal shares in 1875 and his addition of the phrase "empress of India" to the queen's title—showed that he had abandoned the view that colonies were millstones round the mother country's neck, but did not in any way involve him in European entanglements. Nor did the costly campaigns of Lord Lytton in Afghanistan and Sir Henry Bartle Frere in South Africa. It was the near-eastern crisis of 1875–78 which tested his diplomatic skill and produced the liveliest popular discussion of the issues of British foreign policy that had ever taken place. In May 1876 Disraeli rejected proposals made by Russia, Germany and Austria to deal with the armed dispute between the Turks and the Serbian rebels, that had broken out in the summer of 1875 (*see* EASTERN QUESTION). He committed Britain thereby to a pro-Turkish policy, which was denounced by Liberals as a breakaway from the European concert and a violation of all humanitarian principles. Gladstone re-emerged from his retirement—which had begun after his defeat in 1874—and in the famous Midlothian campaign rallied opinion against the "inhuman and despotic Turks." Disraeli's calculations concerned imperial and strategic necessities rather than ideals of conduct, and as the dispute was prolonged and extended, opinion swung to his side. In April 1877 fighting broke out between Russia and Turkey and in Jan. 1878, as the Russians advanced, after considerable difficulties, to the neighbourhood of Constantinople, the cabinet ordered the Mediterranean fleet to steam through the Dardanelles. There was war fever in London, which was intensified when news came through that a treaty had been signed at San Stefano by which the Turks completely accepted the maximum Russian demands. Reservists were mobilized in England and Indian troops were sent to Malta. The crisis passed and at an international conference held in Berlin in June and July 1878 the inroads into Turkish territory were reduced, Russia was kept well away from Constantinople and Disraeli returned with "peace with honour" and the cession of Cyprus. This was Disraeli's golden hour, yet at the general election of 1880 it was the Liberals and not the Conservatives who won a large majority.

LATE VICTORIAN ENGLAND

Economic and Social Change.—From one point of view the Liberal victory of 1880 was the last triumph of "peace, retrenchment and reform," the crusading slogan of the middle years of the century, yet the size of the majority was attributed by contemporaries to industrial and agricultural depression as much as to the spell of Gladstone or his program. The 1870s had brought to an end the period of self-confident prosperity, which had reached its height in the great boom of 1868–73. Although total national production continued to increase and there were improvements in productivity in particular industries, there was a fall in profit margins and an increasing pessimism concerning Britain's industrial future. World economic power was shifting. Britain's monopolistic position in many branches of world industry began to be lost. Finally the challenge of foreign competition and the increasing height of foreign tariff walls began to stimulate demands for "fair" as contrasted with "free" trade.

At the same time, although the real wages of employed labour rose as prices fell, there was heavy unemployment in certain areas and industries and bitter labour disputes about money wage reductions took place. Unskilled labour, organized by new trade unions, adopted methods which had been put aside in the previous 30 years. Collective bargaining and the struggle for higher wages and better working conditions were emphasized rather than the maintenance of mutual benefit schemes and the prudent acquisition of funds. Conflicts in industry were accompanied by a severe depression in agriculture. As a result of improved railway transport in the United States, better storage and shipping facilities and improved agricultural machinery, the competition of transatlantic cereals at last began to be effective in England. In 1877 the price of English wheat, which had averaged 54*s.* 6*d.* in 1846, averaged 56*s.* 9*d.* a quarter; it never came within 10*s.* of that figure again for the rest of the century. During the 1890s there was a sharp fall in rents, a shift in ownership, a challenge to the large estate and an exodus from the village. The landed proprietor was never again to dominate the social scene.

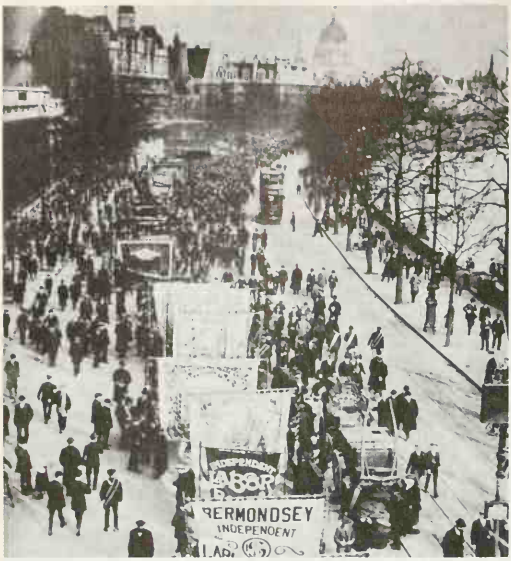
With the shattering of the mid-Victorian social balance there was a marked transformation of economic and social attitudes. Groups with particular economic interests sought sectional advantages while political parties, more bitterly divided than before, adopted new methods and took up new issues. The growth of a Liberal party caucus in some of the constituencies was followed by the creation of national party machines, particularly the National Liberal federation, set up in 1877. Energetic Conservatives, like Lord Randolph Churchill, attempted to rebuild the organization of their own party in the same way, particularly after Disraeli's death in 1881.

The desire to break away from old and accepted modes of thought and behaviour was intensified after the triumph of political democracy in 1884 and 1885. The third Reform bill of 1884 completed the progress toward universal male suffrage by granting the vote to agricultural labourers, while the Redistribution act of 1885 broke up two-member constituencies and robbed 79 towns of less than 15,000 people of their separate representation. The electorate was immediately tripled—indeed more voters were added to the register than had been added by the first two Reform bills taken together. As a result of the redistribution, the dominance of the southern over the northern constituencies and the supremacy of agriculture over all other economic interests at last came to an end. The reforms ignored the traditional influences of property and wealth and embodied the democratic principle that the franchise ought to be conferred upon individuals as a matter of right and not of social expediency.

The acceptance of democracy—although the struggle for women's suffrage was still to come—brought a long period of history to a close and ushered in a new one. The fight for the vote gave way to the battles about its use. As different groups saw the political possibilities of the new situation the shape of politics was inevitably changed.

Labour.—One of the most important changes was the rise of a distinctive labour movement. During the debates on the second Reform bill, the Adullamite leader, Robert Lowe, had prophesied that the workingmen of England would "awake to a full sense of their power" and "set up shop for themselves" as an independent party. For a time labour worked as a pressure group rather than as a political party, determined, in the words of the Labour Representation league, set up in 1869, to "watch the progress of bills in parliament which deal with working-men's interests." A group of working-class members of parliament, led by Henry Broadhurst, were so closely associated with the Liberal party that they began to be called "Lib-Labs"; in the short parliament of 1885–86 the group had 11 members. Broadhurst was bitterly attacked by a rising young Scottish trade unionist, James Keir Hardie, who in 1887 urged the Trades Union congress to sponsor a party independent of both Liberals and Conservatives.

An independent Scottish Labour party was set up in 1888; and the Independent Labour party, most of its support coming from the industrial north of England, was founded in 1893. The



GERNSHEIM COLLECTION

LABOUR AND TRADE UNION DEMONSTRATION AT THE EMBANKMENT, LONDON: 1902

outlook and social philosophy of the party were very different from those of the Marxist parties of the continent and from the Social Democratic federation, led by H. M. Hyndman. The Independent Labour party was securely grounded in British soil, but before it could become an active political force it had to win over the trade unions. The task was not easy, but in 1899 the Trades Union congress itself passed a resolution urging that a special congress of representatives from "co-operative, socialistic, trade union and other working-class organizations" should be summoned in order to devise ways and means of "securing the return of an increased number of Labour members to the next parliament." In Feb. 1900 the projected conference met and the Labour Representation committee was founded. James Ramsay MacDonald was appointed as first secretary, and after early disappointments was able to build up the 20th-century Labour party.

Quite apart from the movement for labour representation there were other tendencies making for an increase in the appeal of socialism in the last 25 years of Victoria's reign. While writers such as William Morris identified socialism with the cult of beauty in art as well as with the development of class solidarity in politics, the Fabian society, founded in 1884, advocated the gradual reorganization of society and the economy on more efficient lines. Skilfully it began to permeate many nonworking-class sections of the community with a belief in administrative socialism. The mid-Victorian emphasis on self-help was giving way to an analysis of the wastes and weaknesses of capitalism. Although it was in the 1890s that the tendencies became plain, it was the 1880s which were the years of the great divide. Traditional forms of leadership were not shattered overnight, and old personalities continued to wrestle with old problems, but Beatrice Webb, one of the new Fabians, was right in describing the Gladstone ministry of 1880–85 as "the no-man's land between the old radicalism and the new socialism."

Gladstone, Chamberlain and the Whigs.—The Gladstone administration, which hastened the processes of change, itself felt many of the strains and tensions which were disturbing society as a whole. In 1880 Gladstone assembled together not so much a compact party cabinet as a creaking coalition. Eight of the members were Whigs; but of the other three, one, Joseph Chamberlain, was representative of a new aggressive radicalism, the product of the cities, less interested in the traditional statement of Liberal individualism than in the strivings and aspirations of a mass electorate. Already, as mayor of Birmingham (1873–76), Chamberlain had embarked upon a large-scale program of civic improvement, which he did not scruple to call "municipal socialism."

The Whigs were the largest group in the cabinet, but the smallest in the country. Many of them were already abandoning liberalism for conservatism before 1880 and the swing increased between 1880 and 1885. Others found themselves increasingly ill at ease in a society where their power was no longer taken for granted. All lost ground as a result of the abolition of two-member constituencies in 1885, for they had long been able to clamber into power as copartners of Radical candidates. While the Whigs feared rapid reform, the Radicals appealed to the people "from whence we come and whose cause we plead." Free land, free schools and free church played an important part in all their many unauthorized programs.

In terms of political logic it appeared in 1880 as if the Gladstonian Liberal party would eventually split into Whig and Radical components, the latter to be controlled by Chamberlain. Gladstone himself held the temporary balance. "I don't see how we are to get on if Mr. Gladstone goes," Chamberlain himself admitted. That English politics did not split in this obvious way, that Whigs and Radicals found themselves allied in 1886 against Gladstone, who was neither Whig nor Radical, was the result of the impact of the Irish question on English politics. But before the Irish question reached a crisis, the government had to face difficult problems of foreign and imperial policy. The Transvaal rebellion of 1880–81 and the Boer victory at Majuba led Gladstone to recognize Boer independence under a vague British suzerainty. In Egypt, where Arabi Pasha had forced the khedive's submission, the government decided that military intervention was necessary to protect British interests, and after a brilliantly successful campaign (1882) Sir Evelyn Baring, afterward Lord Cromer, was appointed British agent there (*see* EGYPT: *History*). In the Sudan however the cabinet followed a disastrously weak and indecisive policy which contributed to the defeat and death of Gen. Charles Gordon (*q.v.*) at Khartoum. This disaster made Gladstone extremely unpopular.

The Irish Question.—Irish politics had influenced England directly at earlier points in the 19th century. The Act of Union of 1800 had served as an expression of Ireland's subjection and of England's supremacy; it provoked a series of Irish political movements in which national, religious, social and economic questions were intermingled. Each movement had an important impact on English politics, precipitating crises and forcing decisions (*see* IRELAND, *History*).

The strength of Daniel O'Connell's Catholic association, formed in 1823, intimidated Wellington and Peel and led them to pass Roman Catholic emancipation, thereby hastening the political crisis of 1830–32. The Irish famine of 1845–46 provided the occasion and to some extent the stimulus for the repeal of the corn laws. The dependence of the Whigs on the Irish Catholic members of parliament complicated the difficulties of government, particularly during the political confusion of 1850–52. Finally, and most important of all, the compact Irish party, built up by Charles Stewart Parnell (*q.v.*) after the death of the more moderate Isaac Butt in 1879, helped to revolutionize the procedure of the house of commons, to destroy the Liberal party and thereby to divert public attention for almost 20 years from possibilities of social reform in England to problems of empire overseas.

Parnell's objective was Home Rule for Ireland. In order to secure it, he was prepared to fight on every front—to organize obstructionist tactics at Westminster and to mobilize Michael Davitt's Land league, founded in 1879, in Ireland. Gladstone met Parnell's first plan of campaign by enforcing stricter rules of closure in parliamentary debate; to meet the second, he turned to land reform. His Land act of 1881, which revolutionized the British concept of property, was based on "the three F's"—fair rent, fixity of tenure and free sale—but it did not go far enough to satisfy Parnell or the Land league. It was accompanied, in consequence, by a coercive policy, which included Parnell's arrest and imprisonment. Irish tension did not slacken even after Parnell's release,

and Lord Frederick Cavendish, the new Irish secretary, who was sent by Gladstone in May 1882 as a harbinger of good will, was murdered, along with the permanent undersecretary, within a few hours of landing in Ireland.

Between 1881 and 1885 Gladstone's administration encouraged a severe policy in Ireland, but the Irish question forced itself into the open again in 1885 when, during the brief Conservative ministry of the summer of 1885, the earl of Carnarvon, the lord lieutenant, followed a more moderate policy than his Liberal predecessors. At the election of 1885 on the wider franchise, Parnell secured every Irish seat outside Ulster but one; Irish voters in English constituencies were urged to vote Conservative. It soon became clear after the election, however, that Lord Salisbury, the prime minister, intended to abandon all Conservative contacts with Parnell and to accept Carnarvon's resignation.

The way was being prepared behind the scenes for the great crisis of 1886. Gladstone was coming to the conclusion that Home Rule was the only solution for the Irish question. His conversion was gradual but profound. It immediately alienated from him most of the Whigs and many of the Radicals. Gladstone hoped that Home Rule could be carried by party agreement, but Salisbury had no intention of imitating Peel. The Conservative government was defeated in Jan. 1886 on a Radical land reform amendment and resigned. Gladstone became prime minister and appointed John Morley, a Home Ruler, as Irish secretary.

In April Gladstone introduced his Home Rule bill. The Liberal party was bitterly divided and 93 Liberals united with the Conservatives to defeat the measure. Gladstone appealed to the country and was decisively defeated at the general election. Whigs and Radicals joined together to form a Liberal-Unionist bloc, which lay between Salisbury's government and the Gladstonian opposition. All attempts at Liberal reunion failed and in 1895 the Liberal Unionists joined the Conservatives in a coalition government. Chamberlain, their leader, like many others of his class and generation, ceased to regard social reform as of the first political importance. While the way was thus left open for the rise of a distinctive "gospel of labour," Chamberlain himself turned increasingly to problems of empire. It was the building of the empire that dominated English politics at the end of the century.

Salisbury.—Salisbury, the Conservative prime minister from 1886 to 1892 and from 1895 to 1902, stood aloof from many of the political tendencies of his times. He was far from believing either in the new political methods or in the often crude statements of popular political doctrines. Aristocratic and detached, Salisbury had been opposed both to the democratic and to the imperialist kinks in Disraeli's policy.

Salisbury's cautious wisdom contrasted at every point with the militant vigour of Chamberlain. Yet important reforms were carried out between 1886 and 1892, particularly the Local Government act of 1888, which set up elected county councils. The influence of the magistrates in the counties was reduced, while the larger boroughs became county boroughs with their own local administration. The way was thus prepared for the development of democratic government in the countryside.

Salisbury's main interests, however, lay in the field of foreign affairs. He was a prudent caretaker. To him the two powers which constituted the chief threat to peace and stability were France and Russia, although he also feared a coalition of all the European powers, who might band together to pounce on the desirable and "divisible booty" of the British empire. Salisbury was reluctant to conclude binding alliances, but he was not averse to reaching ententes with Austria-Hungary, Germany and Italy. He was a skilful negotiator, and his diplomacy was put to good use in the complicated negotiations concerning the partitioning of Africa among the great powers during the 1880s.

At the general election of 1892 Gladstone returned to office at the age of 82, dependent on the support of the Irish members. For three years there was an uneasy interlude in English politics before Salisbury returned to power. Gladstone was still determined to carry Home Rule and with the help of the Irish secured a majority in the commons in 1893. His bill was defeated in the lords and in March 1894 he resigned. The Liberals were clearly an exhausted and divided party and Lord Rosebery, who was prime minister from 1894 to 1895, was neither popular with the rank and file nor able to hold together a cabinet that included imperialists and anti-imperialists, economizers and collectivists. Liberalism was indeed passing through a crisis, just as Whiggery had passed through a crisis 15 years before. At the election of 1895 it was the Conservative party which could claim that it was the popular party, backed by the urban as well as by the rural electorate. Although Salisbury stressed the defensive aspects of conservatism, Chamberlain and his friends were able to mobilize considerable working-class as well as middle-class support for a policy of crusading imperialism.

Expansion of the Empire.—The word imperialism has been used in many different ways and covers many distinct movements, but for most Englishmen in the 1880s and 1890s it implied a sense of national mission and a dream of territorial expansion.

The attitude to the empire changed during Victoria's reign. In the 1830s colonies were little esteemed; they were thought of rather as expensive luxuries, which could not remain, and indeed ought not to remain, as permanent pieces of a single empire. Nonetheless the empire grew, largely as a result of the local pressures of pioneers, settlers, traders and missionaries. Large countries, such as Canada, which in the course of the 20th century became dominions, were given considerable powers of self-government and showed little desire to break away from the larger whole. The Indian mutiny of 1857 led the government finally to force the East India company to transfer its territories to the crown. When in the 1880s writers like Sir John Seeley and politicians like Sir Charles Dilke began to create an imperialist ideal they had the fact of an existing large empire on which to pivot their dreams of expansion in the future. Between 1884 and 1902 2,500,000 sq.mi. of territory fell under British control and when in 1895 Chamberlain chose to take over the colonial office rather than any other position in the cabinet, he was acknowledging the opportunities, both economic and political, afforded by this great "undeveloped estate." The same Radical energies that he had previously devoted to civic improvement were now directed toward imperial problems, and the empire was made significant and attractive to the newspaper-reading electorate.

The cause of empire was associated not merely with the economic interest of businessmen and the enthusiasm of workingmen, but also with the traditional lustre of the crown. Victoria's jubilee in 1887 was the occasion of a demonstration of loyalty to the throne, of national patriotism and of imperial pride. The self-governing colonies sent their prime ministers to London to take part in a conference, which was significant historically not because of the particular matters it discussed, but because it inaugurated a series of similar conferences in the future.

The second jubilee in 1897 had as one of its main features an even bigger conference, which decided that in years to come such conferences should be held periodically as regular instruments of imperial collaboration. There were many differences of opinion between the colonies concerning politics, trade and defense, but the differences seemed less important than the fact that the British empire had established itself as a whole world in itself. Rudyard Kipling's well-known poem "Recessional" summed up the spirit of the year.

Between 1896 and 1902 public interest in problems of empire was intensified not by pageantry but by crisis. The presence of British and Boer elements in South Africa created a precarious and delicate situation which was always liable, particularly after the opening up of the Transvaal gold fields, to lead to an open conflict. The policies of Cecil Rhodes, prime minister of the Cape and pioneer of British imperialism in Africa, and Paul Kruger, Boer president of the Transvaal, were sharply antagonistic, and after the Jameson raid of Dec. 1895, when a force of Englishmen crossed into Transvaal territory, British-Boer relations were extremely tense. Between 1896 and 1899, while British elements in South Africa were trying to put pressure on a not always reluctant colonial office, Kruger was attempting to reach an understanding with other European powers. In Oct. 1899, after negotiations broke down between Kruger and Sir Alfred Milner, the British representative, war began. The early stages of the struggle were

favourable to the Boers and it was not until the spring of 1900 that British numbers and equipment began to count.

In June 1900 British troops entered Pretoria and when Kruger fled to Europe in October it seemed that the war would end. Instead the Boers began to adopt guerrilla tactics and the fighting went on until May 1902 (*see* SOUTH AFRICAN WAR). Just as the Crimean War had focused attention on weaknesses in the whole machinery of the state, so the South African War taught a lesson that might otherwise have been difficult to learn.

Before the war ended, Queen Victoria herself died, after a brief illness, on Jan. 22, 1901, in her 82nd year.

Queen and People.—The death of the queen directed attention toward the relationship between Britain's past and Britain's future. Victoria's long reign of 63 years had covered not one but several Victorian ages, each with its distinctive outlook and momentum. The England of Trollope was so far removed from the England of the six acts that Walter Bagehot complained "we cannot comprehend" that the latter ever existed. The *fin de siècle* England of the 1890s was far removed from both previous periods. All Victorian institutions were challenged: "the old sobriety of mind had left our shores," wrote a contemporary, Richard Le Gallienne, "and we changed from a stolid into a volatile nation." In politics the transformation was just as marked. The England of Peel was separated by a plateau and an abyss from the England of Chamberlain. The queen provided the one element of continuity, and her long reign spanned more social changes than any previous reign of English history.

The richness of the Victorian past was not confined to its growing wealth or to its literary achievement. The whole country had been reshaped. Its population more than doubled between 1837 and 1901, new towns sprang into existence, new assets were acquired abroad and new institutions were created yearly.

Such signs of progress inspired contemporaries in 1901, but did not entirely convince them. The future seemed far less certain than it had in the past. New classes, new nations, new types of men were beginning to make themselves felt. What was the ultimate destination? What indeed was the direction of travel? A self-conscious generation, which had already been searching for novelties in the 1890s, was not sure of the answers, but the queen's death and the beginning of a new century forced it to stop and think. It had little time for meditation, for new problems proved pressing and both the domestic and the international situation soon produced new ideas and policies. (A. BRI.)

XVII. THE 20TH CENTURY

In the 20th century, World Wars I and II were agencies of cataclysmic change as well as periods of exhausting strain and wounding loss. In these struggles for survival Britain expended the last reserves of its Victorian heritage at home and overseas. Insular security disappeared and Britain's primacy as a world power was lost. State control was extended and in every sphere of activity uniformity increased. But constitutional liberties were preserved and national unity was maintained.

EDWARDIAN ENGLAND (1901–14)

The death of Queen Victoria and the accession of Edward VII, a genial man of the world with wide interests and a zest for life, might suggest the passing of a dead weight of tradition. But in social and artistic matters the tradition had for some time been weakening; and in national affairs, once the excitement of these events was over, the old seemed to be depressingly present and the cry for the new depressingly unanswered.

South African War.—The South African War dragged on; while it lasted, it emphasized the differences between the various groups within the Liberal party and consolidated Conservative or Unionist strength. Lord Rosebery's Liberal Imperialists were confronted by a pro-Boer group of Liberals led by a rising young Welsh politician, David Lloyd George. A middle group of Liberals emerged but it was not until after 1903 that party rifts were healed. The Unionists won the "khaki election" of 1900 and secured a new lease of power for six years, but their unity too was to be broken by new issues.

In 1901 the Liberals who were attacking the conduct of the war found a new theme. Lord Kitchener, who was now in command in South Africa, was meeting guerrilla warfare by clearing whole districts and confining their inhabitants in concentration camps, which were often overcrowded and unsanitary. The Liberal leader, Sir Henry Campbell-Bannerman (*q.v.*), denounced them as "methods of barbarism." The whole subject, bringing to such a definite point the irreconcilable difference of opinion between those who felt that the war must be fought to a finish and those who merely demanded that it should be stopped, exacerbated people's minds. Rioting broke up a peace meeting in Birmingham in December and forced Lloyd George to flee for his life.

Within a few months the whole scene was changed. By March 1902 the Boer leaders were known to have opened communications with Lord Kitchener. Peace was signed in May at Vereeniging on terms generous enough to the Boers in the matters of amnesty and repatriation, with a grant of £3,000,000 for re-establishing devastated farms, but also conclusively establishing the unity of South Africa under the British government.

Tariff Reform.—So far the nation had been paying little attention to events which suddenly promoted a much more healthy state of party opposition in the commons. For some time past it had been felt that the school boards of the 1870 act, though they had worked well in towns, were on the whole too parochial in personnel and policy to view education in the large and generous way that was becoming increasingly necessary. The Balfour education act of 1902 abolished them and placed education in the hands of statutory committees of the borough and county councils. For the moment the significance of the fact that Church of England schools were brought into the reorganization on equal terms with the undenominational board schools escaped notice; then the old Liberal Nonconformist spirit awoke, internal divisions lost their importance and the opposition became a fighting force. It was not until the autumn session, when, as a drastic assertion of the rights of conscience against legislation, nonpayment of rates was launched as a definite policy, that this change in the atmosphere of the house of commons had much reverberation in the world outside.

The situation had hardly time to develop seriously before it was lost in a far more startling political storm. On May 15, 1903, Joseph Chamberlain (*q.v.*) made a speech in Birmingham in which he pleaded for an imperial preferential tariff to bind the empire more closely together. At first sight there is something astonishing in the speed with which this pronouncement became the cardinal factor in public affairs, but three considerations provide an explanation. First, there had for several years been increasing complaint of the unfair position of the British manufacturer in an open home market against the protected competition of the foreigner, and of the "dumping" of foreign products, to which much of the unemployment in Great Britain was attributed. Secondly, there had been, in remarks by recent chancellors of the exchequer about a need for broadening the basis of taxation, hints of a reconsideration of fiscal policy, to which the reimposition of a shilling "registration" duty on grain during the South African War had given point. Thirdly, the commanding position of Chamberlain in the cabinet, and his popular reputation as a vigorous driving force beside the somewhat placid, academic figure of A. J. Balfour (*see* BALFOUR, ARTHUR JAMES BALFOUR, 1st Earl of) were enough to bring together in a flash, when he made such a pronouncement, all the half-suggestions of a change in the tariff system. It is no wonder that this speech, though, in fact he had used much the same language to the imperial conference of 1902, became the storm centre; nor is it any wonder that Liberals raised at once in its broadest aspect the old issue of free trade and protection.

During the session of 1903 it remained principally a house of commons question; C. T. Ritchie's attitude in introducing the budget and the discussions on the budget revealed a strong free-trade element among the Conservatives. But Balfour refused to admit any vital difference between himself and Chamberlain, and reiterated that the whole question was one for inquiry and investigation, not for immediate party warfare. In the autumn of 1903 the campaign on both sides developed rapidly outside parliament. It soon became apparent that Chamberlain and the free traders

could hardly remain in the same cabinet. Resignations were to be expected; when they came, they seemed only to make the situation more bewildering. The notable free traders, Ritchie, Lord George Hamilton, Arthur Elliot, Lord Balfour, resigned in September. But Chamberlain had resigned as well. In the reconstruction his son, Austen Chamberlain, succeeded Ritchie at the exchequer. Another distinguished free trader, the duke of Devonshire, resigned in October.

The controversy did not occupy the whole field. In 1904 a liquor licensing bill, the outcome of much discussion about the excessive number of licensed public houses in many towns in proportion to the population, set up a statutory compensation fund for unrenewed licenses in every quarter sessions area. It added a minor weapon to the Liberal armoury inasmuch as it gave for the first time implicit legal recognition to the growth of a legitimate expectation of steady renewal of a license, and limited the discretion of the justices.

The achievement of 1904 that most justified Balfour's continuance in office was undoubtedly the agreement concluded with France. Its terms provided for French recognition of preponderant British influence in Egypt in return for a similar recognition of French interests in Morocco. The agreement was taken as a sign that the policy of "splendid isolation" was being in some measure abandoned. The establishment of better relations with France was attributed largely to King Edward's influence and to his personal popularity in Paris. From that time he was considered to have found a sphere where a constitutional monarch might play an important part.

If 1904 was thus a year that kept some breadth of interest in politics, 1905 was so nearly a year of exclusive concentration upon tariff reform that Balfour, by remaining in office, seemed to do little more than make a present to his opponents of 12 months' electioneering. Under the new Tariff Reform league on one side and the revived Cobden club (the old free-trade Anti-Corn Law league) on the other, the battle was fairly joined. Reputations were made on the free trade side; Lloyd George, already prominent from his opposition to the South African War, greatly increased his power over Liberal meetings; and Winston Churchill, who had refused an equivocal position and left the Unionists in 1904, reached the front rank of politics at a bound. As the year drew on to the inevitable election, argument naturally deteriorated. Tariff reformers' cries of "the foreigner will pay" and "tariff reform means work for all" were matched by Liberal posters of the big loaf and the little loaf. Steadily it became more and more clear on which side the British public as a whole was ranging itself. On Dec. 4, 1905, Balfour at last resigned. Some faint hope existed among the Unionists that the internal divisions of the Liberals, the legacy of the South African War, might revive to weaken them. But when Sir Henry Campbell-Bannerman's government was announced on Dec. 16 it was seen that H. H. Asquith was at the exchequer, Sir Edward Grey at the foreign office and R. B. Haldane at the war office, all these being members of the Liberal Imperialist group to which Campbell-Bannerman himself conspicuously did not belong. A speech by Lord Rosebery a few days later made it clear that nothing had gone on behind the scenes; Liberal Imperialists had simply accepted the prime minister's invitation to join him.

The New Liberal Government.—The general election that followed was the most startling that the country had seen. Member after member of the former government lost his seat, including Balfour himself. The new house was composed of 397 Liberals, 57 labour members (of whom 29 belonged officially to the Labour party, the others being Liberals elected by and representing labour), 83 Irish Nationalists and a mere 157 Unionists. The government had a possible maximum majority of 374. Free trade had beyond all question played the largest part in the results; the introduction of indentured Chinese labour into the Rand gold mines had also provided Liberals with an electoral weapon which they used with effect. It was obvious, too, that Balfour's long continuance in office after his party had become disunited had stimulated the electorate's desire for a change, and had also revived the distrust of the government's efficiency that had been so vocal during the South African War.

There was another important element in the situation. While labour had been rapidly becoming more politically energetic and more articulate it was not yet in a position to appear in full force as a distinct party, and therefore its effect at the polls was a heavy reinforcement of the Liberal vote. Two things above all had profoundly affected labour. One was that in 1899 the various elements in labour politics—trade unionism of the old-fashioned strictly industrial kind, Fabianism and socialism—had been drawn together in the foundation of the Labour Representation committee; in the intervening years this first success was consolidated; the party pledge, after some internal struggles, was imposed, with a compulsory levy, on members of certain trade unions. But what most of all inspirited it and helped toward the acceptance both of the party pledge and the levy was the famous Taff Vale case of 1901 in which, after a railway strike, the unions had to their dismay found that in law they could be held liable for damages for the actions of any individual union member during a strike. The Labour world naturally drew together wholeheartedly to demand amendment of the law; and since Liberal candidates sympathized entirely with this demand, the Taff Vale case must be added to the reasons for the overwhelming Liberal success. Another reason was the labour demand for a more active social policy generally. The early years of the century were on the whole free of serious labour disturbances. But active political labour was aware that employers were working on an ever-narrowing margin of profit and that, with prices slowly but distinctly rising, there must be determined and difficult wage struggles ahead; the wider social conscience was increasingly concerned about unemployment, "sweated" industries, bad housing and the helplessness of the manual worker in sickness and old age. (*See* also LABOUR [TRADE] UNION.)

The narrower social conscience, however, viewed such problems differently. A Conservative party unusually devoid of discriminating leadership sought to recover through the house of lords the position it had lost at the polls; if defeated in the lords it was even prepared to countenance unconstitutional behaviour outside. Thus the history of the next eight years is one of considerable legislative achievement in the presence of forces of opposition which, by the very nature of the case, tended more and more to weaken parliamentary and ministerial authority in the mind of the public.

Much was achieved. The Trade Disputes act of 1906 placed trade union funds above the normal operations of the law in actions of tort; a reorganization of the military system in 1907 created in the expeditionary force and the territorial army a machine that was to meet the first strain of World War I in 1914; the Coal Mines Eight Hours act and the Old Age Pensions act of 1908, the Housing and Town Planning act, the Trade Boards act for "sweated" industries and the Labour Exchanges act of 1909 were all passed. But there was perpetual friction with the house of lords; many bills were lost, and in matters of education and licensing especially the government was not allowed to carry its proposals. To Nonconformist Liberals who chafed under Balfour's 1902 settlement the defeat of a new education bill was particularly vexatious.

Struggle With the House of Lords.—Liberal opinion became convinced that the lords did not even pretend to act properly as a house of parliament when Liberal measures were in question. But since no generalized attack upon the lords could succeed, battle had to be joined on a new and large issue; the issue should be one of social policy, to secure labour support and, if possible, one that would bring into the open the kind of opinion that was behind the house of lords. Liberal strategy based itself, very skilfully, on David Lloyd George's 1909 budget. If the lords threw it out they would startle the nation by an action which, from long disuse, had become constitutionally almost unthinkable. They would appear to be refusing openly at last to foot the bill for necessary reforms and they would give the impression that their hostility to the government was based entirely on selfish capitalist interests. The government's second budget had introduced differentiation between earned and unearned incomes, a graduated income tax and a surtax death duty on estates of more than £1,000,000; the 1909 budget made further differentiation between earned and unearned incomes,

applied the surtax to large incomes as well as large testamentary dispositions and laid a tax on the "unearned increment" of land enhanced in value by industrial or other developments in the neighbourhood. The lords based their opposition on the fact that the "unearned increment" tax involved a land valuation scheme, which was not legitimately part of a finance bill. No doubt, too, they relied upon the obvious ebb of the amazing tide of 1906 and upon a fairly general feeling that Liberal legislation had already gone rather far. Therefore, accepting the challenge, they threw out the Finance bill and opened a struggle which occupied the whole year. Two general elections were held—the first in Jan. 1910 for authority to force through the "people's budget"; the second in Dec. 1910 for authority to carry the measure that was to end such struggles between the two houses—the Parliament bill.

In both cases the government was able to achieve its object; in neither case, however, in such a manner as to put an end to the opposition's questioning of the nature and true extent of its authority. The election of Jan. 1910, on the issue of the budget, seemed to support the view that, free trade apart, the Liberals had no overwhelming mandate from the nation; they lost about 100 seats to the Conservatives, and were consequently largely dependent upon the Labour members and the Irish Nationalists for their majority which was, however, substantial and secure. The opposition accepted this as a decision on the budget, which was passed; but mainly because the holding up of the national finances was an awkward and dangerous way of carrying on a constitutional contest. There remained the question of the obstructive powers of the house of lords as such. To deal with it the government introduced the Parliament bill.

It was an ingenious measure. Postponing for later consideration the possibility of an amended upper house and the powers it might be given, it proceeded to lay down that, until such reform, measures that had passed the commons unaltered in three separate sessions might be presented for the royal assent without the consent of the lords; and that financial measures might be so presented without repeated passage a month after passing the commons. It was the gravest constitutional measure since the Reform bill of 1832—perhaps since the Act of Settlement; there were those on both sides who looked for some agreement that might evade such drastic action. It was significant of the position King Edward had attained that hope, at this crisis, pinned itself to his tact and sagacity; but his death in May 1910 removed the hope centred in him; and it had in fact been cherished rather by the popular mind than by serious politicians. Conferences continued through the year, until in November it was announced that they had failed. The gravity of the announcement of another appeal to the country was enhanced by Asquith's statement (he had become prime minister on Campbell-Bannerman's death in April 1908) that he had only advised dissolution upon such conditions as would enable the Liberals, if successful at the election, to place the Parliament bill on the statute book against all resistance by the lords. This was immediately taken as a reproduction of the guarantee given by William IV to Lord Grey for the passage of the Reform bill; although in the present case this must mean reducing the house of lords to an absurdity by a mass creation of 400 or 500 peers, there could be no other interpretation of Asquith's words. The election of Dec. 1910 made little change; a gain of four seats by the Liberals was hardly conclusive, but the government held steadily on its way with the Parliament bill; the Conservative leaders shrank back at the last and in Aug. 1911, by the abstention of the majority, the bill passed the house of lords by 131 votes to 114.

Years of Unrest, 1911–14.—George V (1910–36) came to the throne (May 6, 1910) in the middle of the crisis, which was only to be the first of many in his reign. Once the constitutional battle was over the cost of the victory began to appear. Extreme Conservative opposition in the country began to feel that against a Liberal and Labour majority thus freed of the power of the house of lords to permanently block legislation, any and every form of obstruction and resistance might be justified. Unfortunately the temper of the nation had become one in which such notions could flourish. Militancy was in the air; it was to be the keynote of the next three years and no class shrank from it. Indeed, restraints of every kind seemed to be disappearing. The old exclusive society was being invaded by new wealth. Cubism in painting and new conventions in sculpture ruthlessly scrapped all the rules that earlier generations had valued; a new music did the same with conventions of rhythm and harmony. In literature and drama George Bernard Shaw, H. G. Wells and John Galsworthy were preaching new and disturbing social gospels.

The government, after the interruption of 1909–10, returned to its social program, passing in 1911 the Shops act, establishing the weekly half holiday, and providing for payment of M.P.'s. But the principal measure—the National Insurance act providing for health and unemployment insurance—was met with the new kind of opposition. A violent newspaper campaign attempted to bring about a strike of the medical profession against the duties which the act would lay upon them, and actually succeeded in inducing a large meeting of respectable, and even important, women at London's Albert hall to pledge itself not to "lick stamps" to affix to the cards of insured persons. It was absurd, but it had its alarming side. The government also set itself to fulfill the statutory conditions for placing Welsh disestablishment and Irish Home Rule on the statute book by passing them through the requisite number of sessions.

There were other causes of anxiety. The hopes raised in the first flush of the agreement with France, that relations with all foreign powers would grow happier, had waned. Grey, the foreign secretary, extended the entente by agreeing to the inception of talks between the general staffs of Britain and France. By a settlement of outstanding disputes with Russia he paved the way for an Anglo-Russian entente which assisted in the consolidation of a triple entente that matched the dual alliance of Germany and Austria. Simultaneously, however, Grey strove for a better understanding with Germany, but his efforts were frustrated by the mounting naval rivalry of the two powers and Germany's attempts to drive a wedge between Britain and France. The launching in 1906 of the "Dreadnought," a new type of battleship more powerful than anything afloat, was at once a response and a challenge to Germany's shipbuilding program.

In 1908 Haldane's "peace mission" to Berlin served only to reveal the persistent incompatibility of British and German policies. The Agadir crisis of 1911 (*see* EUROPE: *History*) over Franco-German rivalries in Morocco was also a German attempt to test the strength of the entente. Lloyd George, in a speech at the Mansion house, warned Germany that Britain, in effect, would not stand idly by if undue pressure were put on France. In 1912 a second Berlin visit by Haldane evoked only a more uncompromising German refusal to abate its naval program.

Though these were ugly portents, it was domestic affairs that mainly engrossed the attention of the British public. The revolutionary trade unionism called syndicalism was finding its way into England from France, and the first Labour daily newspaper, the *Daily Herald*, appeared (1911). Trade union membership was increasing; from 2,500,000 in 1907 to 4,000,000 in 1913 (including those not affiliated with the Trades Union congress). In 1912 a strike of coal miners showed as irreconcilable a temper on either side as the railway strike of the previous year had shown, and, forcing parliament to pass a minimum wage bill for coal mines, raised spectres of what "direct action" (one of the tenets of syndicalism) might come to mean.

With so much hostility in the air, so much social and intellectual restlessness, the nation hardly noticed at first the precipitous slope down which it was hurrying in Irish affairs. Nothing could prevent the Home Rule bill from becoming law in 1914. Unionists concentrated, as in 1885–86, on the injustice to Ulster; but this time Lord Randolph Churchill's phrase of the earlier years, "Ulster will fight and Ulster will be right," was transformed into action. As early as Nov. 1913 armed resistance was in preparation. In March 1914, when some officers stationed at the military training camp at the Curragh, in County Kildare, sent in their papers rather than face the carrying out of orders which the government might shortly have to issue, the public generally woke up to the fact that civil war was at the gates. In April arms and ammunition were being openly landed at Larne and distributed by motor transport throughout

Ulster and a "provisional government" had been organized. During the first stages of the bill in the commons and the lords efforts for peace were still being made; the Home Rule bill itself, if it were to obtain the sanction of the Parliament act, could not be amended; but the government produced in June a bill allowing Ulster counties to "contract out" of Home Rule. It met with so much amendment in the lords that there seemed to be no hope at this stage of reconciliation. When it was announced on July 20 that the king had summoned a conference in the privacy of Buckingham palace, matters were felt to be grave indeed. The graver stage still, when the conference, after four meetings, broke up without having reached an agreement "either in detail or in principle," hardly had time to impress itself on the mind. For on July 28 a Balkan incident that had been occupying some space in the newspapers culminated in a declaration of war by Austria upon Serbia, and after a week of suspense Great Britain was at war.

World War I

The course of the war is described in the article World War I; all that is attempted here is a sketch of events and conditions in England. Altogether the nation entered more smoothly and more unitedly upon war than could have been thought possible. It was satisfied with the immediate changes in the cabinet; Lord Morley, John Burns and one or two others resigned, and the announcement that Lord Kitchener, at home as commander in chief, had become secretary of state for war was thoroughly popular. He appealed immediately for volunteers and, without much regard to his warning that the war would last four years, civilians poured into the recruiting offices. The warning was taken seriously by hardly anyone.

Much of the stimulus to recruiting came from reports of German atrocities in Belgium. It took time for the grim horrors of trench warfare to impinge on the public mind. The most serious check to popular optimism in the winter of 1914 was provided by reverses at sea—the German shelling of Hartlepool and Scarborough and the destruction of three cruisers by German raiders off Coronel in the Pacific. The admiralty's reaction, however, was amazingly swift and restored confidence in Lord Fisher, the first sea lord. He instantly dispatched battle cruisers which fell in with the German ships at the Falkland Islands in December and destroyed them.

The long casualty lists of the winter grew with the failure of the Neuve Chapelle offensive in March and people at home began to realize what the troops in the field had long known—that the war was going to demand wholly new conceptions of massed artillery. This, and what appeared to be delay in the use of the new armies, roused a storm of criticism. One beneficial result was the setting up of a ministry of munitions. Under an extension of the Defence of the Realm act all factories concerned with munitions had been taken over by the government in March.

The Coalition Government.—In May 1915 dissatisfaction with the Liberal government's conduct of the war found expression in a Conservative demand, strongly backed by Lloyd George, for a coalition government. Asquith acceded and reconstructed his cabinet to include the Conservative leaders and also Arthur Henderson as a representative of Labour. Lloyd George's energy was felt, even by those who had in the past regarded it as a national danger, to be in a good place at the head of the new ministry of munitions. Meanwhile, a new phase of the war had been launched by the Dardanelles expedition. The public, knowing little of the divided mind of the government in spite of Lord Fisher's resignation from the admiralty, welcomed the campaign as offering an attack upon the farther side of the Central Powers and as a move to keep the Balkan states from intervention.

The heavy casualties of the Dardanelles landing and the drain of manpower, steady and appalling, of the trench warfare in France, roused the nation to a new realization of its task. The sinking of the passenger liner "Lusitania" on May 7, 1915, roused the public more than the loss of many merchant vessels, but the summer did bring some concern for food supplies. During the summer and autumn the first zeppelin raids brought the war home in a new way. As the pressure on manpower increased, women took over men's jobs on the land, in factories, in canteens and offices. By 1917 they were in uniform as auxiliary forces attached to the army, navy and air service.

Throughout 1916 the real heart of the war was seen to be the western front and it was a year of heavy fighting. In July came the British offensive on the Somme, making new demands for troops. The result was the introduction of a conscription bill, applied to all men between 18 and 40, with exemption for those engaged in absolutely necessary national service and those who had a conscientious objection to military service. In another way the nation's manpower had its problems. There was an amendment of the Munitions of War act, to force upon workshops and factories that still tried to maintain old union rules the "dilution" of labour by unskilled men. This and the growing fear that the Conscription bill would in effect mean military discipline in factories led to a serious strike on the Clyde in March and April, ignited by revolutionary opposition to the whole war. In finance new efforts had to be made. The country's credit had responded amazingly to the strain; but now, besides the raising of the income tax to five shillings in the pound, increase of excess profits duty and other efforts of the budget, a campaign was started for war savings certificates to include the small investor and spread the fiscal burden. Continuance of the zeppelin raids brought pressure upon the government for improvements in the air forces.

Twice in the first half of 1916 the public received sudden shocks. At the end of April armed rebellion broke out in Dublin; it was followed by many executions on the spot, and months later by the trial and execution of Sir Roger Casement for high treason in securing German help for the uprising. The rebellion was quite unexpected; Augustine Birrell, the chief secretary, acknowledged that the significance of the new Sinn Féin movement in Ireland had been underestimated. The Home Rule act had been placed upon the statute book soon after the outbreak of war, but it had been accompanied by an act, suspending its operation until after the war. The other shock came in June when Lord Kitchener, starting out to visit the Russian front, was drowned when the cruiser in which he was travelling struck a mine. Accusations of treacherous betrayal of his plans lasted for years, so sharp was the effect on the public mind. A great deal of uneasiness was also caused by the one great naval affair of the war, the battle of Jutland, which was fought on May 31; not all the power of the censorship could make it appear a victory.

The Deterioration of Morale.—There were, as the year went on, several outbreaks of public hysteria. Members of the government were charged with having treasonable sympathies; Lord Haldane, whose reorganization of the military machine had done more than anything else to enable the country to take the first strain of the war, had had to resign from the government because much of his education had been in Germany. Conscientious objectors were increasingly hounded and many of them were sent to prison. A food controller was established in November, a recognition of the growing menace of the submarine to food supplies. A strike in the South Wales coal field led to putting the coal mines under the Defence of the Realm act. Then in Dec. 1916 came the culmination of press attacks and private intrigues against Asquith's leadership, which, whatever its merits, certainly lacked fire and popular appeal. Conscious of his declining popularity, he resigned. Lloyd George succeeded to the leadership of the coalition and formed a "war cabinet" of only five members with Bonar Law, the Conservative chief, as leader of the house of commons and a ministry largely staffed with "businessmen" new to politics. He set up new ministries of labour, food, shipping and national service.

Crisis of the War.—There was in 1917 the feeling that the crisis of the war was at hand, especially with the first important news of the year, the great withdrawal of the Germans to the Hindenburg line. But for the rest of the year events conspired to make a strained fluctuation of hopes. The capture of Vimy ridge in April had been preceded in March by the Russian Revolution. The third battle of Ypres in July and the capture of hill 70 in August were followed by the Caporetto disaster to the Italian army; and the long dreadful struggle in the mud of Passchendaele in November was followed by Russia's withdrawal from the war in December. The main result at home was renewed agitation about

manpower and pressure upon the government for a re-examination of discharged men and for drastic combing out of all civilian and noncombatant occupations.

The submarine threat to merchant shipping became acute in April and May, but the government still hesitated to impose food rationing. The entry of the United States into the war provided assurance of ultimate victory, freed the navy from restrictions of an all-out blockade caused by a large body of neutral shipping and guaranteed finance and supplies in eventual abundance, but its immediate effect, resulting from the United States's own war needs, was felt in the form of greater stringency. The worst of the war's severities still lay ahead. Huge government loans were floated and taxes were further increased. Prices rose and scarcities developed. There were several bad strikes, largely over "dilution" problems. Dissatisfaction mounted with the virtual stalemate on the western front. In August Arthur Henderson resigned from the government because he was not allowed to attend a labour peace conference at Stockholm. In November Lord Lansdowne, a Conservative leader, advocated a negotiated peace in a letter to the press. The government, which in a different sense looked forward to the peace, prepared for it by passing two measures which before 1914 would have roused tempests of controversy—H. A. L. Fisher's Education bill and a franchise measure establishing adult suffrage for men and also for women over 30 (*see* WOMEN'S SUFFRAGE).

Although compulsory rationing for some foods came before the end of Feb. 1918, the worst of the submarine menace was over. But on land the success attained by the German offensive in April shook even the most confident. Criticism fell on the generals, although the Allied line, reinforced just in time, held long enough to turn the tide. Sir Frederick Maurice, director of military operations at the war office, was dismissed and in the parliamentary debate that followed Asquith led the attack on Lloyd George. The immediate result was the establishment of a unified command under the French general, Ferdinand Foch.

As the war moved to its climax, its demands on the home population intensified. In July rationing was extended until it covered sugar, fats and meat including bacon. Another service bill called out men up to 50 years of age; the comb-out required more and more elderly men to fill up the noncombatant parts of the service; and with this last calling up it became apparent that there was only the comb-out to depend upon now for supplying the wastage of the trenches. Another considerable strike in munition works at Coventry spread to Birmingham and into Yorkshire before the government's threat to draft the strikers into the army caused it to collapse. At last, in August, came the long awaited Allied offensive, and on Nov. 11, 1918, the armistice.

The Peace.—Little was before the electors in the general election of Dec. 1918 except a loose and far too easy summary of the mind of the moment provided by Lloyd George. Some of it was an idealistic, if rather sentimental, program of social amelioration—"Homes fit for heroes." The rest of it was hasty generalization about bringing to book those responsible for atrocities—"Hang the Kaiser"—and about indemnities and reparations from Germany. Loose appeal as it was, it succeeded handsomely.

Those members of parliament, even of Lloyd George's own party, who had supported Asquith in the Maurice debate were denied the so-called "coupon"; *i.e.*, the coalition government's letter of endorsement. Thus the great majority of the coalition government candidates were Conservatives and the Liberals were split into two factions. The coalition candidates were returned 478 strong, of the Asquithian Liberals only 27 survived and Labour became the official opposition party with 59 members. The Labour party, though little heeded at the time, had in fact equipped itself to become a fully national party. In 1918 it had revised its constitution to provide for both individual and trade-union membership and in its program, *Labour and the New Social Order*, drawn up by Sidney Webb and other Fabians, it announced the doctrines of gradual socialization which were to provide its policies for the next 30 years.

Lloyd George, departing with the other British representatives, Balfour, Bonar Law and G. N. Barnes, for the opening of the peace conference in Paris (*see* PARIS, CONFERENCE OF; VERSAILLES, TREATY OF) which held its first session on Jan. 18, 1919, left to his colleagues the troubles which soon began to arise.

INTERWAR YEARS

Domestic Difficulties.—No one paid much attention at that moment to the fact that Sinn Féin had not only secured 73 of the 77 Irish seats, with candidates pledged not to come to Westminster, but had also on Jan. 22 opened a "republican parliament" in Ireland. There were troubles nearer at hand. Miners' grievances broke out again in a strike, only composed for a time by the appointment of a royal commission. Indeed, there was an epidemic of strikes. Worse still, all plans for gradual demobilization were being ruined. Long before the end of the war the government, foreseeing the difficulties of returning 5,000,000 men to civil life, had begun to make plans for demobilization, based as far as possible on the probabilities of the labour market. But this did not meet the impatience of the troops. There were serious troubles in demobilization camps and in the end demobilization had to be carried out wholesale, and problems of unemployment left to chance.

The interest of politicians was centred on the peace conference at Paris, where a good many hopes were vanishing, rather than on the progress of home affairs. The setting up of a ministry of transport and the changing of the local government board into the ministry of health suggested that some lessons of organization had been learned, and a housing bill did something to redeem some of the election promises. In an autumn session a further stage of experiment in the government of India was entered upon by the act giving effect to the proposals of the Montagu-Chelmsford report (*see* INDIA-PAKISTAN, SUBCONTINENT OF: *History*).

Meanwhile the Irish situation deteriorated swiftly. The government was trying to meet it in the old way, with a mixture of coercion and concession to demand in the shape of a new Home Rule bill, introduced on Feb. 26, setting up two Irish parliaments, one for Ulster and one for the rest of Ireland, with a single council of Ireland to bring about harmonious action. This met to some extent the Ulster-difficulty, though the proposal for a single council was mistrusted. It did nothing to meet the demands of the Sinn Féiners, who not only refused to accept any partition of Ireland but demanded nothing less than an independent Irish republic. The government reinforced the Irish constabulary with men who had served in the war; these men went to Ireland without any of the constabulary traditions, and soon acquired a grim notoriety. From their temporary uniforms they were known as the Black and Tans. Ambushings, kidnappings, raids, marked by merciless killing on both sides, with violent street fighting occasionally in Dublin and Belfast, seemed to be working up to a bloody climax of English and Irish relations. With this kind of war on its hands in Ireland, with the rumours of war in Russia, where assistance was still being given to anti-Bolshevik armies, with the incident at Amritsar in the spring, where a threatening Indian crowd had been fired upon in a confined space, with many casualties, the coalition government had little to satisfy it at home. Rising prices provoked wage demands leading in turn to strikes. In 1919 the railway men struck for nine days. In 1920 the miners struck when the government failed to act upon the nationalization clauses of the coal commission's report (clauses, however, which had only been passed on the casting vote of the chairman, Justice John Sankey). In the winter unemployment became so alarming that parliament had to make hurried grants to enable municipal bodies to undertake schemes of work to absorb labour.

Meanwhile mounting budget deficits led to a demand for drastic economy and the limitation of state action in every direction. Controls were lifted on railways, mines, food and drink, prices and agriculture, the last involving the abandonment of the minimum wage for agricultural labourers. The coalition government was giving up every means of achieving that organization of the country for a better national life which had been its appeal to the electorate in Dec. 1918. In Aug. 1921 a committee for economizing in the public service—the "axe" committee—was set up under the chairmanship of Sir Eric Geddes. Decontrol of the railways was, indeed, accompanied by some small remnant of the war's lessons of organization, in an amalgamation of systems which left only four

great railway companies. Decontrol of mines, another indication that the old piecemeal struggle was to be left to go on, led to another strike in April 1921 and to the calling out of the reserves and armed defense of the mines. At the end of June Lloyd George was able to announce a settlement, but he had bought it by a subsidy of £10,000,000 to allow maintenance of wages.

Decline of the Coalition.—In the latter half of 1921 Irish affairs took a turn for the better. Negotiations had been opened in May, but many difficulties had to be met (*see* IRELAND, REPUBLIC OF: *History*) and it was not until the early days of December that terms of peace were arranged. No small part of the difficulty had been caused by an uncompromising attitude on the part of the die-hard Unionists in parliament, who maintained their opposition throughout the debates on the ratification of the Irish treaty. By the beginning of 1922 their opposition had developed into a serious rift in the coalition. Disagreement arose between Lloyd George and Sir George Younger, the Unionist chief whip, about the best date for an election, Lloyd George again being for precipitation; and the rise of a discussion about the distribution of political honours, alleged to have filled an electioneering war chest which was under Lloyd George's control, hastened the desire to draw the Unionist party clear of the coalition. There were other influences at work besides Unionist disaffection. After the boom came the slump. Moreover the economy measures meant not only drastic cutting down of the public service, but the abandonment of any really generous assistance for housing and little prospect of much more subsidy for work for the unemployed—still numbering more than 1,500,000. The failure of the 1922 Genoa conference of European states to "settle Europe" was followed by a period of real tension between Great Britain and France over the operations in Asia Minor which ended in the complete rout of the Greeks by the Turks. A coalition which could give no fruits of peace at home and seemed actually to risk peace abroad had reached the end of its tether. Moreover there had been growing a strong feeling that the power, and especially the electioneering power, behind the coalition was speculative capitalism. This was what gave sharpness to the charges about the sale of party honours. The war with Bolshevik Russia could be regarded as an expensive affair undertaken for the financiers whose loans had been repudiated. The troubles in Asia Minor were openly put down to the scramble of oil interests for territory to exploit. The end came suddenly. On Oct. 19, 1922, the Conservative members of the coalition met at the Carlton club to consider the position. Stanley Baldwin (*see* BALDWIN OF BEWDLEY, STANLEY BALDWIN, 1st Earl), then president of the board of trade, carried a resolution advocating their withdrawal from the coalition. Lloyd George resigned and Bonar Law formed a Conservative government and appealed to the country (Nov. 1922). They routed the Liberals, still divided into Lloyd Georgeites and Asquithians, and returned 347 strong. But Labour was strengthened also, numbering 159.

Economic Problems.—The new government had almost immediately to deal with a grave turn in relations with France. To the French, reparations from Germany had always been a much more urgent subject than to the British and they were moved by no argument of economists on the subject. On the ground that some form of coercion must be applied to Germany for defaulting on its reparations, France proposed to occupy the Ruhr district of mines and metalworks, and expected the active co-operation of Great Britain, which the government, wholly mistrusting alike the wisdom and the use of such a step, could not offer. Conferences in Paris in December and January ended simply in leaving France a free hand, as amicably as might be in the face of French anger at this "betrayal" of the alliance. Stanley Baldwin, the new chancellor of the exchequer, negotiated the funding of Great Britain's debt to the United States. It stood at £978,000,000, and was funded with the interest fixed at 3% for the first ten years and 3½% for another 52 years, with a sinking fund of ½%. Baldwin's first budget, actually with a surplus of £100,000,000, expenditure being now down to slightly more than £800,000,000, enabled him to make the first slight reduction in the income tax; and good sense approved his application of the surplus to debt reduction rather than to any considerable lightening of taxation. But unemployment persisted and the whole session indicated that the government, while making provision for relief work, was mainly relying on the re-establishment of business prosperity as a cure.

In Nov. 1923, amid general surprise, Baldwin (who had become prime minister in May when Bonar Law retired because of ill-health) sought a dissolution to secure a mandate for tariff protection. Although the McKenna duties of 1915 (*see* MCKENNA, REGINALD) put a tariff wall round the motorcar industry and the 1921 Safeguarding of Industries act legalized protection for key industries and those threatened by "dumping," Britain was still substantially the same free trade country that she had been in the previous century. But Baldwin had been a lifelong protectionist and determined to seek approval for a policy of tariffs and imperial preference. The election reunited the Liberals and cost the Conservatives their majority. Although the votes cast differed little from those of the previous election the result in seats was considerably different—258 Conservatives, 191 Labour, 158 Liberals. Baldwin retained office until parliament met in the new year, while Asquith made it clear that the Liberals would offer Labour their support to form a minority government. J. Ramsay MacDonald (*q.v.*), the Labour leader, accepted the offer and formed a Labour government which, while mainly relying on right-wing members of the party, nevertheless introduced a novel element into the governing circles of the nation.

The First Labour Government.—Nine months of office disposed quite sufficiently of doubts whether Labour could govern. Its ministers not only did well in public; they also earned the good will of their departmental staffs. The prime minister (who also assumed the foreign secretaryship) won much credit for the reparations conference which secured general acceptance for the Dawes plan (*see* REPARATIONS). In domestic policy he insisted on moderation. An unemployment bill did little beyond increasing benefits. John Wheatley's Housing act modifying the form of subsidies was more successful, but lacked the full financial backing it required. Philip Snowden, as chancellor, showed himself to be a Gladstonian financier and a Cobdenite free trader. His budget was free of Socialist proposals and concentrated on "freeing the breakfast table," scrapping existing proposals for imperial preference and abolishing the McKenna duties.

But the government was always on sufferance, and the announcement of its Russian policy brought the sufferance to an end. MacDonald had immediately opened negotiations with the U.S.S.R. in a new spirit. In July the results were announced in the form of a treaty in which, besides arrangements for trading facilities, there appeared a limited recognition by the U.S.S.R. of the debts of the tsarist regime, but in return for this a British loan was to be arranged. While there was much to be said for access to the Soviet market, there was still too much distrust of the U.S.S.R. to permit of such a treaty. But another question had also arisen by then—the dropping of the prosecution of J. R. Campbell, a Communist journalist, accused of tampering with military discipline. In the end it was on this issue that it was decided to defeat the government. It fell early in Oct. 1924 and MacDonald dissolved parliament.

In any case it is probable that the Unionists would have secured a safe working majority, but in the end they secured a majority beyond all calculations. A few days before polling day there was published in a newspaper the notorious Zinoviev letter, allegedly sent from the Comintern to the British Communist party, urging the violent overthrow of the British constitution. The bombshell did its work; middle-class voters were thoroughly scared into belief in an imminent red peril. The Unionists returned 413 strong against 151 Labour members and 40 Liberals. Liberals, in fact, suffered more from the Zinoviev letter than Labour; their less stalwart adherents either resented the fact that a Liberal vote had put Labour into power or felt that a three-party balance must not occur again. Labour, on the other hand, had rallied strongly to the denunciation of the letter as an electioneering trick.

Analysis of the voting showed that the Labour poll had actually gone up by 1,000,000, while the Liberal vote had declined by 1,250,000. But the red scare affected more than the election; Austen Chamberlain, MacDonald's successor at the foreign

office, at once put an end to the Soviet treaty negotiations.

The Coal Crisis.—Great Britain at the beginning of 1925 offered no very cheering spectacle to a new government. In many directions there seemed to be no lack of money and yet unemployment was still showing figures of more than 1,000,000; businessmen and manufacturers were gloomy; the labour world restless and aggressive. The government certainly set out to combine fiscal purity with direct social ameliorations. The budget—the production of Winston Churchill, now a Conservative—was notable for the return to the gold standard and the reimposition of the McKenna duties. With the budget came the promise of a considerable extension of the benefit of pensions, the old-age pension to be granted at the age of 65 instead of 70 and widows of insured men to be pensioned without waiting for old age.

The granting of insurance pay to the unemployed was, however, stiffened up somewhat by insistence of a minimum number of contributory payments to qualify for receipt of the pay. In foreign policy the Geneva protocol—an attempt at an all-round international guarantee of security which the Labour government had negotiated—was rejected in favour of the Locarno treaties.

But throughout the year the real concern lay outside parliament. There were clear signs of the approach of a very serious national crisis. The immediate danger point was, as for so long past, the coal mines. A. J. Cook, a leader of the miners, had early in the year been talking of a wide alliance of all labour to back their claims; and this only put into words an uneasy sense that labour and capital were massing opposite one another as never before. Another coal commission had been appointed in September under the chairmanship of Herbert Samuel; and quiet was preserved for a time by the continuance of a subsidy for keeping up wages. But the differences went deep. The mineowners argued that, because of lost markets and rising costs, they could not grant the terms the men demanded. The miners' reply was that the real trouble was the wastefulness of piecemeal ownership, the struggle of small, poor mines and the selfishness of big rich ones; and their demand was for nationalization or, pending that, such a pooling of mines as would, by throwing the returns of rich mines into the common stock, enable the poorer ones to keep up the wage level. On that issue there was not the faintest hope of agreement.

The General Strike.—The commission's report in March 1926 recognized that the parlous condition of the market necessitated some new arrangements between the two sides, calling especially for some sacrifice on the men's part in longer hours of work on a smaller wage minimum. But it recommended that wage agreements should be national, suggested large reorganization of the industry and contemplated some degree of state ownership in certain circumstances. The government immediately professed a readiness to act on the report, even against some of its convictions, if the two parties concerned would agree. They could not. The mineowners posted late in April their new wage offers, which the miners described and acted upon as lockout notices, and when the government subsidy came to an end on April 30 the stoppage began. On May 1 the govenment proclaimed a state of emergency. The instant reply from labour was the ordering of a general strike to begin at midnight on May 3. Last-minute efforts were made, and labour leaders were in consultation with ministers when the negotiations were abruptly broken by the latter because a strike had broken out in the *Daily Mail* office; compositors had refused to set up a certain article on the situation. The ministerial view was that this was treachery to the whole basis of the negotiations; the labour view was that a small and really isolated incident had been seized upon and magnified because now that war had been declared there were too many on the side of capital who were ready to fight it to a finish. So the strike began. (*See* GENERAL STRIKE.)

For a day or two there was intense anxiety; newspapers could not be published and no one knew how near the country might be to armed conflict. Then the tension slightly relaxed. An official organ called the *British Gazette* was published daily, a paper called the *British Worker* was issued, and one or two newspapers managed to publish in much diminished shape. The British Broadcasting company, still in its infancy, provided a skeleton service of news to those with wireless sets able to receive it. Not that the officially controlled organs tended to induce a peaceful state of mind; the relaxation of tension resulted mainly from the good spirit shown in spite of these publications. The response to recruiting for the maintenance of the four essential services was rapid; and as, on the whole, youth brought to its strange tasks the enjoyment of novelty and a good temper, there was far less friction than might have been expected. Yet it was obvious that tempers which could stand a short strain might break under a long one. Fortunately it was not long. On May 8 the prime minister broadcast a conciliatory message; negotiations were opened and on May 12 the general strike was called off. It had not been wholehearted and it was showing signs of crumbling. But the original dispute in the mines remained as it was and the stoppage of work dragged on for months.

With a budget that again, after the better years, was showing a deficit to the mining subsidy, and with business generally bad, even the strong old industries like cotton and steel being in distress, there would, but for the feeling that labour had put itself out of court, have been a more healthy pressure upon both sides. As it was the mineowners held out for their whole position; and the miners, resentful of the collapse of the general strike, stuck to their own extremists. Yet collapse came upon them too. By September a drift back to work had begun, slowly at first, but increasing week by week, in spite of the leaders' efforts, until by late in November half the men at least had gone back. Surrender came on Dec. 1, when the stoppage was declared at an end.

After the Strike.—There remained the aftermath of the general strike. Financially, it was bound to be grave; the chancellor of the exchequer put the total loss at £59,000,000, and he had to budget for an actual deficit of £36,000,000. Nothing could be done to lighten the burden of taxation. Politically, too, the aftermath looked grave. The government introduced a Trades Disputes bill, making a general strike illegal, and also intimidation and threats, with such definition as was possible of the meaning of the words; further, the bill altered the basis of the trade union levy. Labour party finances largely rested on trade-union contributions levied, since the act of 1913, on all members who did not explicitly contract out. The 1927 act reversed this, substituting contracting in. Receipts fell sharply in consequence, and remained low until a Labour government repealed the act in 1946. An unemployment insurance bill and a bill to foster the production of British motion pictures were also passed and broadcasting in Great Britain passed under government control.

In foreign affairs, the trade agreement with the U.S.S.R. was broken off, after a raid upon a London office building revealed evidence of political activities, again under cover of the trading agreement. There was considerable discussion, too, of matters arising out of the League of Nations. Naval expenditure had been coming heavily upon budgets which in every other direction were being cut down; and the fact aggravated the failure of the disarmament conference in the summer at Geneva. It appeared to have broken down largely because of imperfect diplomatic preparation beforehand between Great Britain and the United States. During the summer there was a steadily growing interest in a draft treaty, originated by F. B. Kellogg, the U.S. secretary of state, which aimed at the renunciation of war as a normal feature of international relations. It was eventually signed in Paris on Aug. 27, 1928, by several great powers, including Germany, and it was hoped that others would sign later. Although impotent to check aggression, it inspired much contemporary optimism about the "outlawry of war."

An incident at the very end of the year—the rejection by the house of commons of a proposed new prayer book, presented to it by the convocations and church assemblies under the Enabling act of 1919—was chiefly remarkable for the quite unexpected public interest aroused in matters of doctrine and religion. A second attempt, in which some endeavour had been made to compromise (mainly in modification of the proposals for reservation of the Sacrament), met with a similar fate in May 1928; and as the compromise proposals had alienated some supporters of the proposed new book, the attempt to revise the forms of 1662 failed.

The Second Labour Government.—The Conservatives, when they went to the country again in 1929, had counted confidently on a sufficient majority to carry them through another parliament and it was a shock to them to find their numbers in the new house reduced to 260 against 287 Labour and 59 Liberal members. The Liberals decided that the constitutional course was to give Labour, as the largest party in the commons, an opportunity of forming a government, and Ramsay MacDonald accordingly became prime minister for the second time.

The legislation of the next two years was scanty. A Coal Mines bill reducing hours from eight to seven and a half and instituting the system of quotas for export fought its way to the statute book after some critical divisions. The house of lords used its remaining authority under the Parliament act to wreck bills for raising the school-leaving age and for introducing a system of proportional representation. The last proposal was particularly dear to the Liberals as a third party labouring under the disabilities of the existing electoral system. The minister of agriculture, Christopher Addison, succeeded in passing a marketing boards scheme and a useful but modest measure for housing in rural districts; and the government set up the Indian Round Table conference which met in London and broke the ground for the coming reform scheme.

Colonial and Foreign Affairs.—The imperial conference met again in 1930 and once more renewed arguments on tariffs and preference, but found the Labour government unwilling to abandon free trade and the governments of the dominions reluctant to grant preferences. The chief work of the conference was to prepare the ground for the Statute of Westminster, which was enacted shortly after the next general election (1931).

The Statute of Westminster laid down that "Dominions are autonomous communities within the British empire, equal in status, in no way subordinate to one another in their domestic or external affairs though united by a common allegiance to the crown, and freely associated as members of the British commonwealth of nations." The definition left many difficult legal and constitutional questions unsettled; *e.g.*, the functions of the crown, the possibility of one or more of the autonomous communities remaining neutral while others are at war and so forth, but mutual forbearance and constant consultation between the different units have made the formula remarkably successful in operation.

MacDonald's main interest, as in 1924, was in foreign affairs. Philip Snowden won popular applause by his plain speaking to France at the Hague conference of 1929 about the British share of German reparations under the Young plan, but whatever he gained was wiped out in 1932 when the Lausanne conference made an end of reparations so far as Europe was concerned. In 1930 the naval disarmament conference, which MacDonald held in London, settled the outstanding question of cruisers between Great Britain and the United States and paved the way for a settlement between France and Italy. Friendly relations were restored with Soviet Russia.

The Slump.—However, the most pressing problems were economic. The great U.S. slump, which began in 1929, had seriously affected international trade and caused a sharp rise in the number of unemployed, which mounted to 2,000,000 and more. The government was greatly opposed to diminishing the scale of relief, but with these numbers the unemployment scheme was rapidly piling up debt. In the last weeks of June 1931 Margaret Bondfield, the minister for labour, told the house of commons that the deficit in the unemployment insurance fund was at the rate of £1,000,000 a week and likely to rise. The borrowing on this account was already £125,000,000.

By this time a royal commission appointed in the previous October had issued a report declaring it imperative to increase the contributions and reduce the benefits of the insurance system and pointing to numerous anomalies and abuses in administration. The government introduced a bill dealing with these anomalies but was unable to adopt the other recommendation of the commission in view of the strong opposition of its own supporters.

At this point a committee under the chairmanship of Sir George May, which had been appointed to inquire into the state of the national finances, issued a very alarming report. It said that at the current rate of expenditure the deficit in the next budget would be £125,000,000, apart from the usual provision for debt reduction. This and the analysis of the situation accompanying it came like a thunderclap at home and abroad. Europe was already in the throes of the crisis caused by the collapse of the great Austrian bank, the Credit Anstalt, and foreigners who had regarded British credit as unassailable took alarm and began withdrawing their money from London, with the result that the Bank of England had to seek assistance from the Bank of France and the federal reserve system of the United States.

The cabinet then appointed an economy committee to deal with the situation. But this committee only revealed the deep differences between its members, many of whom held the bankers responsible for the trouble and wished to seize the opportunity of bringing them under public control. While the committee debated the situation was going from bad to worse, and on Aug. 24, 1931, the prime minister, having decided that it was too dangerous for further delay, tendered his resignation and the next day was commissioned by the king to form a coalition national government.

The National Government.—MacDonald's action split his party from top to bottom. He carried with him only three of his Labour colleagues, Snowden, the chancellor of the exchequer, Lord Sankey, the lord chancellor, and J. H. Thomas, the secretary for the dominions and colonies. Of the rank and file a mere handful followed. The rest denounced his behaviour as betrayal. Both the other parties came into the small cabinet which he now proceeded to form, Stanley Baldwin, Neville Chamberlain, Sir P. Cunliffe Lister and Sir Samuel Hoare representing the Conservatives, and Lord Reading, Lord Crewe and Sir Herbert Samuel, the Liberals. On Sept. 8 the new government met parliament and obtained a vote of confidence by a majority of 60. Two days later Snowden introduced a new budget imposing drastic new taxation, and accompanied it by a memorandum showing the cuts and economies affecting the social services and the entire public service, including army and navy, which the government proposed.

Abroad, however, the run on gold persisted. On Sept. 20, 1931, the Bank of England warned the prime minister that withdrawals were again setting in on a serious scale, and on the following day the chancellor of the exchequer found himself compelled to obtain the sanction of parliament for the suspension of payment in gold. The government had told the country that to maintain the gold standard was a high national interest and one of the chief objects of its drastic dealing with finance, and it had now to argue that the departure from gold was a blessing in disguise. That in fact turned out to be the truth. The overvaluation of the pound when the country returned to the gold standard in 1925 had been a serious handicap to foreign trade and its devaluation to a gold standard of about 16*s.* to the pound, which followed the departure from gold, gave an instant fillip to foreign trade, while the internal value of the pound remained unaffected.

The Tariff Issue.—Though the government had been united about immediate steps, it was far from being agreed upon permanent economic policy. The Conservatives wanted a tariff and an immediate election fought on the protectionist issue. A few Liberals, headed by Sir John Simon, announced their conversion to tariffs on the ground that the balance of trade required it; but the great majority of them stood firmly by free trade.

The Conservatives nevertheless had their way and parliament was dissolved at the beginning of October. The prime minister announced no program; instead he asked for a "doctor's mandate" to employ whatever measures the crisis demanded. Liberals had to be content with this and though officially the government denied that a vote for them was necessarily a vote for a tariff, Conservatives left no doubt that if they dominated the new parliament they would institute a protectionist policy. The great majority of the electors were clear only that they did not wish a return to Labour, and the government was returned by the enormous majority of 502, the Labour opposition being reduced to 52. MacDonald now reconstituted his cabinet, bringing it up to the normal number of 20, of whom 11 were Conservatives, 4 National Labour and 5 Liberals of different shades.

Pressure was immediately applied by the Conservatives to induce the government to institute tariffs. It replied with the Abnormal Importations act which enabled the board of trade to impose duties up to 50% on articles said to be dumped, and the free traders in the cabinet accepted this as an emergency measure to redress the adverse balance of trade. But the pressure continued and when parliament reassembled in 1932 the prime minister announced that the government proposed to lay a general ad valorem of 10% duty on all imports and additional duties to be determined by an advisory committee on articles said to be non-essential. At first the Liberals remained in the cabinet on an agreement to differ which left them free to speak and vote against these proposals, but by Sept. 1932 Sir Herbert Samuel, Sir Archibald Sinclair and Lord Snowden decided that the definite adoption of protection and preference as the settled policy of the government at the Ottawa conference had made their position impossible and resigned their offices.

The chief legislative achievement of this parliament was the passing of the great and complicated measure instituting self-government in the provincial councils and central assembly of a federated India. Much of the time of parliament was occupied with the schemes of the minister of agriculture for controlling the output and prices of such commodities as bacon, milk, hops, potatoes, etc., and for regulating foreign imports by quotas. Subsidies played a large part in the minister's schemes and they mounted to a point which presented an anxious problem to the chancellor of the exchequer.

Foreign Policy.—Foreign policy for the next eight years was dominated by the aggressive behaviour of the German and Italian dictatorships and the proportionate decline in the authority of the League of Nations. Ironically it was in the east, where "splendid isolationism" had been first breached by the Anglo-Japanese alliance of 1902, that the frail structure of League security first began to crumble as it stood apparently impotent over the Japanese invasion of Manchuria (Sept. 1931). But if the handling of the Manchuria incident by Sir John Simon, as foreign secretary, betrayed a marked infirmity of purpose, he could at least plead that Britain was distracted by its economic crises. By the time some degree of recovery had been registered (the unemployment figures began to decline for the first time in 1933), the menace of Adolf Hitler had begun to exercise its paralyzing effect upon British diplomacy. The 1930 Allied withdrawal from the Rhineland and the 1932 abolition of reparations payments had done nothing to check Hitler's rise. The collapse of the disarmament and the economic conferences in 1933 served notice of the failure of international action on both fronts and encouraged a British withdrawal into a new isolationism. The stoppage of British war debt payments to the United States ended, not without acrimony, an unhappy phase in Anglo-American relations. The negotiation of the agreement of Britain, France and Italy at Stresa in March 1935 to collaborate in maintaining European peace, the "Stresa front," represented a pathetic attempt at reviving a concert of Europe.

In June 1935 Sir Samuel Hoare succeeded Sir John Simon as foreign secretary and Anthony Eden was given cabinet rank as minister for the League of Nations. A little later Ramsay MacDonald yielded the post of prime minister to Baldwin but remained in the government as lord president of the council. In October the parliament that had now lasted four years was dissolved, and Baldwin appealed to the country. Although afraid, as he later admitted, to advocate full rearmament, he claimed that the European situation made a measure of it urgently necessary, especially the situation in the Mediterranean caused by Italian dictator Benito Mussolini's Abyssinian adventure. This was an embarrassing issue for the opposition, which found itself deeply divided between its dislike of war and armaments and its criticism of the government for being unwilling to risk a war with Italy. Shortly before the election the Labour leader, George Lansbury, resigned on the ground that even for the sake of the League of Nations he could not support a warlike policy. The election held on Nov. 14, 1935, gave the government a majority of 247, a figure which, though considerably below that of 1931, was in excess of the most sanguine estimate of its supporters. The Conservatives came back 387 strong, the National Liberals had 33 and National Labour 8. Labour, with 154 seats, was far below its strength in 1929 and Independent Liberals were reduced to 20.

In Dec. 1935 the acceptance by Sir Samuel Hoare of a scheme proposed by Pierre Laval, the French prime minister, for settling the Abyssinian question by making certain cessions of Abyssinian territory to Italy involved the government in serious trouble. News of this scheme raised a storm in England where it was regarded as a betrayal of the League and a dismemberment of Abyssinia for the benefit of the aggressor. The government withdrew its support and Sir Samuel Hoare resigned and was succeeded by Anthony Eden. Sanctions imposed by the League of Nations against Italy were continued, but the governments could not agree on the one embargo—that on oil—which might have been effective, and Mussolini rushed his campaign to a conclusion with the aid of tanks and poison gas. The League acknowledged defeat and withdrew its sanctions. Sir Samuel Hoare returned to the government and became first lord of the admiralty.

Meanwhile Hitler, his prestige already heightened by the Anglo-German naval agreement negotiated in 1935, moved his troops into the demilitarized zone of the Rhineland in March 1936. British diplomacy concentrated on dissuading France from taking forcible countermeasures. In July 1936 the outbreak of civil war in Spain threatened a new and dangerous complication between the democratic powers, which sympathized with the Spanish government, and Germany and Italy, which sympathized with the insurgents. A French proposal for neutrality and nonintervention was accepted with some qualifications by all the powers, and a committee which sat in London was appointed to see that it was carried out. It met with little success. In spite of its efforts, airplanes, tanks and munitions of all kinds were supplied to both sides from all parts of the continent and many thousands of volunteers joined the two armies. It soon became obvious that Germany and Italy on the one side and the U.S.S.R. on the other regarded their own interests as vitally involved in the Spanish struggle.

The Abdication Crisis.—The year 1936 had opened sadly, for after a short illness George V, whose silver jubilee (May 6) had been the great ceremonial event of 1935, died at Sandringham on Jan. 20. The popular prince of Wales acceded as Edward VIII. In November and December all other events were overshadowed by the question of the king's marriage, which produced a serious constitutional crisis. King Edward was determined to marry Wallis Warfield Simpson, an American lady who had already divorced one husband and in October obtained a decree *nisi* against another. The silence of the British press on this obviously controversial issue was broken on Dec. 2 by a virtually unanimous editorial warning to the king about the difficulties of carrying out his intentions, making it clear that Mrs. Simpson would not be an acceptable queen of England. On Dec. 4 Baldwin presented to the cabinet the king's compromise proposal of a morganatic marriage, but only to obtain confirmation of the intimation he had already given the king, that such a proposal would be totally unacceptable either in Great Britain or in the dominions. Since the king was firm in his resolve to marry Mrs. Simpson no other course remained but abdication. On Dec. 10 an instrument of abdication, signed by the king, was read in the house of commons by the speaker and all the necessary legislation was passed by both houses the following day. (*See* also EDWARD VIII.)

Edward left the country for the continent with the title of duke of Windsor and his brother, the duke of York, became king as George VI (1936–52). The succession was secured to his children, the princesses Elizabeth and Margaret.

Foreign Affairs.—On May 28, 1937, Baldwin resigned the premiership and was succeeded by Neville Chamberlain. By the beginning of 1937 it had become clear that the balance of power in Europe had altered profoundly. With Italy's secession from the League and its adhesion to the new group—afterward called the Axis—of which Germany and Japan were the other principal partners, the great powers were again divided into two camps; and in the absence of any definite sign from the U.S.S.R., it was by no means certain that the powers—mainly France and Great Britain—which adhered to the League were the stronger. Ger-

many had been feverishly rearming; and although the trend of events abroad led to an acceleration of British rearmament, the cost of which was in 1937 estimated at £1,500,000,000 spread over five years, Britain and France had fallen dangerously behind in the armament race, especially in the air. A foretaste of what this might mean was provided on March 11, 1938, when Hitler forcibly annexed Austria to the *Reich*.

Chamberlain nevertheless retained the hope that Italy could be detached from the Axis. His confidence in his capacity to do this had even led him in January to turn down Pres. F. D. Roosevelt's invitation to call a great power conference at Washington, D.C., with all the promise that carried of Anglo-American cooperation. Instead he negotiated, April 16, 1938, an Anglo-Italian agreement by which Britain was to recognize the Italian conquests in Ethiopia. In protest Eden resigned the foreign secretaryship, to be succeeded by Lord Halifax, former viceroy of India.

From that time British foreign policy was virtually Chamberlain's own. Meanwhile, throughout 1938, the farce of "nonintervention" in Spain continued, bringing increasing discredit on the democratic powers and enabling Francisco Franco to establish himself by March 1939 as the dictator of Spain.

Czechoslovakia and the Munich Pact.—Hitler's annexation of Austria opened the door to an attack on Czechoslovakia by a combination of diplomatic and military threats, the British and French reaction to which gave to the word appeasement a new and discreditable connotation. At first Hitler confined himself to demanding autonomy for the Sudeten Germans who were incorporated in the Czech republic and the abandonment by the republic of its alliance with Russia. At the same time he expressed his desire for good relations with the British government, which responded by sending Lord Runciman to Prague at the beginning of August in an effort to mediate between the Sudeten Germans and the Czechs. Runciman's mission failed, and by the beginning of September it was evident that Hitler would be content with nothing less than the annexation of the Sudeten territory to the reich, if he would even stop short at that. On Sept. 14, 1938, he seemed to be on the point of invading Czechoslovakia even at the risk of a European war.

Chamberlain then decided to fly to Munich (Sept. 15) and visit Hitler at Berchtesgaden in a last effort to deter him. He returned the following day and, after consultation with the French premier and foreign secretary who came to London on September 18, agreed with them that the Czechs must agree to the secession of the Sudeten provinces and the setting up of cantons in other districts. This they did "under irresistible pressure from Britain and France" and on Sept. 23 Chamberlain flew again to visit Hitler at Godesberg. But this effort also seemed likely to fail because of Hitler's insistence that the Czechs must by Oct. 1 declare their intention of withdrawing their troops from the Sudeten area. Otherwise he would march in.

Chamberlain was in the act of reporting this with its sinister implication to the house of commons (Sept. 28) when there came an invitation from Hitler to meet him, Edouard Daladier, the French premier, and Mussolini at Munich the following day. When this was announced members of all parties rose and cheered, and the public in the gallery was permitted to join in. Chamberlain accepted Hitler's invitation and on the following day (Sept. 30) came back bringing with him an agreement, accepted under pressure by the Czechs, conceding to Hitler his full demands. On his return Chamberlain was welcomed with rapturous enthusiasm. He himself described the accord reached as heralding "peace in our time"; had not Hitler signed a joint declaration that "they were determined to continue their efforts to assure peace in Europe?"

However, a more critical spirit showed itself in the parliamentary debate that followed a few days later. Winston Churchill, whose disagreements first with the national government's India policy and later with its "appeasement" had kept him out of office, denounced the agreements in sombre and powerful terms. Nor did events fulfill Hitler's promise of amity. The internationally supervised cession of Czech territory turned out to be a continuous surrender to every German demand, while Poland and Hungary seized the opportunity to annex additional areas for themselves. Criticism became increasingly vocal. Duff Cooper, the first lord of the admiralty, had given the cue to it by resigning on Oct. 2 on the ground that he disagreed with the government's policy, and Chamberlain was bitterly attacked by the Labour party and the Liberals for having "betrayed and abandoned" Czechoslovakia.

In March of the following year (1939) Hitler showed the worthlessness of the pledges he had given at Munich by marching into Prague and violently suppressing what remained of the Czech republic. At Munich he had led the world to suppose that his ambitions were limited to bringing Germans within the *Reich*, whereas now he was annexing a population of 10,000,000 Czechs. A week later he annexed Memel, after an ultimatum to Lithuania.

Poland.—At the beginning of 1939 Germany had seemed to be on the best terms with Poland. In January Joachim von Ribbentrop, the German foreign secretary, had visited Warsaw and assured both the Poles and the German colony that the German-Polish agreement of 1934 "had put a final end to enmity between the two peoples." Within a week of the occupation of Prague (March 21), however, the German government presented a memorandum to the Polish government demanding the return of Danzig and a railway through the Polish Corridor.

The Poles replied that the Germans appeared to be demanding "unilateral concessions" and that, though they were willing to discuss "objectively" and with the utmost good will any points raised by the German government, peaceful intentions and peaceful methods would be necessary if negotiations were to succeed.

Neither was forthcoming. Hitler reiterated that Danzig must and would become German, and in a violent press campaign charged the Poles with anti-German atrocities. In Danzig itself the Nazi party raised agitation to the boiling point.

It was apparent that the treatment so successfully applied to the Czechs was being tried on the Poles. The pretences of Munich were thrown aside and on March 31, 1939, Chamberlain went to the opposite extreme and gave Poland a guarantee of British and French support "in the event of any action which clearly threatened Polish independence." Similar guarantees were given to Rumania and Greece. Hitler, however, ignored all warnings and continued to refuse the settlement by negotiation which, on the advice of the British and French governments, the Poles continued to offer. There were few days in July and August in which tension seemed to be relaxed, and on Aug. 23, 1939, Sir Nevile Henderson, the British ambassador in Berlin, flew to London at Hitler's request to "put his case" to the British government.

On the same day the world was taken by surprise by the announcement that Germany had concluded a pact of nonaggression with Soviet Russia. Since the beginning of May a British and French mission had been at Moscow endeavouring to negotiate an agreement, and the British government had been much criticized for its failure to make progress. At each step forward the Soviet government had made stipulations which were unacceptable to either Poland or the Baltic states and had held up progress while alternative solutions were sought. It now appeared that it had all the time been engaged in secret negotiations with Germany on a basis which cut the ground from under the feet of France and Britain. For whereas the French and British were seeking Soviet support in the event of a struggle with Germany for the defense of Polish independence, Germany offered the U.S.S.R. a large slice of Poland without having to fight for it, if the U.S.S.R. would stand aside while Germany conquered the country.

In the meantime Lord Halifax suggested various ways in which contact between the German and Polish governments might be restored, but none was acceptable to the Germans. Hitler's "case," as it was presented in the final stormy interview (Aug. 30–31, 1939) which the British ambassador had with him and Ribbentrop, required the presence in Berlin within 24 hours of a Polish plenipotentiary who was to accept or reject a series of propositions which would then be presented to him. When asked what these propositions were, Ribbentrop read them in German "at top speed," and on being asked for a copy of them in writing said that it was now too late since the Polish representative had not arrived in the specified time. Lord Halifax made a final attempt to bring

the Polish government into touch with Germany, and on the evening of Aug. 31 the Polish ambassador called on Ribbentrop and informed him that his government was ready to enter into discussions; but the Germans were on the march and began the invasion of Poland at dawn on Sept. 1. Sir Nevile Henderson was accordingly instructed to ask for his passports "unless the German government suspended all aggressive action against Poland." The British ultimatum expired at 11 A.M. on Sept. 3 and at 11:15 Chamberlain announced over the radio that Britain was at war with Germany.

During a large part of 1939 the European complications had been attended with trouble in the far east, where the Japanese blockade of Tientsin and Japan's generally hostile attitude to the British in Hong Kong and the international settlements threatened dangerous consequences. In view of the situation nearer home the British government was obliged to keep its resentment within bounds. The situation was temporarily eased (until Dec. 1941) when Japan stood aloof from the war and moderated its anti-British policy accordingly.

Home Affairs.—During these years domestic legislation was overshadowed by the international situation and the drain on the national purse in preparation for war. Sir John Simon, who was now chancellor of the exchequer, made what was generally approved as a just distribution between borrowing and taxing for this purpose. In 1937 he raised the standard rate of income tax from 4*s.* 9*d.* to 5*s.* in the pound; in 1938 from 5*s.* to 5*s.* 6*d.*, and in his second war budget of 1939 to 7*s.* in the pound (7*s.* 6*d.* for the following financial year). Death duties were correspondingly increased.

Expenditure on armaments amounted to £750,000,000 in each of these years, of which one-third was found by taxation and two-thirds by loan. The passage of rearmament legislation was the occasion for much parliamentary controversy. The Labour party accepted in principle the implications of collective security in Oct. 1935 when George Lansbury, its pacifist leader, was defeated at a party conference on the sanctions issue, and then resigned and was succeeded by Clement Attlee. Pacifist sympathies, however, remained, reinforced by distrust of the objectives of Baldwin's and Chamberlain's foreign policy; this led the Labour and also the Liberal parties to vote repeatedly against the government's rearmament proposals. Meanwhile on the Conservative benches a small group under Winston Churchill's leadership criticized the government's tardiness, pointing out that it took three years to secure the establishment of a ministry of supply (which was finally set up in April 1939), and that despite considerable financial expenditure Great Britain lost air parity to Germany in 1935 and when war broke out had barely half the number of aircraft possessed by Germany. An Air Raid Precaution act, reflecting official awareness of the changed character of modern war, was passed in 1937, but it was not until 1939 that the more unpalatable implications of the national danger were confronted.

Then in swift succession the establishment of the territorial army was doubled (March) and conscription was introduced (April). On the day after the signing of the Russo-German pact parliament was recalled, and passed within the day by 427 votes to 2 an Emergency Powers bill giving the government wide powers to cope with the imminent crisis of war.

World War II

Outbreak of War.—It was far from immediately apparent how the war would affect the life of the nation. (For the course of the war *see* World War II.) The first widespread expectations—of aerial bombardment and the disruption of the whole fabric of civilian life, both material and moral—were strangely falsified. The false air-raid alarm which sounded in London within a few minutes of the outbreak of war seemed to symbolize the character of the first six months of the struggle—what came to be known as the "phony war" period. The existence of a national register of manpower and a conscription law meant that there was none of the volunteering rush of 1914; instead there was a planned call up which paid due heed to the balance between military and industrial requirements. The popular mood was marked not by the crusading enthusiasm of the earlier struggle but by an almost fatalistic determination to see the grim business through. The war was accepted like the blackout that accompanied it as something hateful and necessary but at the same time less insupportable than had been expected.

On the first day of war Chamberlain set up a war cabinet on the 1916–19 model, consisting of the prime minister, the chancellor of the exchequer, the minister for co-ordination of defense, the lord privy seal, the minister without portfolio (Lord Hankey) and the service ministers. Of these last one was a new and highly popular appointment, that of Churchill as first lord of the admiralty. The public approved this vindication of his Cassandra-like foresight and relied on him to infuse spirit into the administration, a task to which he gladly set himself. New ministries of home security, economic warfare, information, food and shipping were also set up.

The government announced that it was working on the assumption that the war would last three years. This announcement combined with the strategic caution necessitated by the relative weakness of the Allies to induce a certain *festina lente* atmosphere on the home front which, though far removed from the "business as usual" attitude inculcated in Aug. 1914, was still perilously at odds with the real needs of the moment. The Russo-Finnish winter war of 1939–40, driving an ideological wedge into the sympathies of the left and complicating the diplomatic line-up from whatever angle it was viewed, bemused the nation still further. It was fortunate for the government that the reluctance of Norway and Sweden to have their neutrality breached provided an extenuation in diplomatic propriety for Britain's own inability to frame a clear policy.

The German Offensive.—Into this sultry atmosphere the German invasion of Norway in April 1940 broke with something of the shocking and yet freshening effect of a thunderstorm. Chamberlain, who three days before the invasion had talked confidently of Hitler's having "missed the bus," came in for severe criticism. In the commons a censure debate took place on May 7–8. L. S. Amery from the Conservative benches and Herbert Morrison and the aged Lloyd George speaking for the Labour and Liberal members, denounced Chamberlain's leadership.

The government survived the vote but by a majority so shrunken as to imply a moral defeat. It became clear to Chamberlain that an all-party government must be formed and that he would be incapable of forming it. On May 10, 1940, within one day, came the news of his resignation, of the invasion of the Low Countries and of Churchill's formation of a coalition government in which he was both prime minister and minister of defense. Thus, in the moment of the country's direst peril, its greatest war leader assumed the burden of saving it (*see* Churchill, Sir Winston Leonard Spencer). Chamberlain (until his death six months later) remained in office as lord president of the council, and Lord Halifax (until his appointment to the Washington, D.C., embassy in December) remained as foreign secretary. Labour was represented by seven ministers, of whom the most notable were Clement Attlee as lord privy seal and deputy prime minister, Herbert Morrison as minister of supply (later as home secretary) and Ernest Bevin as minister of labour. The Liberal party was represented by Sir Archibald Sinclair. Of the Conservative members the most notable were Anthony Eden at the war office (later to succeed Lord Halifax as foreign secretary) and Lord Beaverbrook as the dynamic minister of aircraft production.

To parliament, from which he asked and obtained a unanimous vote of confidence, Churchill was eloquently frank. "I have nothing to offer," he said, "but blood, toil, tears and sweat." In return he offered the promise of "victory—victory at all costs, victory in spite of all terror, victory, however long and hard the road might be." Then came the miracle of the evacuation of the British army from Dunkerque, but Churchill chose with brutal honesty to stress the fact of defeat, the imminent risk of invasion and the national determination to resist. In his most famous speech he told parliament and the world that Britain would fight on "if necessary for years, if necessary alone. . . . We shall fight on the beaches . . . we shall never surrender." That these were not

CAMERA PRESS LTD.

WINSTON CHURCHILL VIEWING THE COASTAL DEFENSES NEAR DOVER, ENG., SHORTLY AFTER BECOMING PRIME MINISTER; 1940

idle boasts was shown by the almost simultaneous creation of a part-time volunteer force, the home guard, for defense, and of a new army striking force, the commandos, for attack. Pillboxes and anti-tank obstacles sprang up all over the country and the nation, though somehow never fully believing that shores inviolate so long could ever be assaulted, prepared itself for the worst.

To France, collapsing under the German attack, Churchill made the imaginative but abortive offer of an "indissoluble" federal union (June 16, 1940). A day later the Pétain government asked for an armistice and Britain was left alone. It became imperative to prevent the French fleet from falling into German hands and on July 3, 1940, British ships and planes attacked units of the French navy at Oran and Dakar. Only in the escape to Britain of a junior general, Charles de Gaulle, did hope of a resurgent and allied France survive.

The anticipated German cross-channel invasion did not materialize (though on Sept. 7, 1940, a false alarm spread excitement throughout southeastern England). But on July 10 the *Luftwaffe* began its air attack on Britain and for two months the war raged in British skies and rained down on London and coastal towns (*see* BRITAIN, BATTLE OF). Failure to attain their military objectives led Germany to launch its notorious blitz attack on London, which for 57 nights from Sept. 7 to Nov. 3 was subjected to regular bombardment. The capital of the commonwealth became a front-line outpost, and its citizens acquitted themselves with the courage of soldiers amid the distractions of civilian life. The turn of the provincial cities, Coventry, Birmingham, Liverpool, Bristol, Southampton and others, came next. On Dec. 29 an incendiary attack on London turned the City into a lake of flame and marked the climax of the attacks.

In Aug. 1940 the United States agreed to transfer certain destroyers to Great Britain in exchange for the lease of some naval bases; an agreement valuable as much as an earnest of Anglo-American amity as for its immediate military usefulness. How much depended on this amity became clear by the end of the year when Britain's ability to pay for indispensable U.S. supplies and arms began to be severely strained, and only Pres. Franklin D. Roosevelt's concept of lend-lease enabled the country to go on drawing on "the arsenal of democracy." The U-boats threatened transport, and further German blitzes on the seaport towns in March and April 1941 affected delivery of supplies, but the growth of the convoy system and of anti-aircraft defenses kept the chain of communications from breaking until the U.S. action in extending its security zone to longitude 26 caused the curve of shipping losses to fall away once more. Meanwhile victory over Italy's forces in the middle east was interrupted to provide forces for the assistance of hard-pressed Greece. But soon Erwin Rommel's counteroffensive was followed by the evacuation of Greece and Crete.

Suddenly, however, on June 22, 1941, new prospects opened with Hitler's attack upon the U.S.S.R. Churchill lost not a moment in declaring Great Britain's support for the Soviet Union and sending a military mission to Moscow. Supplies followed and continued to flow in mounting quantities along the hazardous sea route to Murmansk and, later, overland through Iran. British sympathy and admiration for the Soviet struggle remained high throughout the war and found official expression in the Anglo-Soviet treaty of May 1942, but no real intimacy was ever achieved on either the military or the official levels; Russian reserve and secretiveness remained impenetrable.

U.S. Entry Into War.—Fortunately this did not hold true of British relations with the United States. With Harry Hopkins as President Roosevelt's intermediary, a meeting was arranged on Aug. 9, 1941, in Placentia bay, Nfd., between Churchill and Roosevelt. From this emerged the declaration of agreed aims known as the Atlantic charter (*q.v.*).

On the following Dec. 7 the Japanese attack on Pearl Harbor fused understanding into unity. Churchill immediately sailed to Washington and amid the news of the initial misfortunes in Cyrenaica in North Africa and in the Philippines and Malaya the foundations of an Anglo-American strategy for victory were laid. Soon a network of Anglo-American combined boards was coordinating the efforts of the two countries throughout the whole range of military and economic operations.

The addition of these allies, the U.S.S.R. and the United States, did not, however, allow Britain to relax any of its effort. In 1941 industrial conscription for men and women was introduced, operated delicately but firmly by the ministry of labour and national service. By the end of 1941 military conscription was even being applied to women between the ages of 20 and 30, while for men it operated between 18 and 61. Rationing of almost all foods became strict, though lend-lease helped to vary an otherwise plain table. In June 1941 clothes rationing was introduced. In March 1942 gasoline rationing followed. The income tax rose to ten shillings on the pound in April 1941, indirect and purchase taxes increased in scale and extent and an excess profits tax of 100% was imposed. The result was a drastic measure of social and economic levelling—the most rapid, pervasive and irreversible in Great Britain's history. At the same time the government's control over every aspect of the national life became more comprehensive, tight and efficient.

Parliament, however, kept a close watch over civil liberties. Though regulation 18b under the Emergency Powers act, introduced in fear of enemy invasion, permitted detention without trial, it was sparingly used; conscientious objectors incurred less odium and were accorded fairer treatment than in 1914–18; press comment was untrammelled, despite a close security censorship; and parliament, though backing the coalition government, could be outspoken in its censures, as in the confidence debate of July 1942, which the prime minister carried by 476 votes to 25.

At home, 1942 was mainly a year devoted to building up Britain's offensive strength. Abroad it was notable for far eastern reverses and the North African campaign. The U.S.S.R., through Vyacheslav Molotov, who visited Great Britain in May, urged the establishment of a second front in Europe. The U.S., initially

of the same mind, after reconsideration came around to the British view that it was not yet practicable. Instead in Nov. 1942 they combined with the British in the attack on French North Africa. The year 1943 saw the success of that operation, the capture of Sicily and the slow advance up the spine of Italy. Meanwhile in Great Britain U.S. troops and supplies accumulated and under the direction of Gen. Dwight D. Eisenhower as supreme commander an alliance in arms of unparalleled intimacy and effectiveness was forged. Simultaneously combined Allied operations had proceeded in the Pacific theatre, though the "grand strategy" of Roosevelt and Churchill agreed in giving priority to the defeat of Germany.

Meanwhile the strain of war and the social pressures produced by it were stimulating the government to new measures of social reform. In Dec. 1942 the government published Sir William Beveridge's proposals for a comprehensive scheme of unemployment and health insurance to apply to everyone irrespective of salary or occupation, and in 1944 the coalition government committed itself to acceptance of its main proposals. In the same year an Education act raised the school-leaving age to 15.

End of the War.—At a meeting at Tehran in Dec. 1943 Joseph Stalin, Roosevelt and Churchill hammered out the final strategy of the war. This gave priority to the opening of a second front in 1944 by landings in Normandy (June 6, 1944). Hitler's threats of retaliation by a secret weapon, however, were not empty boasts; on June 13 the first V-1 missile, a pilotless aerial bomb, dropped on London.

Not until Sept. 4, when the Pas-de-Calais launching installations were overrun by British and Canadian armies, did the bombardment cease, and not before 30,000 persons had been killed and injured and nearly 300,000 homes either destroyed or seriously damaged. But as the V-1s ceased the V-2s, the rocket bombs, began; they persisted until March 28, 1945, and made the total of casualties and damage still higher.

In diplomacy victory brought its problems, as taxing as those of war. The Russian encouragement of a puppet Warsaw government in rivalry with the exiled Polish government in London served early notice of disagreements to come between the eastern and the western powers. In Dec. 1944 Churchill himself flew to Athens to prevent a *coup d'état* by Communist forces in liberated Greece.

The sense of strain thus developing moderated still further the already sober public reaction to the news of victory in Europe. Weariness after a long effort was matched by a determination to prosecute the war in the far east to a successful conclusion. The atom bomb startlingly abbreviated this second round of the struggle but awareness of the terrifying implications of the new weapon depressed minds that would otherwise have been elated at the thought of peace. Other considerations also contributed to a sombre national mood. Great Britain had emerged from the struggle with 360,000 dead, with a huge housing problem in its battered towns and with the resources built up by Victorian thrift utterly, if splendidly, squandered in its fight for survival.

Its export trade was awry, its shipping was cut by a third, its overseas investments had been sold and it was in debt to foreign creditors to the extent of nearly £3,000,000,000. It was in this condition that Great Britain faced a world in which the balance of power had been totally transformed and heavy imperial and diplomatic tasks awaited its attention.

Postwar Labour Governments

As peace approached, the coalition government began to reveal signs of reviving party spirits and although Churchill offered a continuation of the coalition until the end of the war with Japan, Labour preferred an election in the autumn of 1945. Churchill, however, would accept no such middle course; instead he formed an almost entirely Conservative caretaker government and went to the country in July.

The election, fought by the Conservatives largely on the appeal of Churchill's wartime record and by Labour largely on its program of nationalization and reconstruction, was a striking victory for Labour, 393 of the latter's candidates being successful compared with 213 Conservatives and a mere 12 Liberals. Attlee (*see* Attlee, Clement Richard Attlee, 1st Earl) formed Great Britain's first majority Labour government. Herbert Morrison as lord president of the council and leader of the house and Ernest Bevin as foreign secretary formed his closest advisers. Hugh Dalton at the exchequer, Sir Stafford Cripps at the board of trade and Aneurin Bevan at the ministry of health were other prominent men in the cabinet.

Problems of the Peace.—Within less than a month the new government was confronted with Pres. Harry S. Truman's sudden announcement of the cessation of lend-lease. This made more urgent than before the negotiation of a U.S. loan, and in Dec. 1945 Lord Keynes, the principal British negotiator, secured a credit of $3,750,000,000 available up to the end of 1951 at 2% interest. In addition Britain was spared the payment of all but $650,000,000 of the $12,000,000,000 balance due to the United States on the lend-lease account. As part of the agreement Great Britain undertook to revert to multilateral trading as soon as possible and to free convertibility by July 1947. In the following year Canada extended the United Kingdom a further credit of $1,250,000,000.

In contrast with these agreements was a steady deterioration in western relations with the U.S.S.R. It was not until Dec. 1946 that treaties with the Soviet satellite states of eastern Europe were signed—and then only to be broken by the U.S.S.R. in the following year. The joint occupation of Germany and Austria led to frequent disputes among the occupying powers, and all Ernest Bevin's attempts to reach an accommodation with the U.S.S.R. came to nothing. British policy, irrespective of party, certainly favoured reliance wherever possible on the United Nations, which Great Britain had taken a major part in creating. But as it became increasingly apparent that the use of the veto by the U.S.S.R. was crippling the United Nations and that in any case its evolution as an effective agency for enforcing peace depended on its members' willingness to support it, Britain prepared to meet the increasing international tension by rearming; in May 1947 a year's conscription was imposed on all able-bodied men of 18.

Economic Crisis.—Meanwhile the nation's balance of payments position was seriously deteriorating. Though full employment and increased productivity had been maintained, certain vital sectors of British industry failed to meet requirements. The most serious deficiency was in coal, despite the nationalization of the industry effected immediately on Labour's accession to power. In the cold February of 1947 a nation-wide fuel shortage caused industrial stoppages and export losses.

Consequently, Great Britain was obliged to pass to the United States the responsibility for maintaining Greek and Turkish integrity against Soviet pressure, thus precipitating the "Truman doctrine" (*q.v.*) of March 1947. By midsummer it was apparent that the U.S. loan, intended to last until 1949 or 1950, was almost exhausted. As the promised date of convertibility approached, a run on sterling developed; heavy import cuts were imposed and an intensified "production-for-export" drive began. Sir Stafford Cripps succeeded Hugh Dalton at the exchequer and directed and personified an impoverished Britain's dedication to "austerity." But it was only the further blood transfusion of the European Recovery program (Marshall plan), beginning in 1948, which prevented the chronic unbalance in Britain's trade from producing economic collapse. In Sept. 1949 the pound had to be devalued to a level of $2.80, and although this improved the country's reserves of dollars and gold, this stimulus, in turn, began to wane when the Korean War in 1950 sent up the prices of indispensable imports.

These economic difficulties at no time deterred the Labour government from executing its program of domestic reforms. In addition to coal, the Bank of England, civil aviation, the railways and road transport, gas and electricity and the principal units of the iron and steel industry passed from private to public ownership. An extensive national health and insurance scheme, embracing the entire population, was brought into effect in 1948, providing a wide range of free benefits and bringing both the medical profession and all medical facilities under stricter state control. Both rationing and subsidies of essential foods were retained, and sub-

sidized housing was given a very high priority. Taxation, at almost wartime levels, helped in some degree to restrain inflationary pressure.

Foreign and Commonwealth Relations.—Abroad, the Soviet refusal to let any of its satellites participate in the Marshall plan drew more sharply than ever the line between east and west. Britain played a leading part in organizing the European response to the U.S. initiative, through the Organization for European Economic Cooperation. In Jan. 1948 Ernest Bevin (*q.v.*) announced the scheme for uniting Britain, France, Holland, Belgium and Luxembourg in a common defense plan, the Brussels treaty. The Communist coup in Prague and the Berlin blockade made the strengthening of this urgently necessary; to this end Britain provided bases for U.S. atom-bomb-carrying planes and was largely instrumental in expanding the Brussels treaty into the North Atlantic treaty of 1949.

Meanwhile vast changes had been taking place in Commonwealth affairs. The Indian knot was dramatically cut by Lord Mountbatten who, appointed by Attlee as viceroy in Feb. 1947, put into effect a policy of British withdrawal by Aug. 15. The partition of India which ensued was accompanied with bloodshed on a vast scale but produced two new dominions, India and Pakistan, with a continuing amity toward Great Britain displayed by their remaining within the Commonwealth after attaining independence. Burma, emancipated almost simultaneously, left the Commonwealth. Ceylon passed in Feb. 1948 from the status of a crown colony to a dominion and revised constitutions were brought into effect in all the west African colonies, giving greater powers of self-government.

In the Feb. 1950 election Attlee's government was reduced to a majority of only five over all other parties. Nineteen uneasy months followed. In June the invasion of South Korea by Communist forces produced in Britain the same reaction as in the United States—a grim determination to stand by the United Nations and combine to defeat the aggressor, although the adjustment of British and U.S. far eastern policies proved difficult. New problems appeared in 1951—the burden of a £4,700,000,000 rearmament program (denounced as excessive by some Labour members, including Aneurin Bevan, who resigned in protest), a dispute with Iran which involved the loss of the Anglo-Iranian oil installations, and friction with Egypt over the Suez Canal zone and the Sudan. All these problems were still pressing when Attlee went to the country in October.

Conservative Ascendancy

The Return of Churchill.—The Conservatives won a bare majority of 17 over all other parties combined. Churchill formed a government (with Eden at the foreign office and R. A. Butler at the exchequer) which, despite its slender margin of superiority, proved surprisingly tenacious and stable. Its legislative achievements included the denationalization of steel and of road transport and the authorization of commercial television, but its main emphasis was on administrative efficiency and encouragement of private enterprise. Houses were built at the rate of 300,000 a year, food subsidies were cut, rationing controls were reduced, identity cards were abolished and by May 1954 food rationing was ended. Some modest easement was given to the taxpayer by cuts in purchase and income taxes, but public expenditure remained high and inflationary pressure continued.

On Feb. 6, 1952, King George VI died and was succeeded by his 25-year-old daughter as Elizabeth II. The year 1953 saw her coronation and a royal tour around the world.

In the Commonwealth the movement toward self-government proceeded. In the Gold Coast an African became prime minister in March 1952. In the Sudan in Dec. 1953 a mass electorate voted the pro-Egyptian party (National Unionists) into power. Farther south there was established a Central African federation, uniting Nyasaland and Northern and Southern Rhodesia. But in Kenya outbreaks of Mau Mau terrorism in 1952 led to the establishment of a state of emergency and took more than three years to quell, while in British Guiana the new constitution had to be suspended in 1953 in face of a threatened *coup d'état* by extreme left-wing elements. Churchill himself concentrated his efforts on improving the Anglo-American partnership and seeking a settlement "at the summit," as he put it, of differences with the U.S.S.R. Gen. Dwight D. Eisenhower's assumption of the U.S. presidency in 1953 and Stalin's death later in the same year seemed propitious omens. But although Churchill strengthened the Anglo-American alliance at many points, he could not obtain a sharing of America's atomic secrets (Britain exploded, unaided, her first atomic bomb in 1952 and her first hydrogen bomb in 1957) and, at the foreign secretary level, there was a good deal of disharmony between Sir Anthony Eden (*see* Avon, Robert Anthony Eden, Earl of) and John Foster Dulles. This was reflected in the handling of the Indochina crisis of 1953–54, which the Geneva conference provisionally settled in July 1954, and of the European Defense community crisis of the same period which Eden resolved by his pledge to maintain British forces in continental Europe. The long-awaited "summit talks" did not materialize until Churchill, now 81 and in poor health, had retired to make way for Eden in April 1955.

Eden's Government.—Eden, fortified by an increased Conservative majority at the polls (May 1955), represented Great Britain at prolonged talks with Eisenhower, N. S. Khrushchev and Edgar Faure (and their deputies) at Geneva in July 1955 and later in October–November. However the plans which Eden and his foreign secretary, Harold Macmillan, put forward for a graduated German reunification coupled with guarantees for the security of the U.S.S.R., ran aground on solid Soviet resistance to any lifting of the "iron curtain." In 1956 the visit to Britain of Khrushchev and Marshal N. A. Bulganin produced a short-lived easing of British-Soviet tension which, however, Soviet infiltration into the middle east soon renewed.

A continuing British middle east concern was the future of Cyprus, made more crucial by the 1954 agreement with Egypt to close the British base in Suez. Archbishop Makarios led an intensified nationalist agitation against British rule, provoking in Nov. 1955 the imposition of emergency measures and, in March 1956, his own deportation. For general security in the middle east Britain was relying largely on the Baghdad pact set up in 1955 but this was opposed in Jordan, where in March 1956 Gen. J. B. Glubb, the British organizer of the Arab legion, was expelled, and in Egypt where Col. Gamal Abd-al-Nasser had come to power as a nationalist dictator. In July 1956 irritation at the withdrawal of promised Anglo-U.S. aid for the Aswan dam provoked Nasser into a sudden nationalization of the Suez canal. The threat this constituted to vital British and French interests produced strong reactions in London and Paris; the months that followed saw fluctuating reactions in Washington, while two London conferences, an appeal to the United Nations and an abortive canal-users' scheme were all tried, without avail, in hopes of breaking Nasser's hold.

The Suez Conflict.—Soon it became apparent that there was a mounting divergence of purpose between the U.S. and its allies, particularly on the issue of resorting to force, and a widening distrust between Eden and Dulles. Despairing of U.S. support, Britain and France secretly concerted a separate policy, as a result of which, at the end of Oct. 1956, when Israel attacked Egypt, Anglo-French forces landed and seized positions in the canal zone. This severely strained relations with a critical Commonwealth (especially Canada and India) and a shocked U.S., whose condemnatory resolution in the UN Security council was vetoed by Britain and France.

At home, for a few days while the momentum of his action persisted, the prime minister retained the support of the great majority of his party, but from the outset Labour and most uncommitted opinion was hostile. When the international implications of Eden's course became manifest, doubts grew even in Conservative ranks and soon mounting criticism at home and in the Commonwealth and resolute hostility from the U.S. administration forced the government, with its military operation still incomplete, to announce a cease-fire (Nov. 5). On Dec. 22, with the canal blocked to traffic, the last British forces withdrew, to be replaced by a UN emergency and salvage force.

Macmillan's Government.—In Jan. 1957 Sir Anthony Eden's health (which had been precarious for months) obliged him to re-

sign the premiership. He was succeeded by Harold Macmillan (*q.v.*). Macmillan's major task was to repair the ravages of Suez and to rebuild Britain's shattered relations with the U.S., while avoiding the humiliating recantations which would have destroyed the unity of his party. In this he obtained a high degree of success. At meetings with Eisenhower in March and Oct. 1957 the two countries' defense policies in NATO and diplomatic policies in the middle east were harmonized; in Feb. 1958 Britain supplied the U.S. with land sites for its "Thor" missiles and in Oct. 1960 with a naval base at Holy Loch for its "Polaris" submarines. Revision in March 1957 of the repayment provisions of the Anglo-American loan agreements assisted Britain's recovery from the financial strains of Suez, while cuts in defense expenditure based on a reduction in military manpower eased the budgetary position. By 1959 the U.K. was able to relieve restrictions on imports of dollar goods and in 1960 convertibility for sterling was announced.

At home a vigorous anti-inflationary policy was pursued, but a price was paid in a proportionate reduction in the rate at which industrial productivity increased. On the labour front there were initially some signs of unrest; 1957 saw twice as many working days lost through strikes as in any previous year since the war and unemployment reached a postwar peak of almost 621,000 (about 2.5% of the working population) in Jan. 1959. The general standard of living, however, continued to rise and it was on a general appeal of "Life is better with the Conservatives" that Macmillan in 1959 set a new record in British politics by winning for his party its third consecutive general election with a further increased majority of seats. Macmillan's victory, however, might have been due as much to his conduct of foreign as to home affairs, because 1959 saw him as a kind of roving "summit" ambassador, visiting first Moscow and then Paris, Bonn, Ottawa and Washington in endeavours to arrange an east-west settlement. A ten-power disarmament committee and a three-power nuclear test committee were set up, but 1960 failed to produce further progress.

In Europe the establishment of the European Economic Community (the Six) in 1958 presented a challenge to Britain, whose first response was to form the European Free Trade Association (EFTA) in 1960. By 1961 exclusion from the dynamic Six seemed too costly and Macmillan persuaded his party and the commonwealth prime ministers to let Britain negotiate for entry. However, French opposition proved too strong and in Jan. 1963 the negotiations broke down. In the commonwealth the progress toward independence and racial equality continued. South Africa's persistence in apartheid led to its forced withdrawal in 1961, and the "multiracial" Central African Federation was abandoned in 1963. Nigeria, Sierra Leone, Uganda, Kenya, Malta, Tanganyika and Zanzibar (Tanzania), Nyasaland (Malawi) and Northern Rhodesia (Zambia) became independent within the commonwealth.

At home Macmillan's government was harassed by balance of payments difficulties, labour disputes and by-election reverses. He weathered these, and gained prestige through the successful test-ban negotiations with Moscow in Aug. 1963; but illness forced him to retire in October.

Douglas-Home's Government.—Lord Home, previously foreign secretary, succeeded Macmillan, disclaiming his peerages and becoming known as Sir Alec Douglas-Home. He made few government changes and in a year overshadowed by the forthcoming election introduced little controversial legislation beyond a bill to abolish resale price maintenance. In the general election of Oct. 1964 the Conservatives won 304 seats, Labour 317 and the Liberals 9. Douglas-Home resigned and Harold Wilson, the Labour party leader, formed a government.

See COMMONWEALTH OF NATIONS; *see* also references under "English History" in the Index. (H. G. N.)

BIBLIOGRAPHY.—*The Anglo-Saxon Period: Sources:* Bede, *Historia ecclesiastica; Anglo-Saxon Chronicle.* For a translation of and commentary on essential record and narrative material see *English Historical Documents,* vol. i, ed. by D. Whitelock (1955). *General Histories:* R. G. Collingwood and J. N. L. Myres, *Roman Britain and the English Settlements,* 2nd ed. (1937); F. M. Stenton, *Anglo-Saxon England,* 2nd ed. (1947); R. H. Hodgkin, *A History of the Anglo-Saxons,* 3rd ed. (1952); D. Whitelock, *The Beginnings of English Society* (1952); D. M. Stenton, *English Society in the Early Middle Ages,* 2nd ed. (1952); T. D. Kendrick, *A History of the Vikings* (1930). *Special Topics:* W. R. W. Stephens and W. Hunt, *History of the English Church,* vol. i (1899); D. Knowles, *The Monastic Order in England,* 2nd ed. (1949); W. Levison, *England and the Continent in the Eighth Century* (1946); J. Armitage Robinson, *The Times of St. Dunstan* (1923); H. M. Chadwick, *Studies on Anglo-Saxon Institutions* (1905); J. E. A. Jolliffe, *Pre-Feudal England: the Jutes* (1933); F. Pollock and F. W. Maitland, *The History of English Law Before the Time of Edward I* (1898); J. Tait, *The Medieval English Borough* (1936); C. S. and C. S. Orwin, *The Open Fields,* 2nd ed. (1954); H. C. Darby (ed.), *An Historical Geography of England Before A.D. 1800* (1936); G. Barraclough (ed.), *Social Life in Early England* (1960); T. D. Kendrick, *Anglo-Saxon Art to A.D. 900* (1938), *Late Saxon and Viking Art* (1949); C. W. Kennedy, *The Earliest English Poetry* (1943); E. E. Wardale, *Chapters on Old English Literature* (1935).

The Middle Ages: Sources: for a translation of and commentary on important record and narrative material see *English Historical Documents,* vol. ii, ed. by D. C. Douglas and G. W. Greenaway (1953). *General Histories:* A. L. Poole, *From Domesday Book to Magna Carta,* 2nd ed. (1955); F. M. Powicke, *The Thirteenth Century* (1953); M. McKisack, *The Fourteenth Century* (1959); E. F. Jacob, *The Fifteenth Century* (1961); D. M. Stenton, *English Society in the Early Middle Ages,* 2nd ed. (1952); A. R. Myers, *England in the Late Middle Ages* (1952). *See* also the relevant chapters and bibliographies in the *Cambridge Medieval History,* vol. v–viii (1926–36). *Special Topics:* K. Norgate, *England Under the Angevin Kings* (1887); F. M. Powicke, *King Henry III and the Lord Edward* (1947); T. F. Tout, *The Place of the Reign of Edward II in English History,* 2nd ed. (1936); E. Perroy, *The Hundred Years' War,* trans. by W. B. Wells (1951); A. B. Steel, *Richard II* (1941); E. F. Jacob, *Henry V and the Invasion of France* (1947); C. L. Kingsford, *Prejudice and Promise in XVth Century England* (1925); T. F. Tout, *Chapters in the Administrative History of Medieval England* (1920–33); J. F. Baldwin, *The King's Council in England During the Middle Ages* (1913); T. F. T. Plucknett, *Legislation of Edward I* (1949); F. Pollock and F. W. Maitland, *History of English Law Before the Time of Edward I,* 2nd ed. (1898); D. Knowles, *The Monastic Order in England,* 2nd ed. (1949), *The Religious Orders in England,* vol. i, ii (1948–55); Z. N. Brooke, *The English Church and the Papacy* (1931); A. L. Smith, *Church and State in the Middle Ages* (1913); A. H. Thompson, *The English Clergy and Their Organization in the Later Middle Ages* (1947); G. R. Owst, *Preaching in Medieval England* (1926); H. Maynard Smith, *Pre-Reformation England* (1938); E. Lipson, *Introduction to the Economic History of England,* vol. i, 10th ed. (1949); H. S. Bennett, *Life on the English Manor* (1937); E. A. Kosminsky, *Studies in the Agrarian History of England in the Thirteenth Century* (1956); E. E. Power, *The Wool Trade in English Medieval History* (1941); S. L. Thrupp, *The Merchant Class of Medieval London, 1300–1500* (1948); E. E. Power and M. M. Postan (ed.), *Studies in English Trade in the Fifteenth Century* (1933); E. M. Carus-Wilson, *Medieval Merchant Venturers* (1954); M. Postan and E. E. Rich (eds.), *The Cambridge Economic History of Europe,* vol. ii (1952); T. S. Boase, *English Art, 1100–1216* (1953); J. Evans, *English Art, 1307–1461* (1949); A. W. Parry, *Education in the Middle Ages* (1920); W. L. Renwick and H. Orton, *The Beginnings of English Literature to Skelton,* 2nd ed. (1952); H. S. Bennett, *Chaucer and the Fifteenth Century* (1947).

(F. M. S.; D. M. S.; GY. T.)

The Tudors: Sources: Conyers Read, *Bibliography of British History: Tudor Period,* rev. ed. (1959), continued in *Writings on British History,* ed. by A. Taylor Milne (1937–); G. R. Elton, *The Tudor Constitution: Documents and Commentary* (1960); J. R. Tanner, *Tudor Constitutional Documents,* reissue (1951); R. H. Tawney and E. Power, *Tudor Economic Documents,* rev. ed. (1951). *General:* G. R. Elton, *England Under the Tudors* (1955); S. T. Bindoff, *Tudor England* (1950); J. D. Mackie, *The Earlier Tudors, 1485–1558* (1952); J. B. Black, *The Reign of Elizabeth,* 2nd ed. (1959); A. L. Rowse, *The Elizabethan Age,* 2 vol. (1950–55); J. A. Williamson, *The Tudor Age* (1953). *Special Topics:* G. R. Elton, *The Tudor Revolution in Government: Administrative Changes in the Reign of Henry VIII* (1953); A. F. Pollard, *Wolsey,* rev. ed. (1953); J. E. Neale, *Elizabeth I and Her Parliaments,* 2 vol. (1953–58); Conyers Read, *Mr. Secretary Cecil and Queen Elizabeth* (1955), *Lord Burghley and Queen Elizabeth* (1960); T. M. Parker, *The English Reformation to 1558* (1950); P. Hughes, *The Reformation in England,* 3 vol. (1950–54); G. N. Clark, *The Wealth of England* (1946); Sir John Clapham, *Concise Economic History of Britain* (1949); R. H. Tawney, *The Agrarian Problem in the Sixteenth Century* (1912), *Religion and the Rise of Capitalism* (1926), "The Rise of the Gentry," *Economic History Review,* vol. xi (1941); H. R. Trevor-Roper, *The Gentry 1540–1640* (1953).

The Early Stuarts to the Restoration and the Revolution: J. R. Tanner, *Constitutional Documents of the Reign of James I* (1930); S. R. Gardiner, *Constitutional Documents of the Puritan Revolution, 1625–60* (1906); W. C. Costin and J. S. Watson, *The Law and Working of the Constitution,* vol. i (1952); G. N. Clark, *The Seventeenth Century,* 2nd ed. (1947); Christopher Hill, *The Century of Revolution, 1603–1714* (1961); J. P. Kenyon, *The Stuarts* (1958); G. Davies, *The Early Stuarts, 1603–60* (1937); G. N. Clark, *The Later Stuarts, 1660–1714,* 2nd ed. (1955); C. V. Wedgwood, *The Great Rebellion* (1955–); B. H. G. Wormald, *Clarendon: Politics, History and Religion 1640–60* (1951); D. Ogg, *England in the Reign of Charles II,* 2

vol., 2nd ed. (1955–56), *England in the Reigns of James II and William III* (1955); G. M. Trevelyan, *England Under Queen Anne,* 3 vol. (1930–34); Sir Keith Feiling, *A History of the Tory Party, 1640–1714,* reissue (1950). *See* also bibliographies of *Economic and Social Development 1603–1789* below, and CIVIL WAR, ENGLISH. (R. B. WM.)

The 18th Century: S. Pargellis and D. J. Medley, *Bibliography of British History, the Eighteenth Century 1714–1789* (1951); *English Historical Documents,* vol. x, *1714–1783,* ed. by B. D. Horn and Mary Ransome (1957); W. C. Costin and J. S. Watson, *The Law and Working of the Constitution,* 2 vol. (1952); Basil Williams, *The Whig Supremacy, 1714–60* (1939); J. H. Plumb, *The Eighteenth Century* (1950), *The First Four Georges* (1956); J. Steven Watson, *The Reign of George III, 1760–1815* (1960); J. H. Plumb, *Sir Robert Walpole,* 3 vol. (1956–); Sir Lewis Namier, *The Structure of Politics at the Accession of George III,* rev. ed. (1957), *England in the Age of the American Revolution* (1930); H. Butterfield, *George III, Lord North and the People* (1949), *George III and the Historians* (1957); Richard Pares, *George III and the Politicians* (1953); J. Brooke, *The Chatham Administration, 1766–68* (1956); I. R. Christie, *The End of North's Ministry, 1780–82* (1958); Sir Keith Feiling, *The Second Tory Party, 1714–1832,* reissue (1950); J. A. Williamson, *A Short History of British Expansion,* rev. ed., 2 vol. (1943–45); R. L. Schuyler, *The Fall of the Old Colonial System* (1945); V. T. Harlow, *The Founding of the Second British Empire, 1763–93* (1952). (X.)

Economic and Social Development, 1603–1789: Christopher Hill, *The Century of Revolution, 1603–1714* (1961); D. George, *England in Transition* (1953); T. S. Ashton, *An Economic History of England: the Eighteenth Century* (1955); J. H. Clapham, *A Concise Economic History of England* (1949); M. W. Thomas (ed.), *A Survey of English Economic History* (1957); W. Notestein, *The English People on the Eve of Colonization* (1954); E. Lipson, *Economic History of England,* vol. ii and iii, 6th ed. (1956); E. F. Heckscher, *Mercantilism,* rev. ed. by E. F. Söderlund (1955); E. A. Pratt, *A History of Inland Transportation and Communication* (1912); A. S. Turberville (ed.), *Johnson's England,* 2 vol. (1933); P. Mantoux, *The Industrial Revolution in the Eighteenth Century,* with introd. by T. S. Ashton (1961); G. M. Trevelyan, *English Social History* (1944). (D. C. CN.)

The 19th Century: Sources: English Historical Documents vol. xi, *1783–1832,* ed. by A. Aspinall and E. A. Smith (1959), vol. xii, *1833–1874,* ed. by G. M. Young and W. D. Handcock (1956). *General:* E. Halévy, *A History of the English People in the Nineteenth Century,* trans. by E. I. Watkin and D. A. Barker, 6 vol. 2nd ed. (1949–52); J. Steven Watson, *The Reign of George III, 1760–1815* (1960); E. L. Woodward, *The Age of Reform, 1815–70* (1938); R. C. K. Ensor, *England 1870–1914* (1936); Asa Briggs, *The Age of Improvement* (1959); G. M. Young (ed.). *Early Victorian England,* 2 vol. (1934); W. E. Houghton, *The Victorian Frame of Mind* (1957). *Special Topics:* S. Maccoby, *English Radicalism, 1832–1852* (1935), *1853–1886* (1938), *1886–1914* (1953); N. Gash, *Politics in the Age of Peel* (1953); H. J. Hanham, *Elections and Party Management in the Age of Disraeli and Gladstone* (1959); G. D. H. Cole, *A Short History of the British Working-Class Movement, 1789–1947,* rev. ed. (1948); J. L. Hammond, *Gladstone and the Irish Nation* (1938); R. W. Seton-Watson, *Britain in Europe, 1789–1914* (1937); H. Temperley and L. M. Penson, *Foundations of British Foreign Policy from Pitt to Salisbury* (1938); P. Knaplund, *The British Empire 1815–1939* (1941); J. H. Clapham, *Economic History of Modern Britain,* 3 vol. (1926–38); J. L. and B. Hammond, *The Age of the Chartists* (1930); H. W. C. Davis, *The Age of Grey and Peel* (1929); Asa Briggs, *Victorian People* (1954); H. M. Lynd, *England in the Eighteen-Eighties* (1945). *See* also W. F. Monypenny and G. E. Buckle, *Life of Disraeli,* 6 vol. (1910–20); J. Morley, *Life of William Ewart Gladstone,* 3 vol. (1903); H. C. F. Bell, *Lord Palmerston,* 2 vol. (1936); J. L. Garvin and Julian Amery, *Life of Joseph Chamberlain,* 4 vol. (1932–51); N. Gash, *Mr. Secretary Peel* (1961). (A. BRL.)

The 20th Century: The period dating from 1905 still awaits proper historical treatment and information is often only obtainable by search among the source materials themselves. Of these the most important are the wide range of government and parliamentary papers published by H.M. Stationery office, and the files of the leading national newspapers. The volumes of the *Annual Register* provide convenient digests of the public events of each year. *See* also the wide range of biographies and memoirs published for this period. Of secondary sources, the best general surveys are: R. C. K. Ensor, *England,* 1870–1914 (1936); J. A. Spender, *Great Britain, Empire and Commonwealth, 1880–1935* (1936); C. L. Mowat, *Britain Between the Wars, 1918–1940* (1955). Much of the period is reported upon by an eminent participant in W. S. Churchill, *The World Crisis,* 6 vol. (1923–31), and *The Second World War,* 6 vol. (1948–54). (H. G. N.)

ENGLISH LANGUAGE. The English language is characterized by the extremely mixed nature of its vocabulary, by a simplicity of inflection which has led to the development of a relatively fixed word-order, and by the extension of its system of intonation to express shades of meaning formerly otherwise expressed. No one of these is peculiar to English, unless perhaps the extreme receptiveness of its vocabulary to borrowings from other languages, but taken together they do to some extent characterize it. Certainly there can be little doubt of the outstanding importance of English among the languages of the world. It is not the one used by the greatest number of native speakers, an honour that belongs to the national language of China, a standard language based primarily on Mandarin, and spoken by about 650,000,000 people. But English has more than 250,000,000 native speakers and is far more widespread than Chinese. It is spoken throughout the continent of North America, in the British Isles, South Africa, Australia and New Zealand, and is also the most important of the languages of colonization in Asia, Africa and Oceania. In addition it is the language of commerce, and the second language of many countries which formerly had French or German in that position.

The language has changed considerably during the 1,500 years of its existence. As in all living languages change has been gradual but continuous, and the present language is the result of centuries of development. This is not surprising. Speech is the result of a series of bodily movements, and any such series is naturally subject to change when constantly repeated at different times and by different people. Any change in the action of the vocal organs will result in a difference of sound, with the result that each individual is constantly, if unconsciously, varying from the norm. Under such circumstances there can be no such thing as strict uniformity in speech. Not only will the speech of one community differ from that of another, but within the same community the speech of individuals will differ. Such individual variations within the community will nevertheless necessarily be slight because of the restrictive influence of social intercourse, and in any given community at any particular time there will be a general similarity in the speech. Since any living language is merely the sum total of a number of speech communities, it will naturally change as the language of the individual speakers changes.

But although change is inherent in a living language, it does not conform to a regular and uniform pattern. Changes which in one community find echoes among a majority of speakers, and so come to be accepted in the language, may in another be in a minority and fail to establish themselves. If these communities for any reason come to be separated and remain apart, the conditions will then be present for the differentiation of the originally homogeneous language into dialects. In time, if there is little or no social intercourse between the two communities, the dialects may develop into different languages, since the difference between dialect and language is merely one of degree of differentiation from the original speech. Just as all living languages consist of a number of different dialects, so the languages themselves are but the result of centuries of development, and were themselves at an earlier period merely dialects.

Origins.—Most of the languages of Europe, and many of those of Asia, are descended in this way from a language spoken about 5,000 years ago, perhaps on the steppes of south Russia, or perhaps in no definable geographical region but by a Neolithic people wandering between the Rhine and the Aral sea. This language, which can to some extent be reconstructed from the extant records of the descended languages, is variously known as Indo-European, Indo-Germanic and Aryan, of which the first is probably the best. It should be noted that when using the term Indo-European it means simply the people who spoke that particular language and indicates nothing whatever about their race, about which in fact nothing is known, not even whether it was a pure race or a mixture of races.

Before recorded history this Indo-European-speaking people had dispersed, and their language had been differentiated into dialects. About the beginning of the Christian era one of these dialects, known as Primitive Germanic, or Teutonic, in its turn began to split up into dialects, and it is from these that the modern Germanic languages are descended. The conventional names given to these dialects of Germanic are East Germanic, from which are descended the extinct Gothic, Vandal and Burgundian; North Germanic, later differentiated into East Norse, whence Danish and Swedish, and West Norse, whence Norwegian and Icelandic; and West Germanic. This last is usually divided into High German (whence modern German and the extinct Lombardic) and Low

German by the effects of a sound-change which took place in the mountainous district to the south, but not in the low-lying plains to the north. From Low German are descended the *Platt-Deutsch* dialects of the North sea coast, Dutch, Frisian and English, the last two showing sufficiently close similarities to suggest the possible existence of an earlier Anglo-Frisian dialect. However, work on the subject in the middle of the 20th century has suggested that this traditional view of the development of the Primitive Germanic dialects may be somewhat misleading, and a closer relationship may have existed between North Germanic and the Low German dialects than would be indicated by it.

A Germanic dialect, known variously as Old English or Anglo-Saxon, was introduced into Britain by the Anglo-Saxon conquest, and the history of the language after that time is usually divided chronologically into three main periods of development—Old English, from the time of the earliest records to 1150; Middle English, from 1150 to 1500; and Modern English, from 1500 onward. The dates are necessarily only approximate, since it is impossible to mark off with any precision epochs in a process of change and growth which is continuous. For further convenience the periods are sometimes further subdivided; Old English into Primitive Old English—before the time of the earliest written records—and into Early and Late Old English (before and after 900), and the Middle and Modern periods are also often subdivided into Early and Late, though such subdivisions should not, perhaps, be taken too seriously. (*See* GERMANIC LANGUAGES.)

OLD ENGLISH

The Anglo-Saxon conquest is traditionally accepted as beginning in 449, the date given by Bede for the landing of Hengist and Horsa, but raids on the coasts had been in progress for a considerable time before that, and small settlements of Germanic-speaking peoples were probably already in existence. Exactly what language the invaders found in this country is uncertain. By the middle of the century Britain had been cut off from the rest of the Roman empire for at least a generation, and although some Latin speakers might still have been found in the larger towns, it seems probable that Celtic was normally the language with which the invaders came into contact.

The earliest written records of Old English date from the closing years of the 7th century and the first few years of the 8th, and consist in the main of charters written in Latin but containing personal and place names in English. The Moore and Leningrad manuscripts of Bede's *Ecclesiastical History of the English Nation* (*c.* 737 and *c.* 746) preserve not only very early forms of English names but also a few lines of verse in the Northumbrian dialect, said to be the original version of Caedmon's "Hymn on the Creation." Also from the 8th century are some inscriptions, along with some glosses and glossaries. But the great bulk of documents in the oldest form of English does not begin before the 9th century. However, by the time of the earliest written records the language is already differentiated into at least four main dialectal types—Northumbrian; Mercian; Kentish; West Saxon.

Northumbrian and Mercian, which have much in common, are often classed together as Anglian. Northumbrian was spoken north of the Humber, and at its greatest extent the kingdom of Northumbria included much of the Lowlands of Scotland. Mercian was spoken in the greater part of the area between the Humber and Thames rivers. From the Mercians a group of Angles who settled Norfolk and Suffolk are definitely distinguished by name as East Angles. Unfortunately their dialect is hardly known before the Norman Conquest, but its Middle English descendant shows marked differences from the Mercian type. Kentish was spoken over a larger area than the modern county of Kent, and although Saxon dialects were presumably spoken in Sussex, Middlesex and Essex, as well as in Wessex, West Saxon is the only one of which much is known.

How far back this division goes, and to what extent it reflects differences of origin among the invaders is a difficult question. According to Bede, "Those that came from beyond the seas were from three of the most valiant nations in Germany: that is, the Saxons, the Angles, and the Jutes. From the Jutes came the people of Kent and of the Isle of Wight, and they which in the province of the West Saxons are called unto this day 'the nation of the Jutes,' right over against the Isle of Wight. From the Saxons, that is from that region which is now called that of the Old Saxons, descended the East Saxons, the South Saxons, and the West Saxons. From the Angles, that is from that country which is called Angul, and from that time to this is thought to stand deserted in the midst between the lands of the Jutes and the Saxons, descended the East Angles, the Middle Angles, the Mercians, and all the progeny of the Northumbrians, together with the other people of the Angles."

This would seem definite enough, but Bede was writing nearly 300 years after the events which he describes, and there are difficulties in the account itself. Moreover it seems clear enough that in their nomenclature the invaders themselves knew nothing of such clear-cut divisions. Elsewhere Bede uses the phrase *Angli sive Saxones* (Angles or Saxons)—implying that both terms mean the same thing—though he usually refers to the people of the country generally as *Angli*, and to the language as *sermo Anglicus*. Similarly Alfred invariably refers to himself as king of the West Saxons, but speaks of his language as *Englisc*. Nor is such a threefold division supported by archaeology or social history. Archaeologically it seems possible to distinguish the Jutes from the other invaders, and the social systems give a similar result, but there seems no reason for assuming any difference of race between so-called Angles and so-called Saxons, and there is nothing to show that, at the time of the migrations, there was any difference in language between the two peoples. Even in the case of the Jutes, who were almost certainly of a different origin, it is doubtful whether at this period the various branches of West Germanic had diverged to any great extent. It is far more likely that the dialects which later appeared had sprung up on English soil and were occasioned more by geography than by origin.

When literature and learning first emerged in England, Northumbria was the cultural centre of the country, and it was there that a vernacular literature first developed. However, culture and learning in Northumbria were completely destroyed by the Viking invasions of the 9th century; Wessex alone preserved its independence, and only there could the native literary traditions survive. Later writers had to rely on West Saxon models, so that that dialect became the literary language of the country. Moreover, during the 10th century, the Danelaw was gradually reconquered by the kings of Wessex, and West Saxon became the official and legal language of the country. Consequently most of the Old English works still surviving are in West Saxon, which, by the late Old English period, can be regarded as the standard literary language of the whole country.

Orthography.—The Anglo-Saxons, together with the other Germanic peoples, originally used the runic alphabet. (*See* RUNE.) This alphabet persisted in use, especially in short inscriptions, throughout the Old English period, but for ordinary purposes it gave place after the conversion to Christianity to a form of handwriting which was an adaptation of the Roman alphabet in use among the Irish, and which was passed on to the English by Christian missionaries from that nation. This alphabet was made more suitable for expressing the sounds of Old English by the introduction of certain symbols from the runic alphabet, *þ* (thorn) for *th* and *ƿ* (wynn) for *w*, for which *th* and *uu*, or *u*, were written in the earliest records. In addition the former sound is often represented by *ð* (eth), while the vowel sound approximating to the *a* in Eng. *cat* is represented by *æ* (ash), Old English *a* itself representing the sound of the vowel in German *Mann*. Marks of quantity are used in many manuscripts, though by no means consistently, and sometimes a vowel is doubled to express length. Doubling of consonants is consistently carried out to express a long or double consonant. In modern editions of Old English texts only *þ*, *ð*, *æ*, are retained; *ƿ* is usually represented by *w*, and length marks over the vowels are consistently given or omitted. Otherwise modern forms of the letters are used.

In the main the spelling of Old English adequately represented the pronunciation. It possessed the vowels *a*, *e*, *i*, *o*, *u*, *æ*, *y*, all both long and short, the last being the high front rounded vowel

the short variety of which is that represented by *u* in French *lune*. In unaccented syllables the vowels still retain something of their original quality, not yet having been uniformly weakened to an indeterminate sound. The diphthongs, *ea, eo, io, ie*, long and short, were falling diphthongs (*i.e.*, with the stress on the first element) and were in the main pronounced as spelled. The same is true of the consonants. Only in the cases of *c, g, h*, is there difficulty. Originally all three had been back consonants, but in Primitive Old English they became fronted before the front vowels that then existed, remaining as back consonants before the original back vowels; even if these back vowels were later fronted (*e.g.*, as in *Cent, gylden*) the consonants retained their back pronunciation. Hence *c* represents the [k] sound initially before a back vowel, finally, and medially between back vowels, as in *cuman, bæc, wacor*. But initially before an original front vowel, medially between front vowels, and when doubled, it approximated to the sound usually represented in modern English by the *ch* of *choose*, as in *cēosan, rice, þæcc*, etc.; *sc* in *scip, wascan, fisc*, etc., was always the sound represented by *sh* in modern English. When initial before back vowels *g* was the voiced back stop, as in *gān, gōd*, but medially between back vowels it was the voiced back open consonant—a sound which no longer exists in English—as in *lagu, sagu;* before an original front vowel it approximated to the sound represented in modern English by the *y* of *yell*, and similarly when medial between front vowels, or when final after a front vowel (*e.g., gieldan, fæger, dæg*, etc.). *cg* was the sound represented in modern English by the *dg* of *edge*, as in *licgan*. Initially *h* was always the aspirate, but medially or finally after a back sound it was the back voiceless open consonant (the sound represented by *ch* in Scottish *loch*), and after a front sound the front voiceless open consonant—the sound of *ch* in Scottish *nicht*. These were the chief inconsistencies which prevented Old English spelling from being strictly phonetic. Otherwise the only difficulty lies in the fact that, apart from *h*/*g*, no distinction is made between the spelling of the voiceless and the voiced open consonants. Normally *f, s, þ*, (*ð*), represented the voiceless sounds, but between voiced sounds, unless doubled, they too became voiced.

Accidence.—Old English was an inflected language—though less so than Primitive Germanic and very much less than Indo-European—in which the relation of the words in the sentence was indicated mainly by the use of case endings. Grammatical gender was preserved, and word order was comparatively free so that subject and predicate could and frequently did change places. In its grammar Old English resembles modern German. The noun and adjective are inflected for four cases in the singular and four in the plural (nominative, accusative, genitive, dative), though the forms are not always distinctive. In addition, the adjective has separate forms for each of three genders. A former ablative, and usually a locative or instrumental, have merged with the dative, and in the same way the vocative is normally identical with the nominative. Although Old English retained a fairly elaborate system of declensions, three-fourths of the nouns followed three main strong declensions according to gender, while the so-called weak declension accounted for most of the remaining ones. Following are the four main declensions of nouns:

		Strong			*Weak*
		masc.	neut.	fem.	
Sing.	nom.	*stān*, stone	*scip*, word	*giefu*, gift	*guma*, man
	acc.	*stān*	*scip*	*giefe*	*guman* (m.)
	gen.	*stānes*	*scipes*	*giefe*	*guman*
	dat.	*stāne*	*scipe*	*giefe*	*guman*
Plur.	nom.	*stānas*	*scipu*	*giefa*	*guman*
	acc.	*stānas*	*scipu*	*giefa*	*guman*
	gen.	*stāna*	*scipa*	*giefa*	*gumena*
	dat.	*stānum*	*scipum*	*giefum*	*gumum*

In addition five of the nouns of relationship, *fæder, mōder, brōþor, sweostor, dohtor*, had a separate declension to themselves, and the 7 mutated plurals of modern English had 25 counterparts in Old English.

The adjective, with rare exceptions, had two forms, so-called strong and weak, according to function and position. The weak declension was normally used after the definite article and after the demonstrative and possessive pronouns, and had the endings of the weak nouns. The strong declension contained a mixture of nominal and pronominal endings. The comparative of adjectives usually ended in *-ra*, the superlative in *-ost, -est*. Like modern German, Old English possessed a fully inflected definite article, which was really a demonstrative pronoun, its pronominal character appearing in its not infrequent use as a relative. The personal pronouns of the first and second persons and of the third person feminine singular (*ic; þū; hēo*) agree with those of other Germanic languages, but the forms of the third person masculine and neuter (*hē; hit*) differ from those of Old High German, Gothic and Old Norse but agree with the Old Frisian and Old Saxon forms. In addition to the ordinary two numbers, singular and plural, Old English had forms of the first and second personal pronouns for use in referring to two persons or things—the dual. But already in Old English these dual forms were falling into disuse, though they survived sporadically until the 13th century.

In the verb Old English distinguished only two simple tenses by inflection, a present and a past. It recognized the indicative, subjunctive and imperative moods, and had the usual two numbers and three persons, though in the plural one form sufficed for all three persons. A distinctive feature of the Germanic languages was the division of the verb into two main conjugations, strong and weak. The former indicated a change of tense by a modification of the root vowel (as in mod. Eng. *sing, sang, sung*), while the latter indicated the change by the addition of a suffix containing *-d-* or *-t-* (as in mod. Eng. *seem, seemed, seemed*). The strong verbs are usually grouped into seven different classes according to the variation of vowel in four principal parts, the infinitive, past tense singular and plural, and past participle. The weak verbs are normally divided into three classes, forming their past tenses by the addition of *-ede, -ode, -de, -te*, to the stem, and the past participle by the addition of *-ed, -od, -d*. It should be noted that already in Old English the weak conjugation had become the dominant one. Some of the strong verbs had already passed over to it, and practically all new verbs added to the language were inflected in accordance with the weak conjugation. Old English agrees with other German languages in having certain anomalous auxiliary and preterite-present verbs, as *wesan* "to be," *sculan* "shall," *witan* "to know," *magan* "to be able," etc.

Vocabulary.—The greatest number of loan-words in Old English comes from Latin, and these can be divided into those borrowed while the pre-English tribes were still on the continent, and those borrowed in England. To these may perhaps be added a third class consisting of words borrowed in England from Latin-speaking Celts.

The first class tends to appear also in other Germanic languages and, having been borrowed on the continent, naturally undergoes the Primitive O.E. sound changes. At other times evidence of date may be furnished by sound changes which took place in Vulgar Latin, *e.g.*, an intervocalic *p* became *v* in parts of Gaul during the 7th century, so the fact that Lat. *cuprum*/*coprum* appears as O.E. *copor* indicates a borrowing before this change took place. Many of these very early loans have long since become an integral part of the language, and it is often difficult to realize that they are not native. The chief occupations of the Germans in the later Roman empire, agriculture and war, are reflected by such words as *mīl* "millet," *næp* "turnip," *peose* "pea," *culter* "coulter," *camp* "battle," *segn* "banner," *weall* "wall," *trifot* "tribute," etc. Even more numerous are the words connected with trade; *cēap* "bargain" and *mangian* "to trade," along with their derivatives, are fundamental, and here also belong such words as *pund* "pound," *sēam* "burden," *mynet* "money," etc. One of the most important branches of trade with the Germanic peoples was the wine trade, hence such words as *wīn* "wine," *must* "new wine," *flasce* "bottle," etc. A number of words refer to domestic life, *cytel* "kettle," *disc* "dish," *belt, cuppe* "cup," *cycene* "kitchen," etc. Latin words for certain foods were adopted, *cȳse* "cheese," *pipor* "pepper," *butere, plūme* "plum," etc. Roman contributions to building are evidenced by *cealc* "chalk," *pic* "pitch," *tigele* "tile," etc., and in addition there are such mis-

cellaneous loans as *mūl* "mule," *draca* "dragon," *cāsere* "emperor," *Sæternesdæg* "Saturday."

The Latin loans borrowed through the Celtic inhabitants of this country are unimportant. Apart from a few elements found in place names, *e.g.*, *ceaster, port, munt, torr,* hardly any Latin words can be proved to owe their presence in the language to the Roman occupation, though it is possible that some of the Latin words acquired on the continent were reinforced by the presence of the same words in Celtic.

The third type of Latin loans, those borrowed in England, fall into two classes; those due to the introduction of Christianity in the 7th century, and later loan-words due to the monastic reforms of the 10th century and the consequent renewed literary activity. Most of the first type were needed to express the new ideas and sentiments due to the conversion, but the church also exercised influence on domestic life and education, and what medical knowledge there was came mainly from Latin sources. Consequently, in addition to such purely religious words as *abbud, mynster, dēacon, munuc* "monk," *bisceop, prēost*, etc., there are also such words as *cæppe* "hood," *purpur* "purple," *cyst* "chest," *sæcc* "sack," *scōl* "school," *fers* "verse," *mēter* "metre," *pīn* "pine," *alewe* "aloes," *hysope*, etc. In expressing these new ideas and sentiments the Anglo-Saxons did not entirely neglect the resources of their own language. Native suffixes were added to the foreign words, *e.g.*, *prēosthād, bisceoprīce*, etc.; words used formerly of the old pagan religion might be modified to fit Christian ideas, *e.g.*, *hūsl* "eucharist," *wēofod* "altar," *Ēastron* "Easter," etc., and not infrequently Latin words were translated literally into Old English, *e.g.*, *handbōc* (*manualis*), *þrȳnes* (*trinitas*), *bōcere* (*scriba*), etc. The second class of Latin loans differs from the earlier Christian borrowings in being of a less popular kind, and in often expressing ideas of a scientific and learned character. As with the earlier class a considerable number had to do with religious matters, *e.g.*, *alb, apostol, cantere, dirige* "dirge," but words relating to everyday life characteristic of the earlier period were less numerous. Literary and learned words predominated, *e.g.*, *accent, punct* "point," *organ* "song," *tītul*, and many others which were never assimilated. A few words appear to have been borrowed directly into West Germanic from Greek, but the only certain example in Old English is *cirice* "church," though *engel* and *dēofol* may belong there, but could have come through Latin.

The conquest of a Celtic population by the Anglo-Saxons might have been expected to result in numerous Celtic loans into Old English. The old idea of an almost complete extermination of the Celtic inhabitants may come very near the truth in the districts first conquered, but it is certain that later there was a considerable contact, not always hostile, between the two peoples. Such Celtic influence is mainly perceptible in place names. The O.E. kingdoms of Kent, Deira and Bernicia are derived from Celtic tribal names; Cumberland is the "land of the Cymry," London and York are Celtic, while the first element of Winchester, Salisbury, Exeter, Gloucester, Worcester, Lichfield and others is traceable to a Celtic source. But it is in the names of rivers and hills that the greatest number of Celtic words survives. Thames is a Celtic river name, and various Celtic words for "river, water" are preserved in the names Avon, Exe, Usk, Esk, Ouse, Stour, Wye, etc. Celtic words meaning "hill" are found in place names such as Bar, Bredon and Pendle, while other Celtic elements such as *cumb* "a deep valley," *pyll* "a tidal creek," *brocc* "badger," occur more or less frequently. In other than place names, however, the influence of Celtic upon the O.E. vocabulary is almost negligible. Less than a score of words can be traced with reasonable probability to a Celtic source. These include popular words such as *binn* "manger," *bratt* "cloak," *bannuc* "piece of bread," *luh* "pond," *dunn* "gray," along with learned words such as *drȳ* "magician" (*cf.*, *druid*, a later borrowing of the same word), *cine* "sheet of parchment," *cros, clucge* "bell," and perhaps also *āncor* "hermit," *stær* "history" and *cursian* "to curse." Few of these Celtic words, whether learned or popular, attained a permanent place in the O.E. vocabulary.

The next important influence on the vocabulary was that of the Scandinavian invaders who settled in England during the 9th, 10th and 11th centuries, though the effects are not particularly noticeable before the Middle English period. The Norse settlements had been completed before the Norman Conquest, but the areas where most of the invaders were situated produced no literature that survived before Middle English times. Most of the settlers appear to have been Danes, especially in the north and the east midlands, but settlements of Norwegians were to be found along the west coast, especially in Cumberland and Westmorland. However at this period East and West Norse have little to distinguish them, and it is only rarely that it can be said definitely that a loan-word is distinctively Danish or Norwegian. The similarity between Old English and Old Norse makes it at times difficult to decide whether a given word in Middle or modern English is native or borrowed. Many of the commoner words in the two languages were identical, and if there were no O.E. literature surviving from the period before the Viking invasions it would often be impossible to say whether a particular word was of Scandinavian origin or not. Hence such words as *man, wife, house, thing, come, see, stand, wit, wise, over, under*, could equally well be derived from Old Norse or Old English, though they are in fact English. In certain cases, however, there are fairly reliable criteria by which a borrowed word can be recognized, the most certain depending on the differences in development of certain sounds in North and West Germanic. The simplest is probably the development of *sk* which in Old English had early been fronted to the *sh* sound, whereas the original sound remained in North Germanic. Thus words with *sk* in Middle English are Scandinavian loans, *e.g.*, *sky, skill, skin, scrape, bask*. In the same way the retention of back *c*/*g* before a front vowel, as in *kid, dike, get, gild, egg*, is an indication of O.N. origin, though this test is not so certain as the preceding one since the Anglian dialects in certain cases may have back *c*/*g* corresponding with front *c*/*g* in West Saxon. More rarely the vowel of the word may prove it to be a loan. For example Germanic *ai* gave *ā* in Old English but *ei* in Old Norse; hence such words as *aye, nay, swain*, are clearly from Old Norse. Such tests as these, based on the different sound developments in the two languages, are the most reliable means of distinguishing Scandinavian from native words, though occasionally the meaning may give some hint. Thus, so far as the form is concerned, mod. Eng. *bloom* could be from either O.E. *blōma* or O.N. *blóm*. But the O.E. word meant "lump of metal," whereas the O.N. one meant "flower, bloom." The O.E. meaning has survived as a term in metallurgy, but it is the O.N. sense that has come down in ordinary use. When neither the form of the word, nor its meaning, proves its O.N. origin, it can never be certain that the word is a Scandinavian loan.

Comparatively few O.N. words appear in Old English, in all about 40. The largest single group is composed of words dealing with battle or the sea, as might be expected from the character of the relationship between the two peoples. Hence *barda* "beaked ship," *cnearr* "small warship," *flēge* "ship," *scegþ* "vessel," *lið* "fleet," *hā* "rowlock," *dreng* "warrior," *hofding* "chieftain," *orrest* "battle," *fylcian* "to marshal troops," etc. A little later a number of words dealing with law, or characteristic of the social and administrative system of the Danelaw, appear, *e.g.*, *lagu* "law," *māl* "lawsuit," *hold* "freeholder," *wæpentæc* "wapentake," *hūsbonda* "householder," *prǽll, wrang*, etc. But it is not until the M.E. period that there are any considerable additions to the English vocabulary from Old Norse, and if these are examined in an attempt to discover in what respects the Danes more especially contributed to English culture, no significant result is obtained. Their civilization was very much like that of the English, and so the Scandinavian elements that entered the language were such as would make their way into it through everyday intercourse. This is the great difference between the earlier borrowings from Latin or Celtic and the later O.N. element. In the former cases the borrowings had been necessary to supply words for new ideas for which no term existed in the vocabulary. In the latter, new words were introduced for ideas which were already, in the main, adequately expressed in the native language.

The O.N. and O.E. words were used side by side, and the sur-

vival of one or the other must often have been a matter of chance. Where there were differences in form between the two words it was often the O.E. word that survived, *e.g., bench (benk), goat (geit), gray (graw), leap (loup)*, though the corresponding O.N. forms are common in Middle English and in some cases still exist in the modern dialects. Similarly where the same idea was expressed by different words it was often the O.E. form that survived, *e.g., think (attlen), swell (bolnen), lose (tinen), mist (roke), path (reike)*. In other cases the O.N. word replaced the native word, often after the two had long been in use concurrently. Hence Scand. *sky* replaced O.E. *lyft*, and *die* replaced O.E. *steorfan/sweltan* (though *steorfan* survived in a specialized sense as the modern *starve*), the new verb being more easily associated with *dead* and *death*. Similarly with *take (niman), fellow (gefēra), leg (scanca), low (niþer), egg (ǣg), awe (ege)*, etc. Occasionally both the O.E. and the O.N. words were retained, usually with a slight difference of meaning or use, *e.g., rear/raise, from/fro, craft/skill, hide/skin, sick/ill, shirt/skirt*. Or the Scand. word might impart its meaning to the English cognate; O.E. *brēad* meant "crumb, morsel," while the sense in modern English is due to O.N. *brauð*, O.E. *plōh* was a measure of land, but O.N. *plógr* was the implement, a meaning transferred to the modern *plow* which has dispossessed O.E. *sulh*. O.E. *eorl* "a man of noble birth, warrior" gradually acquired the sense of Scand. *jarl*, a title applied to the governor of a province, and later still became the name of a grade in the peerage.

The English borrowed from the Danish settlers some names of territorial divisions and kinds of jurisdiction. Thus North, East and West Riding (Yorkshire) from O.N. *þriðjungr* "third part"; *by-(law)*, and the same element in Derby and other place names, from O.N. *býr* "town, village." All this merely goes to show that in the O.N. influence on Old English there is an intimate mingling of the two languages, further evidence for which is shown by the fact that the loan-words were not confined to nouns, adjectives and verbs, but extended to pronouns, prepositions, adverbs, and even part of the verb *to be*. It is only rarely that such parts of speech are transferred from one language to another. The pronouns *they, their, them*, came from Old Norse, and their later extension throughout the country was probably because they were less subject to confusion than the native forms. O.N. prepositions are *til* and *fro*, and from the same source comes the conjunction *though*. The O.N. use of *at* as a sign of the infinitive is still to be seen in *ado (at do)*; it was widely used in Middle English, and is still found in some dialects. As adverbs such forms as *aloft, athwart* and *seemly* still survive.

Moreover, O.N. influence affected not only the vocabulary but extended to matters of grammar and syntax. A certain number of inflectional elements peculiar to Northumbrian have been attributed to O.N. influence, among others the *-s* ending of the third person singular present indicative, and the *-and* ending of the present participle. The words *scant, want, athwart*, preserve in their final *-t* the O.N. neuter adjective ending. More important is the fact that in many words Old English and Old Norse differed chiefly in their inflectional endings. The stem of the word was the same in both languages, so that the inflectional endings would be the main obstacles to mutual understanding. In the mixed population of the Danelaw such endings would tend gradually to become obscured and then lost. It is probable that the tendency toward the loss of inflections, inherent in English, was strengthened and accelerated by the conditions prevailing in the Danelaw.

The only other point to be noticed in connection with Old English is the existence of a highly developed poetic diction, characterized by a wealth of synonyms. Not only is there a large number of words whose use is confined to poetry, but also a most elaborate system of metaphorical expressions and round-about ways of referring to such things as the sea, the sun, ships, battle, a prince, a sword, etc.

These expressions are known as kennings and include such words and phrases as *wælwulf* "wolf of slaughter," *i.e.,* "warrior," *sæhengest* "sea horse," *i.e.,* "ship," *æsc-plega* "sport of spears," *i.e.,* "battle," *bēaga brytta* "distributor of treasure," *i.e.,* "chieftain," *sēoles pæþ* "seal's path," *i.e.,* "sea." Many of these kennings became stereotyped and recur again and again; some of them are purely conventional and have no particular poetic value, while others are vivid and bear the impress of the poet's individuality.

MIDDLE ENGLISH

The Norman Conquest marks an important stage in the history of the English language, though its main effects were not apparent until later than might have been expected, and were often to be seen in other than the obvious places. Numerous French loan-words were borrowed, though these had begun to appear before the Conquest and they did not become numerous before the 14th century. More important is the fact that with the Conquest the literary supremacy of West Saxon was lost; no one dialect had any superior literary claims, and each author was free to use his own. The increasing importance of London during the M.E. period meant that henceforth the dialect of the capital would be supreme. This had originally been a southern—mainly southeastern—dialect, but by the 14th century it had become mainly east midland in character, so that modern Standard English is derived from the East Mercian rather than from the West Saxon dialect.

On the orthography, too, the Conquest had a considerable influence. By the end of the O.E. period, spelling had become more or less fixed, and about 150 to 200 years behind the actual pronunciation. With the breakdown of the old spelling conventions, a new orthography was developed based on the actual spoken language. As a result, sound changes which had taken place in Old English, but had not been represented in the written language, are first regularly recorded in the more phonetic spelling of post-Conquest scribes who had been strongly influenced by French scribal habits. The Conquest was responsible too for a complete change of script. From the 10th century onward occasional use had been made of the new Carolingian script developed on the continent. After the Conquest this script was naturally used by the French-trained scribes, and remained in use, with some changes, until the 15th and 16th centuries. On English syntax also it is probable that French influence was appreciable, though as yet it is hardly possible to assess it properly.

On the other hand, there is not the slightest evidence to show that the Conquest had any effect on the accidence or the phonology. Although Middle English is marked by a general tendency to level all inflectional endings under *-e*, it is clear that the tendency to analogize grammatical forms has always been inherent in the language. Similarly, Middle English is characterized by widespread changes in phonology, but the beginnings of these changes are to be found in the O.E. period.

Orthography.—Many of the O.E. vowels and diphthongs continued to be written long after they must have changed in sound, and in general it is important to remember that many of the changes in the form of a word in Middle English are merely orthographical. Where the O.E. *y̆* sound remained it was normally written *u;* and *y* then came to be used as a spelling for *i* in positions where the retention of *i* might have led to confusion, *i.e.,* before and after *m, n, u*. In similar positions *u* was written *o*, and *ū* in all positions came to be written *ou, ow*. In addition many ambiguities of the O.E. consonant system were cleared up. Front *c* was represented by *ch*, while the back sound continued to be represented by *c* and occasionally *k*. Later the use of *c* and *k* was conventionalized; *c* being used before back vowels and liquids, *k* before front vowels and *n*. In addition, under French influence, *c* was sometimes used as a spelling for *s; e.g., ice, cinder, mice*. The use of *g* for the back stop continued, but for the front sound a new symbol, ȝ (yogh), developed from the O.E. insular *g*, was used. By the 14th century this tended to be replaced by *i/y*. The sound indicated by O.E. *cg* was written *gg* medially and finally, and the same sound in French loan-words was written *i/j* initially. The *h* was used as a spelling for the aspirate only; the open consonant, whether front or back, being represented by ȝ in the early period and by *gh* later. In the northern dialects *hw-* was represented by *qu-, quh-, qw-*, and by *wh* elsewhere, while *sc* was usually written *sh*. An *f* was used for the voiceless sound only, the voiced sound being written *v/u* initially

and *u* medially. The symbol *ð* fell into disuse in the 13th century, but *þ* continued throughout the period, though *th* came to be used side by side with it more and more frequently. Similarly wynn (*ƿ*) continued in use until the end of the 13th century, often confused with *þ* and *y*, but after the Conquest *u, uu, w*, were also used, and by the 14th century *w* had become the normal spelling.

Dialects.—The usual grouping of the M.E. dialects corresponds roughly with that of the O.E. ones; Northumbrian splits into Scots and Northern English, Mercian into East and West Midland, West Saxon is the basis of the South Western and Central Southern dialects, and Kentish the basis of the South Eastern ones. Exact boundaries cannot of course be given for these dialects, though for the sake of convenience approximate ones are often indicated on a map. But in reality a dialect rarely has precise boundaries, the characteristics of the different dialects shading gradually into one another.

Knowledge of the M.E. dialects is also very uneven; nothing is known, for example, of Northern English before the end of the 13th century. The distinction between Scots and Northern English probably did not arise before the end of the 14th century, but it may well be that only the union of the crowns and the adoption of the authorized version of the Bible by the Scottish reformers prevented the development of Scots as a separate language.

Phonology.—The sound changes which distinguish Middle from Old English are conveniently divided into quantitative and qualitative. The most important of the former was the lengthening of short vowels before certain consonant groups, notably *-ld, -mb, -nd, -ng;* and in open syllables of dissyllabic words. These changes were particularly important for the later history of English, since, while the M.E. short vowels underwent comparatively little change, the long vowels were subsequently completely altered in character. Before double consonants, and most combinations of consonants, long vowels were shortened. This also had a considerable effect on the later language; *e.g.*, O.E. *fēdan* "to feed," had a past tense *fēdde*.

In Middle English the long vowel remained in *fēden* but was shortened in *fedde*, and this explains the difference between modern *feed:fed*. Similar shortenings took place also in numerous other weak verbs.

The changes in quality were much more extensive and important than the changes in quantity. The more important of them were: O.E. *æ* gave *e* in the southeast and southwest midlands, but *a* elsewhere; *y̆* gave *ĭ* in the north and east midlands, *ĕ* in the southeast, but remained spelled *u* in the southwest and west midlands; *ā* was rounded to *ō* south of the Humber, but remained in the north until the 14th century when it had probably become open *ē* (the sound of Fr. *ê*); O.E. *ǣ* was from two sources: from the fronting of O.E. *ā* (Germanic *ē*), and from the mutation of O.E. *ā* (Germanic *ai*). These had different developments in the Middle English dialects; both appear as *ā* in Essex and in a part of the adjacent counties, but the former became open *ē* in the southwest, close *ē* (the sound of Fr. *é*) elsewhere; the latter appears as close *ē* in the southeast, as open *ē* elsewhere; *ō* remained except in the north where it became [*ȳ*], represented by *u, ui*.

The diphthongs were monophthongized; *ea* usually to *a*, but before *-ld* lengthening took place. The lengthened vowel then fell in with *ēa* in the south and, like it, became open *ē*, but in the north and midlands the vowel in such circumstances was *a* which was lengthened and like original *ā* gave *ō* in the midlands but remained in the north. The *ĕo* gave *ĕ* in the north, east midlands and southeast, but in the southwest and west midlands a mid-front rounded sound represented by *o, eo, u*, etc.; *ēa* gave open *ē* in all dialects except the southeast where it often appeared as *ya, yea* (perhaps *ie*), especially after dentals; *ēo* which normally existed only in Kent gave *ie/ye* in the southeast. New diphthongs were developed from the O.E. combinations *-ǎh, -āh, -ǒg, -āg, -æ̆g, -ǣg, -ĕh, -ēh* which became respectively *ouh, auh/ouh, ou, au/ou, ai/ei, aih/eih, eih*. The *ēi* diphthong was monophthongized to *ī* in many of the southern and midland dialects as early as the 13th century, and these *ī* forms spread into the London dialect during the 14th, hence such forms as *high* (O.E. *hēah*), *nigh* (O.E. *nēah*), etc.

Accidence.—In most dialects grammatical gender appears to have been lost early in the period, but remnants survive in the south until the end of the 13th century. In accidence Middle English occupies an intermediate position between the well-developed inflectional system of Old English and the almost complete flexionless system of modern English. Most masculine nouns fall within the pattern, singular in a consonant or *-e*, plural in *-es*, and most feminine and neuter nouns later came over into this declension. More of the mutation plurals survived than in modern English, and a relic of the O.E. weak declension is still to be found in mod. Eng. *oxen*, such plurals being fairly frequent in the south until the 14th century. In general the adjectival inflections were lost at an early date. The strong declension of an adjective ending in a consonant had no inflection in the singular, but otherwise *-e* was usual whether strong or weak, singular or plural. In the pronouns the most important variations were in the feminine pronoun and in the plural of the third personal pronoun. The feminine pronoun could be a descendant of O.E. *hēo*, appearing in a variety of forms, *hi, he, ha, heo, hue, ȝhe*, etc., or it could be the ancestor of mod. Eng. *she*. The latter is first found as *scæ* in the east midlands about the middle of the 12th century and spreads rapidly into other dialects in the forms, *sche, scho, she*, etc. Similarly the plural of the third personal pronoun could have one of two basic forms. It could be the descendant of O.E. *hīe, hēo*, appearing in Middle English as *hi, he, heo, hue, ho*, etc., or it could be derived from the O.N. loans *þei(r), þeir(r)a, þeim*, of which *þei* appears first in the east midlands about the beginning of the 13th century, and the oblique forms in the same dialect later in the century. The elaborate inflections of the definite article gave way very early in the north and midlands before the newly formed invariable *þe* used in all cases, genders and numbers. In the verbs the following were the more important variations:

	North	Midlands	South
Inf.	*bind*	*binde(n)*	*binden: louien, lokin*
Pres. part.	*-and*	*-ende*	*-inde*
3rd sing. pres. ind.	*-is*	*-eþ, -es*	*-eþ*
Pres. ind. pl.	*-e, -is*	*-en*	*-eþ*
Past part.	*bounden*	*bounden*	*ybounde*

In the south the *-n* of the infinitive remained longer than in the north or midlands, and the old second weak class was kept separate until late in the period with an infinitive in *-ien, -in*. In the present participle northern *-and* was a borrowing from Old Norse; midland and southern *-ende, -inde* were from Old English, though raising of *e* to *i* before the nasal had taken place in the latter. Modern *-ing* appeared first during the 13th century, and spread rapidly through the south and midlands. The northern *-is* of the third person singular present indicative spread rapidly and was early found in the north midlands; from it is derived the *-s* ending of the modern form. The midland ending of the present indicative plural derived from the O.E. subjunctive plural; and the initial *ge-* of the past participle, originally perfective in meaning, remained in the south as *y*, but was early lost in the north and midlands where, however, the final *-n* remained. Characteristic of the north and west was the leveling of the singular and plural of the past tense, when they differed in vowel, under one or the other form.

Vocabulary.—Undoubtedly the most important influence on the vocabulary during the M.E. period was from French. Some words had been borrowed before the Conquest, *e.g., capun, prud* "proud," *sot*, but the vast majority entered the language during the 13th and 14th centuries. Two distinct periods of borrowing can be observed, with 1250 as the approximate dividing line. The borrowings of the first period differ from those of the second in being much less numerous and in being in the main borrowed from the northern French dialects, especially from Norman-French. In the second period loans from the literary central French dialect began to appear.

During the earlier period many of the words borrowed were such as would become familiar to the lower classes through con-

tact with a French-speaking nobility, *e.g., baron, noble, feast, prison.* Others, such as *story, letter, lay,* etc., owed their introduction to the overwhelming influence of French literature. The largest single group among the words that came in early was associated with the church, which is sufficiently explained by the influx of French-trained priests after the Conquest, *e.g., miracle, procession, passion, image, pilgrim, virgin.* In the later period the conditions under which French words had been making their way into English were supplemented by the fact that those who had been accustomed to speaking French were increasingly turning to the use of English. Partly to supply deficiencies in the English vocabulary, and partly from habit, an astonishing number of common French words were carried over into English. These included much of the governmental and administrative vocabulary, ecclesiastical, legal and military terms, as well as familiar words dealing with fashion, food and social life, together with much of the vocabulary of art, learning, literature and medicine. The fact that the earlier words were usually borrowed from northern, the later from central, French meant that occasionally the same word was borrowed in a slightly different form from both dialects, and if the two words became differentiated in meaning both might survive in modern English. Hence *catch/chase, canal/channel, real/royal, leal/loyal, wage/gage, warden/guardian, warrant/guarant-ee,* etc., in which the first is the northern, the second the central French form. The influx of French loans naturally led to the loss of many O.E. words, though this loss was rarely immediate. Usually both continued in use for a time, and the English word sometimes still survives in dialect. Where both English and French words survived they were generally differentiated in meaning, *e.g., hearty/cordial, ask/demand, answer/reply, house/mansion, ghost/spirit, room/chamber, shun/avoid, seethe/boil, yearly/annual, wish/desire, body/corpse, ship/vessel, spring/fountain.* An interesting group of words illustrating the principle is *ox, sheep, swine, calf,* by the side of the French equivalents *beef, mutton, pork, veal.* The words in French denoted, and still denote, the animal, but in English they were used from the beginning to distinguish the meat from the living animal.

The importance of the French element in the M.E. vocabulary tends to overshadow the words borrowed from Latin. Many came through French, but there were also a number direct from Latin, the 14th and 15th centuries being especially prolific in such borrowings. These included terms relating to law, medicine, theology, science and literature, as well as words which, in the beginning were specifically technical, but later acquired a more general application; *e.g., diocese, client, psalm, equivalent, allegory, formal, library, dissolve, ascension, impediment, alienate, dissent,* etc. The introduction of unusual words from Latin became a conscious stylistic device during the 15th century. It was extensively used by poets who, by means of such words as *dispone, diurne, equipolent, tenebrous,* etc., attempted what has been described as a kind of stylistic gilding, and consequently this feature of their language is known as "aureate diction." Most of the words introduced in this way never took root in the language and were soon lost. Yet with some the novelty wore off, and words such as *laureate, mediation, oriental, prolixity,* which were aureate in Chaucer, have become part of ordinary speech.

The languages of the Low Countries—Flemish, Dutch and Low German—formed another important source of loan-words during the M.E. period, due to the constant and close relations between these countries and England.

Some words were borrowed comparatively early, *e.g., dote, huckster, tackle, booze, splint, wainscot, skipper,* but most of them came in during the 15th and 16th centuries and were commercial, as *mart, hop, pickle, sled, spool, excise, hawker,* etc., or nautical as *buoy, deck, hoist, dock, splice, yacht, reef,* etc.

MODERN ENGLISH

By the second half of the 15th century London English had been accepted as a standard literary language in most parts of the country, and in the modern period it is with the history of this particular dialect that this article is concerned. It is, of course, only one of a number of varieties of spoken English in Britain which, in general, may be divided into three classes: Received Standard English, *i.e.,* the English of the educated classes used without self-consciousness; modified standard, *i.e.,* a compromise between the former and some regional or occupational dialect; and regional dialects, *i.e.,* those of localized use in rural areas. In addition there is the mixed speech of industrial towns, occupational dialects, and such varieties as slang, jargon, etc. It is not, however, possible to deal in any detail with all these varieties, for most of which available information is in any case far too slight.

One of the most important influences in establishing London English as the standard literary language was the introduction of printing in 1476. From the beginning London was the centre of book-publishing in England; William Caxton in his numerous translations used the current London speech, and the books published by him and by his successors gave an impetus which more than anything else led to the rapid adoption of London English. Not that complete uniformity was quickly attained. The standard of the 16th century was that variety of English spoken by the court and the upper grades of society. But far into the 17th century this standard was nothing so fixed and definite as at present, when there is a single accepted pronunciation for the vast majority of words in ordinary polite use, deviation from which is regarded either as a provincialism or as a vulgarism. From the 16th to the 18th centuries considerable variation of pronunciation existed, and was tolerated, among those who were regarded as speaking according to the fashionable standard. Rustic forms of speech, for example, were certainly in use among many of the country gentlemen who came to London from time to time from their estates. To have an appreciable rustic smack to one's speech might provoke the mirth of town wits and make a man the butt for writers of comedies, but it did not disqualify him from frequenting polite society. But from the middle of the 18th century—the age of Lord Chesterfield and Samuel Johnson—new and more rigid standards of correctness began to appear. The letters of the former show how desirous he was of standardizing the language: "A man of fashion takes great care to speak very correctly and to pronounce properly—that is, according to the use of the best company." Samuel Johnson speaks of "a double pronunciation, one cursory and colloquial, the other regular and solemn. The cursory pronunciation is always vague and uncertain, being made different in different mouths by negligence, unskilfulness and affectation. The solemn pronunciation, though by no means immutable and permanent, is yet always less remote from the orthography and less liable to capricious innovation." Even more significant is his dictum that "For pronunciation the best general rule is to consider those the most elegant speakers who deviate least from the written words." This process of deliberately substituting a new, careful pronunciation, based on spelling, for the traditional pronunciation is always at work, though many of the older types of pronunciation die hard, and linger on for generations.

It was the early modern period too which saw the struggle of English for recognition as the language of scholarship. The traditional importance of Latin in this respect had been strengthened by the revival of learning; in addition it had the advantage of universal currency, while in comparison with it the vernaculars were felt to be immature and unstable, unpolished and incapable of expressing abstract ideas. Despite such views the vernaculars now began to have their champions. Translations and original works in them began to pour from the presses, and although earlier writers in English are often apologetic over using the vernacular instead of Latin, by the end of the 16th century English had been accepted as the language of serious scholarship.

Orthography.—The modern period saw the final development of a fixed spelling, although the introduction of printing did not immediately lead to any great changes in the orthography. Caxton simply took over the conventions of the best scribes, and in general the early printers adhered to medieval scribal traditions. Nor did they as yet see any particular virtue in consistency.

In fact, not infrequently Elizabethan compositors varied the forms of words in order to justify their lines of type, so that readers were accustomed to see the same word spelled in different ways on the same page, and consequently felt free to spell much as they pleased themselves. Despite this, the influence of the printers gradually tended toward the establishment of a fixed spelling, and the process is said to have been hastened by the English Civil War. In the haste of producing political pamphlets, compositors had no time to bother with more than one way of spelling a word, even for the sake of obtaining a better-spaced line. Unfortunately it was at this time, just when the printers were gradually evolving their own spelling conventions, that English pronunciation, especially in its vowels, underwent considerable changes. The result was that by the time the printers had managed to fix the orthography it was already completely unphonetic. Further discrepancies between sound and symbol arose when letters were inserted in words, where they were not pronounced, for purely etymological reasons (*e.g., debt, doubt*), or by analogy (*e.g., delight, spright-ly*).

Nevertheless the 17th century saw most of the current spelling conventions established in printing, while during the 18th century the immense influence of dictionaries finally brought about the definite acceptance of a permanent fixed spelling. Much of the credit for this usually goes to Samuel Johnson's *Dictionary* (1755), not without justification. By his time spelling was already well on the way to becoming fixed, but his dictionary was the first to receive universal acceptance as the arbiter of usage and spelling, and it was more or less imitated by dictionaries during the following century. The result is that in the main it is the spelling of Johnson's dictionary that is at present in use, a system not very far removed from the conventions of the best scribes during the late M.E. period. It was in the 18th century too that the Gothic or "black-letter" type used by the early printers finally gave place to Roman characters and long *s* (ſ) was thus replaced by ordinary *s*.

Apart from the establishment of a fixed spelling, there were few innovations in the orthography. Perhaps the most marked feature has been the inevitable tendency to restrict freedom of choice, one of several possible spellings having become the invariable one, though variation still exists in a few words, *e.g., show/shew, gray/grey*. Sometimes the variant spellings became associated with different meanings and are now regarded as separate words, *e.g., metal/mettle, flour/flower*. The sounds developed from M.E. open *ē, ō*, came in the 16th century to be represented by *ea, oa* (*e.g., eat, deal, loaf, broad, etc.*), though *oa* is avoided finally (*e.g., foe, toe*). Similarly it became usual to represent the sounds developed from M.E. close *ē, ō*, by *ee, oo*, and in addition the spelling *ie* was also used for the former, due partly to French influence, hence *keen, see, cool, doom, field*, etc. The M.E. tendency to use *y* as a spelling for *i* next to *m, n, u*, did not long survive the invention of printing. In modern English *i* is preferred initially and medially, but *y* finally. In the same way the M.E. preference for *v* initially and *u* medially gave way after the 16th century to the practice of using *u* for the vowel and *v* for the consonant. Similarly *j* came to be used for the consonant, *i* for the vowel.

In early modern English there was a good deal of fluctuation between the spellings *ou/ow;* today *ow* is generally used finally (except in *thou*) and before a vowel, *l* or final *n*, while elsewhere *ou* is usual. There was also a good deal of uncertainty in the use of final *-e;* sometimes it indicated a preceding long vowel, but often its addition or omission seemed to depend entirely on the writer or printer, and this confused state of affairs has remained.

Phonology.—In Old and Middle English the spelling remained a fairly reliable source for obtaining knowledge of the pronunciation, but the development of a fixed spelling early in the modern period meant that this was no longer the case, and other types of evidence have to be used. The most important of these are: (1) the spellings in private writings, since these are not bound by conventions, *e.g.*, the *Paston Letters, Verney Memoirs*, and the like; (2) the rhymes of poets in the early period, with which may be included puns. But it is often difficult to be certain whether the rhyme or pun is exact or approximate, and by the 18th century eye-rhymes were beginning to appear. In any case such evidence, at the best, can merely show that two words have a common sound; the rhyme or pun by itself cannot tell us what that sound was; (3) early writers on pronunciation, whether foreigners writing to explain how the sounds of English compare with those of their native language, or English writers interested in the subject as phoneticians, spelling reformers or practical teachers. Their works vary in value; of the spelling reformers, J. Hart (1551–70), R. Hodges (1643–44) and especially A. Gill, Sr. (1619) are the most useful. Few of the phoneticians provide trustworthy evidence before J. Wallis (1652) and C. Cooper (1685); (4) historical considerations, *i.e.*, from a knowledge of the sound at the end of the M.E. period, and its present representative it is possible to deduce to some extent the stages through which it must have passed. From these various sources a good deal of evidence can be obtained, but it must be emphasized that no one source is sufficient by itself, and any presupposed pronunciation must be corroborated by more than one type of evidence.

The most important change distinguishing the pronunciation of modern from Middle English was the Great Vowel Shift which took place during the 15th and 16th centuries and modified the entire vowel harmony of the language. By this change the long high vowels *ī* and *ū* were diphthongized, while the other long vowels were raised. Hence M.E. *ī* became [ɔi], later [*ai*]; close *ē* gave [*i:*]; open *ē* gave [*e:*] which later became [*i:*]; *ā* was fronted to [ɛ:], raised to [*e:*], and then diphthongized to [*ei*]; open *ō* was raised to [*o:*] and later diphthongized to [*ou*]; close *ō* was raised to [*u:*]; and *u* diphthongized to [ʌ*u*] later [*au*]. Hence

Chaucer		*Shakespeare*	*Contemporary*	
Spelling	Pronunciation	Pronunciation	Pronunciation	Spelling
tyme	ti:mə	taim	taim	time
dede	de:də	di:d	di:d	deed
clene	kle:nə	kle:n	kli:n	clean
make	ma:kə	me:k	meik	make
stone	stɔ:nə	sto:n	stoun	stone
mone	mo:nə	mu:n	mu:n	moon
house	hu:sə	hʌus	haus	house

The short vowels were much less extensively changed, the only two at all seriously affected being *a* and *u*. Short *a* was fronted to [*æ*], except after *w* where it was rounded to [ɔ], and before *f, s, th*, where lengthening eventually resulted in [*a:*]; hence *hat, wash, castle*. Short *u* was unrounded and lowered to the sound heard in mod. Eng. *shut, hurry*, except after a labial when the full *u* sound often remained; *e.g., put*. In addition, at an early period *er* had become *ar*, sometimes with a change in the spelling, sometimes without; *e.g., star, hart, clerk, Derby*, etc. Where the spelling with *-er* remained there was a tendency for a new pronunciation to arise based on the spelling, as in *servant, merchant*, but such rhymes as *served: starved* (Dryden), *charge ye: clergy* (Swift), etc., should be noted. Rather more extensive changes occurred in the diphthongs; *au* became [ɔ*u*] which gave [ɔ:], as in *law, draw*, but before a lip consonant the result was [*a:*], as *laugh, laughter, draught*, and similarly earlier *a* before *l* plus a lip consonant which had also become *au* eventually became [*a:*]; *e.g., half, calf, balm*. The *ou* before a following *h* gave [ɔ:], as in *brought, sought*, but otherwise it still appears as [*ou*], *e.g., bow, flow*, though what, if any, may have been the intervening changes is still uncertain. M.E. *ei* and *ai* had fallen together in *ai* during the M.E. period, and this has given [*ei*] in modern English, *e.g., clay, day, eight*, etc., except before *r* where the result is [ɛə]. M.E. *eu* and *iu* had similarly fallen together in *iu*, and the result in modern English is [*ju:*], as in *duke, due, Tuesday*, but after *l* and *r* [*u:*] is usual, as in *look, salute, rude*, and after *s* variation between [*ju:*] and [*u:*] is not uncommon, as in *supreme, suit, assume*. The *oi* appears in modern English as [ɔi], but it is clear that this is not the historically regular development. In the 16th century it had become [*ai*] and, as shown by rhymes such as *refines: joins: lines* (Dryden), *toil: pile* (Pope), had fallen together with M.E. *ī* which had attained the same stage of development, and both should then have developed to [*ai*]. The modern pronunciation with [ɔi] is due entirely to the influence of the spelling. The other diphthongs appearing in modern English

[eə], [iə], [əə], [uə], are due to the special development of early mod. Eng. open *ē*, *ī*, open *ō*, and *ū*, before a following trilled *r*. In unstressed syllables, vowels and diphthongs have usually been weakened, front vowels, including *a* which had been fronted to *æ*, and *ai*, to *i*, the back vowels and *au*, *ou*, to the neutral vowel [ə]. The *oi* diphthong became *i*, but in modern English it has usually been restored in the unstressed as in the stressed positions. In fact, since the original spelling of the unstressed vowels was usually retained it has not infrequently influenced the pronunciation, so that in modern English the quality of the vowel may be restored to something like its earlier character.

Few of the consonant changes are of importance, the only ones worth noting being the loss of front *h*, the loss or change to *f* of back *h*, and the loss of *r* and *l*. Front *h*, usually spelled *gh* in late Middle English, was lost sufficiently early for the preceding *i* to be lengthened and to fall in with M.E. *ī*, like it being diphthongized to [*ai*], but the *gh* still remained in the spelling, as in *night*, *fight*, *might*, etc. Back *h* either disappeared or became *f*, but again *gh* is retained in the spelling; *e.g.*, *cough*, *enough*, *though*. In Standard English *r* disappeared before consonants and finally, often with lengthening of the preceding vowel, as in *hard*, *harsh*, *car*, *hear*, but finally it is retained before a word beginning with a vowel. Where it remains the trill has been lost and, except in some of the dialects, *r* is now merely a variety of weak open consonant. The *l* was lost between a back vowel and a following lip consonant or *k* (*e.g.*, *calm*, *calf*, *walk*), though in some cases its retention in the spelling led to its reintroduction into the pronunciation (*e.g.*, *almanac*, *palfrey*, etc.).

Grammar.—By the end of the M.E. period the only inflections retained in the noun were those marking the plural and the possessive singular, while the adjective had lost all its endings. In both little change took place in the modern period. In the pronouns the 16th century saw the establishment of the personal pronouns in their present form. This involved the disuse of *thou*, *thee*, *thy*, the substitution of *you* for *ye*, and the development of *its* as the possessive of *it*. The same century saw also the development of the use of *who* as a relative. In the verb, *-eth* as the ending of the third person singular present indicative gradually gave way to the *-s* ending; some of the original strong verbs developed weak forms, while the subjunctive mood was largely lost.

In syntax the general loss of inflections resulted in a wider use of prepositions and a more rigid word-order. In the 16th and 17th centuries syntax was still comparatively free, and it was the 18th century which saw the beginnings of prescriptive grammar, many of the conventions accepted in the present textbooks having been established at that time. This formalizing of the language has to some extent continued, and has led for example to a tendency toward the revival of the subjunctive. The progressive forms of the verb (*I am going*, *was going*, *shall be going*, etc.) are first found fully developed in Joseph Addison's writings, and the 18th century sees the rapid spread of this idiom by which so many different shades of meaning can be expressed. During the following century developed the *get* passive (*he got hurt*) and the very great extension of verb-adverb combinations (*e.g.*, *keep on*, *keep up*, *keep out*, *keep in*, etc.).

Vocabulary.—During the modern period the most notable addition to the vocabulary has been the influx of words from classical sources. This was particularly noticeable during the 16th century, when there was not only a wholesale importation of Latin loans, but also a tendency to refashion French-derived words on Latin models; *e.g.*, *descrive*, *perfit*, *verdit*, were transformed into *describe*, *perfect*, *verdict*. The frequency of borrowing led to a reaction against the use of such "inkhorn terms," but despite this, borrowing from Latin and Greek has continued throughout the modern period, and these two languages have become an indispensable quarry for modern technical and scientific terms. In addition most of the languages of the world have contributed to the modern vocabulary. French has been particularly influential, and during the 20th century borrowings from French have continued; *e.g.*, *garage*, *revue*, *fuselage*, *camouflage*, etc. From Italy has come much of the technical vocabulary of music, painting and architecture (*soprano*, *trombone*, *fiasco; studio*, *replica; baroque*); the Dutch contribution contains many words connected with commerce and the sea (*smack*, *sloop*, *cruise*, *taffrail; brandy*, *gin*, *colza*); and modern German is particularly notable for mineralogical terms (*cobalt*, *shale*, *quartz*, *wolfram*, *nickel*, *loess*). From farther afield Arabic has given *loofah*, *divan*, and India contributed *gymkhana*, *dinghy*, *loot*, *polo*. Other words come from Turkey (*fez*, *bosh*), Japan (*tycoon*, *ju-jitsu*), Malay-Polynesia (*raffia*, *gutta-percha*, *taboo*, *tattoo*, *ukulele*), South America (*quinine*, *sisal*, *tapioca*) and China (*ketchup*, *kaolin*, *kowtow*).

Side by side with this widespread borrowing English has retained the power of coining new words from its own resources. It can still form self-explaining compounds, such as *searchlight*, *aircrew*, *Caterpillar tractor;* it joins familiar prefixes and suffixes to existing words, thereby forming new ones; *e.g.*, *transformer*, *postgraduate*, *antenatal*, *superman*, and proper names, whether personal or place names, continue to provide new words, *mackintosh*, *bakelite*, *sandwich*, *dahlia*, etc. Particularly modern is the use of initial letter coinages such as UNESCO, ASLIB, PEN and similarly TB, VD, GI and BBC.

See AMERICAN ENGLISH; BASIC ENGLISH; SLANG; *see* also references under "English Language" in the Index volume.

BIBLIOGRAPHY.—*General Works: General Works on Language and on Indo-European Languages:* L. Bloomfield, *Language* (1933); H. A. Gleason, *An Introduction to Descriptive Linguistics* (1955); L. H. Gray, *The Foundations of Language* (1939); L. Hjelmslev, *Principes de Grammaire Générale* (1928); H. Hirt, *Handbuch der Urgermanischen* (1931–34); O. Jespersen, *Language: Its Nature, Development and Origin* (1922); A. Meillet, *Introduction à l'étude comparative des langues indo-européennes* (1937), *Caractères généraux des langues-germaniques* (1930); A. Meillet and M. Cohen, *Les Langues du Monde* (1952); E. Prokosch, *A Comparative Germanic Grammar* (1939); E. Sapir, *Language* (1921); E. Sturtevant, *An Introduction to Linguistic Science* (1947); J. Vendryès, *Language*, trans. by P. Radin (1925); H. C. Wyld, *Historical Study of the Mother Tongue* (1913). *General Bibliographical Works: Annual Bibliography of English Language and Literature* (Modern Humanities Association); *Cambridge Bibliography of English Literature*, ed. by F. W. Bateson (1940); A. G. Kennedy, *A Bibliography of Writings on the English Language From the Beginning of Printing to the End of 1922* (1927); *Linguistic Bibliography*, published by a permanent International Committee of Linguists (vol. 1–2 for 1937–48 and annually thereafter); *The Year's Work in English Studies* (1921 *et seq.*). Following are some of the more important periodicals publishing articles on the English language, many of them containing bibliographies: *Anglia; English Studies; Language; Modern Language Notes; Modern Language Quarterly; Modern Language Review; Modern Philology; Publications of the Modern Language Association of America; Philological Quarterly; Word.* General Works on the History of English: A. C. Baugh, *A History of the English Language* (1935); N. Bøgholm, *English Speech From an Historical Point of View* (1939); K. Brunner, *Die Englische Sprache* (1950–51); R. Huchon, *Histoire de la langue anglaise* (1923–30); O. Jespersen, *Growth and Structure of the English Language* (1905); C. Luick, *Historische Grammatik der englischen Sprache* (1914–29); G. H. McKnight and B. Emsley, *Modern English in the Making* (1928); A. H. Marckwardt, *Introduction to the English Language* (1942); F. Mossé, *Esquisse d'une histoire de la langue anglaise* (1947); C. T. Onions, "The English Language," *The Character of England*, ed. by Sir Ernest Barker (1947); S. Potter, *Our Language* (1950); C. L. Wrenn, *The English Language* (1949); H. C. Wyld, *A Short History of English* (1927). Works on Vocabulary: J. F. Bense, *A Dictionary of the Low-Dutch Element in the English Vocabulary* (1926–32); C. T. Carr, *The German Influence on the English Vocabulary* (1934); G. N. Clark, *The Dutch Influence on the English Vocabulary* (1935); J. B. Greenough and G. L. Kittredge, *Words and Their Ways in English Speech* (1901); B. Groom, *A Short History of English Words* (1934); G. H. McKnight, *English Words and Their Background* (1923); M. Serjeantson, *A History of Foreign Words in English* (1935); J. A. Sheard, *The Words We Use* (1954); J. C. Smock, *The Greek Element in English Words*, ed. by P. W. Long (1931).

The Old English Period: Dictionaries: J. Bosworth and T. N. Toller, *An Anglo-Saxon Dictionary*, with *Supplement* (1898–1921); J. R. Clark Hall, *A Concise Anglo-Saxon Dictionary* (1931). Grammars: K. Brunner, *Altenglische Grammatik nach der angelsächsischen Grammatik von Eduard Sievers* (1942); K. Bülbring, *Altenglisches Elementarbuch*, 1. Teil, "Lautlehre" (1902); N. Davis, *Sweet's Anglo-Saxon Primer* (1953); R. Girvan, *Angelsaksisch Handboek* (1931); R. Quirk and C. L. Wrenn, *An Old English Grammar* (1955); E. E. Wardale, *An Old English Grammar* (1926); J. and E. M. Wright, *Old English Grammar* (1925). *Readers and Collections of Extracts:* P. S. Ardern, *First Readings in Old English* (1953); J. W. Bright, *An Anglo-Saxon Reader*, rev. by J. R. Hulbert (1935); G. T. Flom, *Introductory Old English Grammar and Reader* (1930); R. Kaiser, *Altund Mittelenglische Anthologie* (1954); F. Mossé, *Manuel de l'anglais du Moyen Age, 1. Vieil anglais* (1945); H. Sweet, *Anglo-Saxon Reader,*

rev. by C. T. Onions (12th ed., 1950); A. J. Wyatt, *An Anglo-Saxon Reader* (1919). *Vocabulary:* O. Funke, *Die gelehrten lateinischen Lehn-und Fremdwörter in der altenglischen Literatur* (1914); H. Gneuss, *Lehnbildungen und Lehnbedeutungen im Altenglischen* (1955); R. Jordan, *Eigentümlichkeiten des anglischen Wortschatzes* (1906); H. S. MacGillivray, *The Influence of Christianity on the Vocabulary of Old English* (1902); A. Pogatscher, *Zur Lautlehre der griech., latein., und roman. Lehnwörte im Altenglischen* (1888). *Handwriting and Runes:* H. Arntz, *Handbuch der Runenkunde* (1944); N. Denholm Young, *Handwriting in England and Wales* (1954); R. Derolez, *Runica Manuscripta* (1954); W. Keller, *Angelsächsische Palaeographie* (1906); O. von Friesen, *Runorna* (1933).

Middle English: Dictionaries: H. Kurath and S. M. Kuhn, *Middle English Dictionary* (1952–); F. H. Stratmann, *A Middle English Dictionary*, rev. by H. Bradley (1891). *Grammars:* G. Fridén, *Studies in the Tenses of the English Verb From Chaucer to Shakespeare* (1948); R. Jordan, *Handbuch der mittelenglischen Grammatik* (1925); M. M. Long, *The English Strong Verb From Chaucer to Caxton* (1944); L. Morsbach, *Mittelenglische Grammatik* (1896); J. and E. M. Wright, *An Elementary Middle English Grammar* (1928). *Readers and Collections of Extracts:* A. Brandl and O. Zippel, *Mittelenglische Sprach- und Literaturproben* (1927); A. S. Cook, *A Literary Middle English Reader* (1915); B. Dickins and R. M. Wilson, *Early Middle English Texts* (1952); O. F. Emerson, *A Middle English Reader* (1915); J. Hall, *Selections From Early Middle English 1130–1250* (1920); F. Mossé, *A Handbook of Middle English*, trans. by J. A. Walker (1952); K. Sisam, *Fourteenth Century Verse and Prose* (1937). *Vocabulary:* E. Björkman, *Scandinavian Loan Words in Middle English* (1900–02); R. Feist, *Studien zur Rezeption des französischen Wortschatzes im Mittelenglischen* (1934); J. Orr, *The Impact of French Upon English* (1948); H. T. Price, *Foreign Influences on Middle English* (1947); E. Slettinger, *Contributions to the Study of French Loanwords in Middle English* (1932). *Handwriting:* C. Johnson and H. Jenkinson, *English Court Hand From 1066–1500 A.D.* (1915). *London English:* R. W. Chambers and B. M. Daunt, *A Book of London English 1384–1425* (1931); H. M. Flasdieck, *Forschungen zur Frühzeit der neuenglischen Schriftsprache* (1922), *Mittelenglische Originalurkunden (1405–1430)* (1926); B. A. Mackenzie, *The Early London Dialect* (1928); L. Morsbach, *Ueber den Ursprung der neuenglischen Schriftsprache* (1888). *Bibliographical:* J. E. Wells, *A Manual of the Writings in Middle English 1050–1400*, with *Supplements I-IX* (1916–51).

Modern English: Dictionaries and Lexicography: J. R. Hulbert, *Dictionaries: British and American* (1955); M. M. Mathews, *A Survey of English Dictionaries* (1933); J. A. H. Murray *et al.*, *A New English Dictionary on Historical Principles* (1884–1933); E. H. Partridge, *A Dictionary of Slang and Unconventional English* (1949); C. T. Onions, *A Shakespeare Glossary* (1919); W. W. Skeat and A. L. Mayhew, *A Glossary of Tudor and Stuart Words* (1914); D. W. T. Starnes and G. E. Noyes, *The English Dictionary From Cawdrey to Johnson 1604–1755* (1946); D. W. T. Starnes, *Renaissance Dictionaries* (1954); H. C. Wyld, *The Universal Dictionary of the English Language* (1936). Grammars: C. C. Fries, *The Structure of English* (1952); H. W. Fowler, *A Dictionary of Modern English Usage* (1926); G. O. Curme, *A Grammar of the English Language* (1931–35); W. Horn and M. Lehnert *Laut und Leben* (1954); O. Jespersen, *A Modern English Grammar on Historical Principles* (1909–49); A. Kihlbom, *Contribution to the Study of Fifteenth-Century English* (1926); H. Poutsma, *A Grammar of Late Modern English* (1914–29); J. Steinki, *Die Entwicklung der englischen Relativpronomina in spätmittelenglischer und frühneuenglischer Zeit* (1932); H. Sweet, *New English Grammar* (1892–98); H. C. Wyld, *A History of Modern Colloquial English* (1936), *Studies in English Rhymes From Surrey to Pope* (1923); R. E. Zachrisson, *Pronunciation of English Vowels 1400–1700* (1913); R. W. Zandvoort, *A Handbook of English Grammar* (1957). *Pronunciation:* A. Cohen, *The Phonemes of English* (1952); D. Jones, *An Outline of English Phonetics* (1940), *An English Pronouncing Dictionary* (1957); W. Franz, *Die Sprache Shakespeares in Vers und Prosa* (1939); H. Kökeritz, *Shakespeare's Pronunciation* (1953); I. C. Ward, *The Phonetics of English* (1945); I. C. Ward and L. E. Armstrong, *Handbook of English Intonation* (1926); D. Lloyd James, *The Broadcast Word* (1936); E. J. Dobson, *English Pronunciation 1500–1700* (1957). Spelling: H. Bradley, *On the Relations Between Spoken and Written Language, With Special Reference to English* (1919); Sir W. A. Craigie, *English Spelling: Its Rules and Reasons* (1927), *Some Anomalies of Spelling* (1942). *Writing:* H. Jenkinson, *The Later Court Hands in England From the 15th to the 17th Century* (1927). *Vocabulary:* A. A. Dayrush, *Persian Words in English* (1934); M. Förster, *Keltisches Wortgut im Englischen* (1921); A. Ksoll, *Die Französischen Lehn- und Fremdwörter in der englischen Sprache der Restaurationszeit* (1933); P. Leidig, *Französische Lehnwörter und Lehnbedeutungen im Englischen des 18 Jahrhunderts* (1941); Mario Praz, "The Italian Element in English," *Essays and Studies*, xv; A. A. Prins, *French Influence in English Phrasing* (1952); G. Subba Rao, *Indian Words in English* (1954); W. Taylor, *Arabic Words in English* (1933). *Dialects:* A. J. Ellis, *The Present English Dialects*, vol. v in "On Early English Pronunciation" (1889); A. MacIntosh, *Introduction to a Survey of Scottish Dialects* (1952); W. W. Skeat, *English Dialects* (1911); J. Wright, *English Dialect Dictionary* (1898–1905), *English Dialect Grammar* (1905). *Place names:* O. S. Arngart, *The English Hundred Names* (1934–39); E. Ekwall, *The Concise Oxford Dictionary of English Place-Names* (1947), *English River Names* (1928); publications of the English Place-Name Society; Sir A. Mawer, *Problems of Place-Name Study* (1929). *Personal Names:* C. W. Bardsley, *Our English Surnames* (1873), *A Dictionary of English and Welsh Surnames* (1901); G. F. Black, *The Surnames of Scotland* (1946); P. H. Reaney, *Dictionary of British Surnames* (1958); E. C. Smith, *Personal Names: a Bibliography* (1952); E. Weekley, *Surnames* (1936); E. G. Withycombe, *The Oxford Dictionary of English Christian Names* (1950); H. B. Woolf, *The Old Germanic Principles of Name-Giving* (1939).

(R. M. Wn.)

ENGLISH LAW. In the language of English jurisprudence, "legal memory" is said to extend as far back as, but no farther than, the coronation of Richard I (Sept. 3, 1189). This is a technical doctrine concerning prescriptive rights, but it expresses an important truth. For the last seven and a half centuries and more, English law has had not only a continuous but a matchlessly well-attested history, and it has, moreover, been the subject matter of rational exposition. The plea rolls, which recorded the daily doings of a central tribunal which was controlling and forming the entire system, are extant from 1194; and from then onward it is the enormous bulk rather than any dearth of available materials that makes it difficult to plot accurately the transformation of every old doctrine and the emergence and expansion of every new idea.

Continuity and Derivation of English Law.—Looking backward through the legal textbooks, it is possible to trace the history of English law from the present day through the work of Sir William Blackstone (d. 1780), Sir Matthew Hale (d. 1676), Sir Edward Coke (d. 1634), Sir Anthony Fitzherbert (d. 1538), Sir Thomas Littleton (d. 1481) and Henry de Bracton (d. 1268) to Ranulf de Glanville (d. 1190) in the reign of Henry II. It is clear that all these writers treat of one body of law, although in the course of time it has grown vastly and ideas that were once few and indefinite have become many and explicit.

Behind these seven and a half lucid centuries lies a darker period. Nearly another 600 years divide Glanville from the laws of Aethelbert of Kent (d. 616), and nearly 700 from the *Lex Salica* (*see* SALIC LAW). The Norman conquest of England marked the confluence of two streams of law, the French or Frankish and the Anglo-Saxon. The French stream, traced backward, passes through the capitularies of Carolingian emperors and Merovingian kings to the time when Clovis (d. 511), conqueror of Gaul, adopted Catholic Christianity and the Franks began to record their customary usages in written Latin. The Anglo-Saxon stream of law runs backward though the code of Canute (d. 1035), the ordinances of Alfred (d. 899), and the laws of Ine (d. *c.* 726) in newly converted Wessex to those of Aethelbert in newly converted Kent. Bede records that Aethelbert's judicial ordinances were made in accordance with Roman precedents. Aethelbert had perhaps heard from St. Augustine and other Roman missionaries of the great legal work of Justinian. It is, at any rate, interesting to note that in time Justinian's lawbooks came between the *Lex Salica* and the earliest Kentish laws, and that Pope Gregory I, who dispatched the missionaries to England, was one of the very few men living in the western empire who was acquainted with Justinian's *Digest*.

In the west, the Germanic folk laws (*Leges Barbarorum*) were being compiled, and it is as members of this Teutonic group of law that the Anglo-Saxon dooms are best studied. Those of Aethelbert and Ine are closely related to the European folk laws, especially to the *Lex Saxonum* and the laws of the Lombards. The laws of Offa of Mercia (d. 796) are unfortunately lost, but those of Alfred and his successors strongly resemble the Frankish capitularies. But whereas the age of legislation in the Frankish dominions was already over by the time of Alfred, English rulers continued to publish laws; those of Canute were probably the most comprehensive issued in 11th-century Europe.

A high degree of racial purity can be claimed for the Anglo-Saxon dooms. Celtic elements have been sought in them but have never been detected. At certain points, notably in the regulation of the blood feud and the construction of a tariff of atonements, the laws of primitive peoples will always be to some extent similar. But the existing remains of old Welsh and old Irish law resemble the dooms of Aethelbert and Ine far less than do the edicts of the

Lombard kings Rothari (d. 652) and Liutprand (d. 744). It is, indeed, very dubious whether distinctive Celtic customs played any considerable part in the evolution of that system of rules of Anglian, Scandinavian and Frankish origin which became the law of Scotland. Within England itself, although for a while there was fighting enough between the various small kingdoms, the tribal differences were not so deep as to prevent the formation of a common language and a common law. Even the strong Scandinavian strain seems to have blended rapidly with the Anglian. If, for example, in modern times we can distinguish between *law* and *right*, we are debtors to the Danes; but *law* soon ceased to be distinctive of eastern or *right* of western England. In the first half of the 12th century the author of the *Leges Henrici primi* could still say that the country was divided between the law of Wessex, of Mercia and the Danelaw (*q.v.*); but he had also to point out that the law of the king's own court stood apart from, and above, all partial systems. The local customs were those of shires and hundreds, and shaded off into each other; although some counties were more and others less affected by Danish influence, it was a matter only of degree, for rivers were narrow and hills low. England was meant by nature to be the land of one law.

In England and elsewhere, Germanic law developed in an atmosphere that was charged with traditions of the old world, many of which had become implicit in the Christian religion; but there is no evidence of the survival in Britain, or of the reintroduction into England, of anything that could be called Roman jurisprudence. The Anglo-Saxon legal texts show no trace of Roman rule except in the ecclesiastical sphere. Even there, the church-property law of the middle ages, the law of the ecclesiastical benefice, can be shown to be permeated with Germanic ideas. This is true of Gaul and of Italy, and yet truer of an England in which Christianity was for a while extinguished. Moreover, the laws that were written in England were, from the first, written in the English tongue; and this gives them a unique value among Germanic folk laws, for even the very ancient and barbarous *Lex Salica* is a Latin document, although it contains many Frankish words. In England no Romano-British population survived as a subject race, permitted to live under their own law like the *Romani* of the continent in Burgundy or in Visigothic Spain, where Alaric II (d. 507) issued the *Breviarium Alarici* or *Lex Romana Visigothorum* for his Roman subjects. This collection, excerpted chiefly from the Theodosian code, became for a long while the chief representative of Roman law in Gaul and was, indeed, known in England, where St. Aldhelm (d. 709) read it in Canterbury. The Frankish kings ruled over many men whose law was to be found not in the *Lex Salica* or the *Lex Ribuaria* (the Ripuarian law) but in what was called the *Lex Romana*. Law was a matter of race; the *homo Romanus* handed on his Roman law to his children, while Frankish or Lombardic, Swabian or Saxon law would run in the blood of the *homo Barbarus*. But these laws could not continue in juxtaposition and remain quite separate. On the one hand, distinctively Roman rules crept into the laws of the barbarian tribes, while on the other, a decay and debasement of Roman law took place which ended in the formation of what modern historians have called a Roman "vulgar-law" (*Vulgarrecht*). For a short while, round about the time of Charlemagne (d. 814), it seemed possible that the Frankish kings, who were becoming Roman emperors, would be able to rule by their capitularies almost the whole of the Christian west. But the dream vanished before fratricidal wars and heathen Viking invasions, which in turn encouraged the growth of a feudalism at any rate potentially disruptive of centralized government. The personal laws of Roman or barbarian began to transmute themselves into local custom, the Roman *Vulgarrecht* becoming that of those districts where *Romani* were the preponderating element in the population. Meanwhile the Norse pirates acquired possession of what was to become northern France, a land where *Romani* were few. These Normans had a peculiar genius for appropriating and adapting the usages of an alien civilization. In their language, religion and law they had become French many years before they subdued England. It is possible to contend that they preserved some of the traditions of the best days of the Frankish monarchy, and that the Norman dukes rather than the German emperors or French kings were the truest heirs of Charlemagne.

One of the greatest acquisitions which the Normans made when they settled in France was their assimilation of feudalism (*q.v.*). As yet, it was a complex of customs and practices of varied origins, partly Roman and partly Germanic, which was slowly producing a new social system for Europe. As soon as Roman law came to be studied in Bologna, the lawyers were deeply interested by the novelties of the feudal fief; already Irnerius, early in the 12th century, referred to it, the emperor Frederick I legislated about it a little later, and by the end of the century the *Libri feudorum* took shape in Lombardy. In short, feudalism had acquired its own peculiar system of law, thanks to the lawyers, who first tried to interpret it in terms of Roman law but soon accepted it as an independent system which had reached results (such as the theory of *dominium* being divisible between lord and tenant into *dominium utile* and *dominium directum*) which Romanists found hard to reconcile with their older books.

The Normans were already established in England, however, when this was taking place, and they must not be accused of introducing either feudal law or feudalism itself, which developed in England partly from native resources; and the growth of the law of estates served to replace the learned speculations which in Europe tried to create a Romanized feudal law.

HISTORY AND DEVELOPMENT

The Anglo-Saxon Legal System.—With these sources in mind, we can examine the laws by which England was governed before the Norman conquest. Anglo-Saxon England was a primitive agricultural community, newly converted to Christianity, with superstition still everywhere. Orderly government was unknown, and violence, robbery and sudden death were everyday occurrences. Although the community still bore many of the marks of the old Germanic tribal society, some of the characteristics of feudalism were developing. At the bottom of the social pyramid were the slaves. Above them were the freemen or ceorls. These were increasingly becoming dependent on the thegns (*q.v.*), or noblemen. It became usual for every man to have a lord over him. Thus the relationship of lord and man, which was one of the fundamental tenets of feudalism, was already established.

For purposes of government the whole country was, by the 11th century, divided into shires, a division which still remains. Each shire was broken up into districts called hundreds (*q.v.*), or, in certain parts of the north, wapentakes or wards, and each hundred was divided into townships or vills, which corresponded to the later parish. The larger fortified towns were called boroughs, and it was there that such commerce as there was usually took place. Each shire had a governor, called the ealdorman, who not only governed the shire but commanded its military forces and presided in the shire court with the bishop. However, the everyday government of the shire was carried on by the ealdorman's deputy, who was also the king's representative. This officer was called the king's reeve, and later the shire reeve. Under his later name of sheriff (*q.v.*), he was to become for a time the principal governmental official in Norman times.

Gradually the sheriff took over all civil government from the eorl, who, in the late Anglo-Saxon period, often ruled over several shires. His two principal tasks were the maintenance of the peace and the collection of taxes, both difficult in the extreme. The shire court was the main centre of governmental and judicial activity and met two or three times a year to deal with the more important judicial business. There were also the hundred courts, which met once a month and were presided over by the sheriff or one of his deputies. The judges of the court were the freemen of the hundred, usually represented by 12 or 24 of the more substantial ones. Here may be one of the roots of the jury. The hundred court had jurisdiction over every kind of dispute, whether civil or criminal. The hundred court also supervised what was later called the frankpledge (*q.v.*) system, which was the primary means of keeping order. Every freeman had to be a member of a tithing, which was a group of ten men under a tithingman. Each member of the tithing could be made responsible for the misdeeds of any of the others, thus giving him a personal interest in the proper working of

the judicial system. The hundred and shire courts administered customary law and that promulgated in the dooms of the kings.

The main purpose of the law, indeed, was to control and mitigate personal revenge by providing instead, wherever possible, a practice of compensation by money payment. The basis of the legal system was thus the nature and amount of these money payments. At the root of everything was the wer (blood money), which defined in monetary terms the value of a man's life. If he were murdered, then the murderer would be liable to pay the wergild. The task of the court was to fix this price. The enforcement of its payment was left to the dead man's family, who might, if they were not paid, seek direct revenge, thus perhaps provoking a prolonged blood feud. This principle of compensation extended to all offenses.

The few written laws or dooms which have come down to us are mainly concerned with laying down the tariff of payments due for personal injuries and larcenies. Like the wer, the compensation or bot payable varied according to the social rank of both the victim and the wrongdoer. The only help which the community could give to assist the victim or his relatives in enforcing the law was that in some serious cases the wrongdoer might be declared an outlaw, in which case anyone might kill him with impunity.

The judicial process was slow and unreliable. It might be very difficult to get the accused before the court. Apart from the tithing system, an aid to this end was to make some local lord liable for the production in court of his man. This made it even more desirable to have a powerful lord to attest the respectability of his man. Inevitably the lord soon came to share in such reparation as his man might recover.

Even when the parties could be got to the court, there was the problem of proof. The Anglo-Saxon methods of proof generally consisted of direct appeals to the supernatural. The principal method was that of oath helpers. The party who was proving had to produce a number of witnesses (usually 12) who would swear that he was telling the truth. It was understood that the value of a man's oath might vary with his status (*i.e.*, with his wer) and so it might be necessary for a defendant to meet a charge by assembling oaths of a prescribed monetary value. This procedure, under its later names of "wager of law" and "compurgation," was to last for centuries, though it was open to the most flagrant abuse. In more serious cases, resort was had to the ordeal. Ordeal was generally by fire or water; ordeals by battle were as yet unknown, for they were a Norman importation.

In civil cases, proof was by various forms of oath. Claims to the ownership of land might in some cases involve the landboc or charter. To avoid litigation, sales of property were attended with the utmost publicity. Oaths of the parties and witnesses to the transaction might be required.

It is important to notice that there were already gaps in the system of compensation for wrongs by money payments. Some crimes were recognized as being so serious that they were "bootless"; *i.e.*, could not be compensated for. The most important of these was breach of the king's peace. The king's peace was later extended to cover all parts of the country at all times, and any crime constituted a breach of it. The list of bootless crimes was thus extended, the system of compensation disappeared, and the state acquired jurisdiction over all crimes. (*See* also ANGLO-SAXON LAW.)

The Norman Age.—There is much to support the view of German historians of the 19th century, who tended to regard English law as of Frankish derivation. During a long period English lawyers wrote in French and even thought in French, and to this day many of the technical terms of the law, more especially of the private law, are of French origin. When English law took shape in the 13th century, it was very like one of the *coutumes*, or bodies of customary law, of northern France; and, even apart from linguistic difficulties, the *Saxon Mirror* of Eike von Repgow (d. *c.* 1233) will seem far less familiar to an Englishman than the so-called *Establissements de St. Louis*. This French character presented by English law in the 13th century was the outcome of a slow process of change over two centuries and owed much to the reforming energy of Henry II (d. 1189), who, in addition to England, ruled a good half of France. William I seems to have intended to govern Englishmen by English law. There was much in the English system that he was anxious to retain, notably the system of land taxation and the prestige and authority of the crown. One of his motives for the compilation (1086) of Domesday Book (*q.v.*) was certainly the desire to fix upon every landholder, French or English, the amount of geld (tax) that had been due from his predecessor in title. Although William introduced military feudalism into England, he was well aware of its potentially disruptive tendencies and took stern measures to enforce direct obedience to himself. At Salisbury (1086) he took oaths of fealty and homage from all the mesne tenants as well as from his tenants in chief. The sheriffs, though called *vice-comites*, were to be and to remain the king's officers; the shire moots might be called county courts, but were not to be the courts of counts.

After the misgovernment and tyranny of William II, his brother Henry I promised a restoration of King Edward's law, that is, the law of the Confessor's time. Various attempts were then made, mostly it would seem by men of French birth, to state in a modern and practicable form the *laga Eadwardi* which was thus restored. The result of their labours is an intricate group of legal tracts, of which the best is known as the *Leges Henrici primi*. This work, planned as a comprehensive law book, derived most of its content from the code of Canute and from older English dooms, although the author shows some knowledge of the *Lex Ribuaria* and of the *Breviary* of Alaric. Neither the Conqueror nor his sons had issued many ordinances; the invading Normans had little, if any, written law to bring with them, and they had come to a country where the previous kings had been lawgivers.

Nevertheless, although there was little immediate radical change in the legal system after 1066, it is possible to say that the Norman conquest determined the whole future of English law. England had now acquired the strongest central government in Europe, and the effects of this were felt at once. There was an immediate tightening up of the enforcement of public order, a matter of special interest to the Normans, who were to begin with a small group of foreign conquerors in a large and potentially hostile native population. The frankpledge system was vigorously overhauled and the tithings brought up to strength. The duty of belonging to a tithing, and of joining in the "hue and cry" in pursuit of offenders, was impressed on each community. As an additional safeguard for the Normans, who were in daily danger of assassination, William introduced the system of "presentment of Englishry." Any murdered man was presumed to be a Norman, unless four freemen of the hundred could show that he was an Englishman. Unless the presumption was rebutted, the heavy *murdrum* fine was imposed on the community; hence the modern term murder.

Another immediate change was that French became the language of the upper classes; for legal documents it ultimately superseded Latin and was to remain the language of the law for centuries. The Normans brought with them a new system of judicial proof: trial by battle. At first only used between Normans, its popularity spread to the indigenous population, though it was generally invoked only by the feudal aristocracy.

These changes, however, were on the surface. More fundamental was the establishment of a well-organized feudal system and the creation of the *curia regis*. The *curia regis* ("royal court"), the Norman version of the Anglo-Saxon witan, was at first an all-purpose governmental body, the instrument of central government. It was composed of the great officers of state and church, and the king's personal advisory staff. Though in origin somewhat formless, this institution was to prove fertile in producing other government bodies. It is indeed the earliest ancestor of nearly all the organs of central government in the English constitution; and in particular, the three great common-law courts of the king's bench, the common pleas and the exchequer were originally branches of it. So was the chancery, where the writs which determined the all-important forms of action in the early common law were formulated and issued.

Just as vital as the *curia regis* was the establishment of feudalism, made to serve the purposes of a strong central government. As has been seen, some of the elements of feudalism were already

present in England before the Conquest, and the fundamental principle of a personal bond between lord and man was well known. The Normans, however, greatly elaborated and refined this basic social institution. William I was determined to reap the full reward for his victories. The source of wealth and power was land; he therefore maintained that all the land in England was in his gift. Since then, in theory at any rate, all English land has been ultimately owned by the crown. Most of it was granted out at once to William's followers as his tenants (or tenants in chief as the king's immediate tenants were called) in return for specified services; they in their turn granted their land out to lesser tenants, who made further grants, all in return for services. Thus, by this process of subinfeudation a feudal pyramid was established which was the skeleton of the social order.

Among the rights which usually went with being a landlord was the right to hold a court to deal with the affairs of tenants. Many of the new feudal courts now sprang up to take their place beside the old shire and hundred courts. The feudal courts were generally only entitled to deal with civil matters, but occasionally the king would grant a landlord the right to try criminal cases. His court then became a franchise court. The lowest unit of landholding was the manor; and the lord of the manor held his own court among his tenants. The great majority of these tenants were usually of unfree status—villeins—and where they were the suitors the manor court was called a customary court. If the lord had enough free tenants to provide judges for disputes between free tenants, his court was called a court baron (*q.v.*). Some of the manorial courts were given power to supervise the local frankpledge system; if so, they were called courts leet (*q.v.*). A great network of courts and jurisdictions grew up over England, with all sorts of local rules and customs. These were nearly all destined ultimately to be swept away by the tremendous centralization of royal justice under the common law. The office of justiciar (*q.v.*), or chief minister of the king, was a Norman institution which died out in the 13th century. (*See* also MANOR; SERFDOM; VILLEINAGE.)

Norman Feudal Land Law.—The system of land tenure upon which feudalism was based was so important to the early development of English law that it deserves special mention. As we have seen, the king and his tenants in chief were at the top of the landholding pyramid. Below them might be a succession of mesne lords, the lowest of whom, the tenant in demesne (who actually lived on the land), held the land he worked. The rules relating to tenure, or the terms on which the land was held, were standardized quite soon after the Conquest.

Each tenure had its own distinctive set of services and "incidents." Apart from a few miscellaneous tenures, the most important distinction was between free and unfree tenure. The unfree tenant, who was usually a villein, was virtually at the mercy of his lord; whereas the free tenant was protected by the law. The main types of free tenure were knight service (*q.v.*) or military tenure, which involved providing the lord with a number of armed soldiers; grand or petty serjeanty (*q.v.*), which was the duty of doing some specified personal service to the king; spiritual tenure, which might be either frankalmoign (no specific duties except to pray for the repose of the grantor's soul) or "divine service" (a specified number and type of spiritual services); and finally the commonest of all, socage (*q.v.*) tenure, where the services might be anything but were usually agricultural.

The great mark of free tenure was that services were always predetermined. In unfree tenure they were not, and the unfree tenant never knew what he might be called to do for his lord. Although at first the villein tenant held entirely at the will of the lord and might be ejected at any time, the royal courts later protected him to the extent that he held at the will of the lord and according to the custom of the manor, so that he could not be ejected in breach of the existing customs. Tenure in villeinage (as unfree tenure at first was generally called) then became known as tenure by copy of the court roll, or copyhold (*q.v.*) tenure. Many of the "incidents" of free and unfree tenure were the same in some places; an example is the incident of heriot, by which the lord was entitled on the tenant's death to his best beast or other chattel. There were also local customs which might apply to both kinds of tenure. For example, although the general rule of inheritance was primogeniture, yet in places, such as Kent, where the custom of gavelkind (*q.v.*) applied, all the sons of the deceased tenant shared his estate between them. The system of tenures gradually decayed as services were commuted into money payments. No new tenures could be created at all after the Statute of *Quia Emptores* (1290); and in the centuries that followed, the various types of tenure were one by one abolished, until in the early 1960s there were only two: socage (in a purely theoretical form) and the nonfeudal tenure of leasehold.

Tenure showed the terms on which land was held; the tenant's "estate" showed how long he was entitled to hold it. In the perfected feudal land law, there were three estates of freehold. Freehold is here contrasted with leasehold, which was not yet properly developed as a means of holding land. The estates of freehold were the fee simple, the fee tail and the life estate. The word fee denoted that the estate was inheritable and might theoretically last forever. The fee simple was and is the nearest possible to absolute ownership; on the tenant's death, the land descended to his heirs, and was freely alienable. The estate thus continued until there were no heirs of a deceased tenant, when it went back to the lord.

The fee tail or entail was a creation of the statute *De Donis Conditionalibus* (1285), and was an estate which continued for as long as the original tenant or any of his lineal descendants survived. Only lineal descendants, not just heirs, sufficed. When lineal descendants died out, the land reverted to the lord. The life estate, as its name indicates, lasted only for the life of the tenant, and then reverted to the lord. The estates of freehold all had certain features in common. They were all of uncertain duration. Only the owner of a freehold estate could be seised of the land, and thus have the right to recover it by a real action. They could originally only be conveyed to a new tenant by the formality of feoffment with livery of seisin, a ceremony in the presence of witnesses whereby the feoffee formally took possession of the land, and the feoffor solemnly gave up possession by handing him a piece of earth, or some such symbolic act. In contrast, the mere leaseholder had no seisin, and therefore no right to bring a real action for the recovery of the land.

Apart from criminal law, the early common law was primarily concerned with land law; and it was around these basic rules of tenure and estate that the first "forms of action" were developed to give remedies for their breach. The law of real property was for centuries the chief if not the only branch of the civil law. (*See* also REAL PROPERTY AND CONVEYANCING, LAWS OF.)

The Growth of Royal Justice.—The gulf that divides the *Leges Henrici primi* (*c.* 1115) from the texthook (*c.* 1188) ascribed to Ranulf de Glanville seems at first sight very wide. The one represents a not easily imaginable chaos and clash of old rules and new; it represents also a stage in the development of feudalism which in other countries is represented chiefly by a significant silence. The other is an orderly, rational book, which through all subsequent centuries has been readily understood by English lawyers. Making no attempt to tell us what goes on in the local courts, its author (who may have been Henry II's chief justiciar, Ranulf de Glanville, or Glanville's nephew, Hubert Walter, or Geoffrey FitzPeter) fixes attention on a novel element which was beginning to subdue all else to its powerful operation. Glanville wrote of the justice that was done by the king's own court. Henry II had opened the doors of his French-speaking court to the mass of his subjects. Judges chosen for their ability were to sit there, term after term; judges were to travel in circuits through the land, and in many cases the procedure by way of "an inquest of the country," which the Norman kings had used for the ascertainment of their fiscal rights and which was not unlike some practices familiar in the Anglo-Saxon age, was to be at the disposal of ordinary litigants. All this had been introduced in a piecemeal, experimental fashion by ordinances that were known as "assizes." There had not been, and was not to be, any enunciation of a general principle inviting all who were wronged to bring in their own words their complaints to the king's audience. The general prevalence of feudal justice, and of the old methods of supernatural probation (ordeals, battle, oaths sworn with oath helpers), was to be theo-

retically respected; but in exceptional cases, which would soon begin to devour the rule, a royal remedy was to be open to anyone who could frame his case within the compass of some carefully worded and prescript formula. With allusion to a remote stage in the history of Roman law, a stage of which Henry's advisers can have known little or nothing, we may say that a "formulary system" was established which was to preside over English law until modern times. Certain actions, each with a name of its own, were open to litigants. Each had its own formula set forth in its original (or originating) writ (*see* PRACTICE AND PROCEDURE); each had its own procedure and its appropriate mode of trial. The litigant chose his writ, his action, and then had to stand or fall by his choice. Writs had already been frequently used by the Anglo-Saxon kings, who had made some slight progress in adapting them to judicial use; but it is to the credit of the Norman kings that they should have laid the foundations of a skilful formulary system. Thus a book about royal justice tended to become, and Glanville's book already was, a commentary on original writs and, in a sense, a register of writs.

The precipitation of English law in so coherent a form as that which it had assumed in Glanville's book is not to be explained without reference to the revival of Roman jurisprudence in Italy. Out of a school of Lombard lawyers at Pavia had come Lanfranc, archbishop of Canterbury and the Conqueror's adviser, and the Lombardists had already been studying Justinian's *Institutes*. Then at length the *Digest* came to be studied. About the year 1100 Irnerius was teaching at Bologna, and by about 1149 Vacarius (*q.v.*) was teaching Roman law in England. He spent the rest of a long life there, and faculties of Roman and canon law took shape in the nascent University of Oxford. Whatever might be the fate of Roman law in England, there could be no doubt that the canon law, which was crystallizing in the *Decretum Gratiani* (*c.* 1139) and in the decretals of Alexander III, would be the law of the English ecclesiastical tribunals. The great quarrel between Henry II and Thomas Becket, archbishop of Canterbury, brought this system into collision with the temporal law of England. The result was the system of benefit of clergy, which was not abolished until 1827, nor its last vestiges removed until 1841 (*see* CLERGY, BENEFIT OF). The king's ministers must have seen that they had much to learn from the canon-law systems. There was healthy emulation of the work that was being done by Italian jurists, but no meek acceptance of foreign results.

The Age of Bracton.—By the beginning of the 13th century a great constructive era had opened, and its outcome was an important book, the *De legibus et consuetudinibus Angliae* ("The Laws and Customs of England") by Henry de Bracton (*q.v.*), who died in 1268, after having been for many years one of Henry III's justices. The model for its form was the treatise on Roman law of Azo of Bologna, from which were taken many of the generalities of jurisprudence, maxims that might be regarded as of universal and natural validity. But the true core of the work was the practice of the English central court (*coram rege*), which had yearly been extending its operations in many directions. For half a century past, diligent record had been kept on parchment of all that this court had done, and from its rolls Bracton cited numerous decisions. He cited them, when he saw need, as precedents, paying special heed to the judgments of two judges who were already dead, Martin Pateshull and William Raleigh; but he obviously did not feel in any sense bound by what he found in the rolls—he frankly tells us that the judges of his own day had not always gained his respect. For this purpose he compiled a large *Note Book*, which was discovered by Sir Paul Vinogradoff in the British museum in 1884. Thus at a very early time English common law showed a tendency to become what it afterward became; namely, case law.

The term common law (*q.v.*) was being taken over from the canonists by English lawyers, who used it, however, to distinguish the general law of the land from local customs, royal prerogatives and, in short, from all that was exceptional or special. Since statutes and ordinances were still rarities, all expressly enacted laws were also excluded from the English lawyers' notion of the common law. Magna Carta (*q.v.*; 1215) had taken the form of a grant of "liberties and privileges," comparable to the grants that the king made to individual men and favoured towns. Nonetheless, because of its importance and solemnity, it was in after ages regarded as the first article of a statute book. There it was followed by the Statute of Merton (1236) and the Statute of Marlborough (1267), issued after the end of the Barons' war.

During Henry III's long reign, however, the swift development of English law was due chiefly to new "original writs" and new "forms of action" devised by the chancery and sanctioned by the court. Already in the time of Glanville it had been possible to assemble a goodly collection of writs commonly used in the king's courts, and his book was in effect a commentary upon them; he seems to have had no doubt that the most authoritative statement of English law was to be found in the writs. Forty years later Henry III caused a manuscript register to be sent to Ireland (1227), for he too regarded such a register as the best introduction to English law; his son Edward I felt the same and enacted a register of the principal writs for introduction into Wales (Statute of Wales, 1284). About this time there were also registers for use in the Channel Islands and in Chester.

The Reign of Edward I.—The comparison that is suggested when Edward I is called the English Justinian cannot be pressed very far. Nevertheless, as is well known, it was in his reign (1272–1307) that English institutions finally took the forms they were to keep through coming centuries. The parliament of the three estates, the convocations of the clergy, the king's council, the chancery or secretarial department, the exchequer or financial department, the courts of king's bench and common pleas and the commissioners of assize and gaol delivery were already in existence. By that time there was also a small group of professionally learned judges, and a small group of professionally learned lawyers, whose skill was at the service of those who would employ them. Moreover, the statutes that were made in the first 18 years of the reign, though their bulk seems slight to us, were so fundamental in character that in subsequent ages they appeared as the substructure of huge masses of superincumbent law. Coke commented upon them sentence by sentence, and even now a student with the merest smattering of English law must profess some knowledge of *Quia Emptores* and *De Donis Conditionalibus*. If some U.S. states have, while others have not, accepted these statutes, that is a difference which is not unimportant to citizens of the United States in the 20th century.

In the new reign the novel statutes of Edward I occasioned a great increase in the number of writs; sometimes, indeed, a statute itself enacted the precise form of words to be used. Thus the control of the vital writ-making power passed from the crown to parliament. Great social changes were also taking place; laymen were placing clerics as judges of the king's courts and as the legal practitioners appearing before them. This change was reflected in the legal literature compiled for use in the humbler branches of the profession as well as at the top. The work of estate management, court-keeping, business correspondence and the multifarious duties of stewards and bailiffs fell into the hands of laymen who possibly knew a little Latin but who were much more at home in French. The clerical Latin of Bracton naturally placed his great book beside many other books of romanists and canonists, some of which he used and thus came into touch with the main stream of European legal culture. At the very end of his life the tide turned, and the Latin, Roman and clerical culture of Bracton was replaced by a culture which was insular, lay and French. Little tracts from around the year 1260 make this clear, and it is obvious that the future lay with *Fet Asaver*, a tract on procedure written in French, and with similar works. They seem closely connected with teaching—not with that of the universities but with the practical and vocational instruction which was to create the "common lawyer" even in the days before the inns of court (*see* INNS OF COURT AND CHANCERY) seem to have existed. Professional education of such a kind was centred upon practice and the study of forms (writs) and the verbal exchanges which in due course produced the pleadings which were enrolled in Latin, not by lawyers, but by the clerks of the courts. The altercations in French, designed to produce the required results upon the Latin plea roll, were often of great ingenuity. Teachers would give examples, and students

would collect newer and better ones from listening in court. At first these new cases would be inserted in copies of these tracts, but they soon came to form separate collections of cases.

Thus the first law reports that have descended to us come from the early years of Edward's reign; the oldest that have been printed belong to the years 1272–78. These were the precursors of the long series of Year Books (Edward II–Henry VIII) which runs through the residue of the middle ages. Lawyers were already making and preserving notes of the discussions that took place in court, French notes that would be more useful to them than the formal Latin records inscribed upon the plea rolls. From these reports it is clear that there were already, in modern terms, a few "leading counsel," some of whom would be retained in almost every important cause. Meanwhile, the church was becoming anxious to withdraw the clergy from secular employment. The clerical element had been strong among the judges of Henry III's reign: Bracton was an archdeacon, Pateshull a dean, Raleigh died a bishop. Their places were beginning to be filled by men who were not in orders but who had pleaded the king's causes for him—his serjeants or servants at law—and beside them there were young "apprentices at law," who were learning to plead. Also, certain "attorneys at law" were making it their business to appear on behalf of litigants. The history of the legal profession and its monopoly of legal aid is intricate, and at some points still obscure; but the influence of the canonical system is evident: the English attorney corresponded to the canonical proctor (*q.v.*), and the English barrister to the canonical advocate (*q.v.*). The main outlines were being drawn in Edward I's day; the legal profession became organic, and professional opinion became one of the main forces that molded the law.

The Legislation of Edward I.—The profession, therefore, already had its practitioners' books—a register of writs and a volume of statutes; before the reign closed it was beginning to create the earliest Year Books, and from the little books of pleading which already existed it was possible to see how much could be done with these homemade tools. The crown had already shown by its use of legislation how much could be accomplished out of purely native resources. From Magna Carta (1215, 1225) downward an immense number of details concerning feudalism and tenures had been settled by statute, so that it became clear that the crown could profoundly influence the course of legal development. The first 20 years of Edward I's reign produced the greatest outburst of statute law that was to be seen until the 16th century. The first Statute of Westminster (1275) had much to say of miscellaneous matters (mostly but not entirely criminal); in the next few years the Statute of Gloucester (1278) tried (like Edward's legislation generally) to protect the spirit and main principles of feudalism against the determined attempts to evade or misuse it by unscrupulous persons, be they husbands or wives, lords or tenants, clerics or laymen, particularly tenants or reversioners. The procedure of the courts was at the same time revised in the hope of curing its defects. The second Statute of Westminster (1285), ch. 1 (which is often called, from its opening words, *De Donis Conditionalibus*), is typical of much of Edward I's legislation. As a matter of pure practice, and without much legislation, it had become usual for landowners to make gifts to their daughters on the occasion of their marriage, gifts which were primarily intended to provide for the young couple and their prospective family. The recent habit of the court of interpreting these gifts as conditional, and regarding the condition as fulfilled as soon as issue was born, had the result that restrictions ended at the moment of the birth of issue, the donees acquired an unrestricted fee simple, the legal protection of the infants ceased and, if the new family eventually failed, the reversion, which settlors had hoped would take place when the gift no longer fulfilled its purpose, was often no longer possible. The statute *De Donis*, in spite of manifest defects of draftsmanship, did succeed in establishing the entail at least until the next century, when elaborate means were devised for evading it. Its immediate result was to create some doubts and immense masses of technicalities. The statute itself mentions, and extends its protection to, other schemes of settlement besides the "frank marriage" just described, for it was often desirable to make some provision for younger sons (now that primogeniture had become the rule); this became possible through the entail, which could be made to cease when the failure of descendants made further provision for them unnecessary. Thus the results achieved by the "appanage" system of royal and noble houses in Europe could be achieved in England by the entail, which in the prescribed circumstances would return to the parent stock which had first created it.

The statute called *Quia Emptores* (1290) was the first great conveyancing act. Its primary object was to check certain practices which had been devised in order to avoid feudal obligations to an overlord. Its policy was to forbid subinfeudation, and in this it succeeded. Its unexpected results, however, were momentous: it was held that as a result no new manor could be created and the class of free tenants of a manor could not be replenished. The immediate object of the Statute of Mortmain (1279) was of the same nature, since a lord who had an ecclesiastical corporation for tenants could obviously receive no wardships, marriages, reliefs or escheats from a tenant who never died or married and was unable to commit felony. This statute was indeed part of a dispute between Edward I and the archbishop of Canterbury, but it was not primarily an anticlerical move by the king; like the rest of his statutes its main purpose was to protect feudalism from the ingenious machinations of tenants. Municipal corporations became subject to the mortmain laws in 1391 as also did uses, churchyards, etc.

Besides the clergy, there were other classes who were outside the common law or in other respects had a distinctive status, such as the Jews (expelled in 1290) and the merchants. These latter received close legislative attention from Edward I in the Statute of Acton Burnell (1283), drastically revised by the Statute of Merchants (1285). The course of legal procedure was constantly being amended in his legislation, and in the times of his successors the sheer bulk and practical importance of his statutory reforms seemed to suggest that it would be hard to place any limit to legislation: crisis after crisis seemed to lead up to yet another statute. The civil conflict under Edward II produced first the "ordinances" (1311) and then the Statute of York (1322); the social upheaval which accompanied the Black Death also produced the Statute of Labourers (1351), and ten years later the justices of the peace were firmly established. At about the same time, the old conflict of church and state produced the first of several statutes of provisors and praemunire (*q.v.*), while in 1352 came the famous Statute of Treasons.

An obvious criticism of English legislation during the middle ages was its reluctance to abolish anything; if some abuse needed correction, the favourite course was to add a new procedure which it was hoped would correct the mischiefs produced by earlier tradition or legislation. In the end, vast piles of statutory or judicial reforms eventually defeated themselves. As legislation became a parliamentary process concerned with the deepest mysteries of the common law, the feeling grew that it might be more prudent to leave it to the experts—the judges and serjeants in the courts. Thus the law grew even more technical and incognizable to Englishmen in general, and even more isolated from European learning as students of the common law applied all their efforts to the Year Books, whose technical pages alone could help them to interpret aright the masses of statute law which in turn were being overlaid with masses of case law.

The study of English law thus fell apart from all other studies, and by the beginning of the 14th century the impulse that had flowed from Italian jurisprudence was ebbing. There are but two comprehensive textbooks from Edward's reign: the one known as *Fleta*, the other as *Britton* (*q.v.*); both quarried their materials from Bracton's treatise. There also survive two little books on procedure, attributed to Chief Justice Ralph Hengham, and a few other small tracts of an intensely practical kind. *Britton*, being in French, was much used, especially since it embodied the statutes of Edward I; Hengham's work, however, was in Latin and never gained a public, except for one part of it which circulated separately under the name of Walter of Henley's *Husbandry*—in a French version. Under the cover of fables about King Alfred, the author of the *Mirror of Justices* made a bitter attack upon King Edward's

judges, some of whom had fallen into deep disgrace. English legal history has hardly yet been purged of the leaven of falsehood that was introduced later by Coke's undue reliance on this fantastic and unscrupulous pamphleteer. The *Mirror* ended that literate age which began with Glanville's treatise. Between Edward I's day and Edward IV's, hardly anything that deserves to be called a book was written by an English lawyer.

The 14th and 15th Centuries.—During the later middle ages the body of statute law was growing, but not very rapidly. Acts of parliament intervened at a sufficient number of important points to generate and maintain a persuasion that no limit, or no ascertainable limit, could be set to the legislative power of king and parliament. There are very few signs that the judges ever permitted the validity of a statute to be drawn into debate. Thus the way was being prepared for the definite assertion of parliamentary omnicompetence later made by the Elizabethan statesman, Sir Thomas Smith, and for those theories of sovereignty associated with the names of Thomas Hobbes and John Austin. Nevertheless, English law was being developed rather by debates in court than by open legislation. The most distinctively English of English institutions of the later middle ages were the Year Books and the inns of court. Year by year, term by term, lawyers were reporting cases in order that they and their fellows might know how such cases had been decided. The allegation of specific precedents was indeed much rarer than it afterward became, and no calculus of authority as definite as that which exists in modern times had been established in Coke's day, far less in Littleton's. Still, it was by a perusal of reported cases that a man would learn the law of England. A skeleton for the law was provided, not by the Roman rubrics (such as public and private, real and personal, possessory and proprietary, contract and delict), but by the cycle of original writs that were inscribed in the chancery's *Registrum Brevium*. A new form of action could not be introduced without the authority of parliament, and the growth of the law took the shape of an explication of the true intent of ancient formulas. Times of inventive liberality alternated with times of cautious and captious conservatism. Coke could look back to Edward III's day as a golden age of good pleading. The otherwise miserable time of the Wars of the Roses produced some famous lawyers and some bold doctrines which broke new ground. It also produced Littleton's *Treatise on Tenures*, an excellent statement of law in exquisitely simple language. Meanwhile, English law was being taught scholastically.

Now that the heroic age of new writs formed by enterprising chancellors or by great remedial statutes had come to an end, other means had to be devised to meet the demands for law that would satisfy the needs of the newer age. The old "real actions" for land had from the first been handled freely by the courts, which in the end had created, mainly by decision and practice, a complicated and closely knit system which in effect covered the principal needs of the landowning classes. The question of personal actions was rather different. By this time there was little help to be had from chancellors or from parliament, and indeed some of the personal actions had already been directed into undesirable channels by judges and lawyers who still thought in terms of land rather than of trade and commerce. Thus covenant (which might have been developed into a much-needed action on contract generally) came to be annexed by the property lawyers; the action of debt also had obvious possibilities, and to some extent the courts encouraged new ideas in it, but they never completely freed it from the age-old wager of law or compurgation—timidly, they left this hoary survival until chancery treated it for what it was, a denial of that justice which the subject had a right to expect from a rational legal system and which chancery would supply if the common law persistently refused it. In the end it was the action of trespass which proved to be most promising. From its early days it was strongly tinged with criminal elements and thus had a swift and stringent procedure, as compared with the real actions, which were slow; in its simplest form it called upon the defendant to show why he had assaulted the plaintiff, gone off with his goods or invaded his land. Those, no doubt, were the commonest kinds of trepass which vexed the countryside, but there were many others demanding redress, and plaintiffs were allowed to call upon defendants to show "why" ("whereas . . ."). The clause beginning "whereas . . ." would set out the circumstances in which the defendant had caused damage to the plaintiff. The scope of the writ of trespass was thus greatly enlarged, and it became a writ of "trespass on the case" or, more slightly, an action of "case." As such clauses increased in variety, a great many more torts became actionable. But the greatest exploit of the action of "case" was not in the field of tort (*q.v.*) but of contract. It was already necessary in certain types of case to say, and to prove, that the defendant had undertaken (*assumpsit super se*) to produce a certain result; what he in fact did only became tortious because he had previously undertaken to act otherwise. The underlying theory was still that of tort; but the ground changes when it is said that the defendant, being indebted, undertook to pay a debt (*indebitatus assumpsit*). The 14th and 15th centuries were much occupied with the development of assumpsit; the 16th developed, from a minute study of the pleadings of *indebitatus assumpsit*, the doctrine of consideration, which is still a remarkable feature of the English law of contract (*q.v.*).

The Growth of Equity.—A danger then threatened: the danger that a prematurely osseous system of common law would be overwhelmed by summary justice and royal equity. Even when courts for all ordinary causes had been established, a reserve of residuary justice remained with the king. Whatever lawyers and even parliaments might say, it was seen to be desirable that the king in council should, with little regard for form, punish offenders who could break through the meshes of a tardy procedure and should redress wrongs which corrupt and timid juries would leave unrighted. Papal edicts against heretics had made familiar to all men the notion that a judge should at times proceed summarily. And so extraordinary justice of a penal kind was done by the king's council upon misdemeanants, and extraordinary justice of a civil kind was ministered by the king's chancellor (who was the specially learned member of the council) to those who "for the love of God and in the way of charity" craved his powerful assistance. It is now well established that the chancellors started upon this course, not with any desire to introduce rules of equity which should supplement, or perhaps supplant, the rules of law, but for the purpose of driving the law through those accidental impediments which sometimes unfortunately beset its due course. The wrongs that the chancellor redressed were often wrongs of the simplest and most brutal kind; assaults, batteries and forcible dispossessions. However, he was warned off this field of activity by parliament; the danger to law, to lawyers, to trial by jury, was evident.

But just when this was happening, a new field was being opened for him by the growing practice of conveying land to trustees. In England, when the forms of action were crystallizing, this practice had not been common enough to obtain the protection of a writ; but in the 14th century many cases conspired to make it common, and with the general approval of lawyers and laity the chancellors began to enforce by summary process against the trustee the duty that lay upon his conscience. In the next century it was clear that England had acquired a new civil tribunal. Negatively, its competence was defined by the rule that when the common law offered a remedy, the chancellor was not to intervene. Positively, his power was conceived as that of doing what "good conscience" required, more especially in cases of "fraud, accident or breach of confidence." His procedure was the summary, the heresy-suppressing (not the ordinary and solemn) procedure of an ecclesiastical court; but there are few signs that he borrowed any substantive rules from legist or decretist and many proofs that within the new field of trust he pursued the ideas of the common law. It was long, however, before lawyers made a habit of reporting his decisions. He was not supposed to be tightly bound by precedent. Adaptability was of the essence of the justice that he did. (*See* also LORD HIGH CHANCELLOR; EQUITY.)

The Tudor Age.—The 16th century was a time of danger for English common law. After the lawlessness and the political and administrative chaos of the Wars of the Roses, strong "governance" was both necessary and popular; but it presented some threat to the parliamentary system and to the preservation and growth of

the common law. New courts appropriated large fields of justice and proceeded summarily; the Star Chamber (*q.v.*), the chancery, the courts of requests, of wards, of augmentations, the Councils of the North and of Wales and, a little later, the court of high commission (*q.v.*). Some of these courts used torture to obtain evidence. Under Henry VIII the stream of law reports become intermittent; few judges of his or his son's reign left names that are to be remembered. In an age of humanism and of profound scholarship, the best works of English lawyers were alphabetically arranged abridgments of medieval cases, such as those made by Anthony Fitzherbert (d. 1538) and Robert Brooke (d. 1558).

At this time there was a widespread reception of Roman law in Germany; the same thing might well have happened in England, had a decisive supercession of English by foreign law seemed attractive to Henry VIII. As it was, the academic study of canon law was prohibited; regius professorships of civil law were founded and civilians were to sit as judges in the ecclesiastical courts. But the vast constitutional changes of Henry's reign, although created to further the policies of an imperious ruler, in fact served to stress the sovereignty of parliament. The omnicompetence of an act of parliament was more plainly seen when it was used to settle the succession to the throne, to annul royal marriages, to cancel royal debts, to define religious creeds, to alter the entire legal status of the church and to lend the force of statute to the king's proclamations. Meanwhile, the courts of common law worked only in obscurity, for jurors feared fines, and matters of state were reserved for council or star chamber. The inns of court were allowed to continue; their moots and readings did not hinder royal policy.

The Age of Coke and Selden.—The second half of the 16th century brought a marvelous resuscitation of English medieval law, described by F. W. Maitland (*English Law and the Renaissance*, 1901) as "a little Renaissance of our own, a Gothic revival." It was foreshadowed by the commentaries of Edmund Plowden (d. 1585), who reported cases at length and lovingly. Bracton's great book was printed (1569) and some Anglo-Saxon laws were brought to light and considered. The new national pride that animated Elizabethan England issued in boasts touching the antiquity, humanity and enlightenment of English law. Resuming the strain of Sir John Fortescue's *De laudibus legum Angliae* ("Praises of the Laws of England," *c.* 1470), Sir Thomas Smith (d. 1577), himself a civilian, wrote a book on the commonwealth of England that claimed the attention of foreigners for its law and its polity. Finally, the common law acquired an unrivaled champion in the person of Edward Coke (*q.v.*; 1552–1634). With an enthusiastic love of English tradition, he ranged over nearly the whole field of law, commenting, reporting, arguing, deciding—disorderly, pedantic, masterful, an incarnate national dogmatism tenacious of continuous life. Imbued with this spirit, the lawyers fought the constitutional battle against James I and Charles I; historical research was employed by both parties in the struggle. For the modern student of medieval law, many of the reported cases of the Stuart era are storehouses of valuable material, since the lawyers of the 17th century were eager students of records. William Prynne (d. 1669), the fanatical Puritan, published ancient documents with fervid zeal, and made possible the writing of a history of parliament. John Selden (*q.v.*; d. 1654) was in all Europe among the very first to write legal history as it should be written. His book about tithes is still a model and a masterpiece. When this accomplished scholar had declared that he had laboured to make himself worthy to be called a common lawyer, it could no longer be said that the common lawyers were "the least learned of learned men." Even pliant judges, whose tenure of office depended (until 1689) on the king's will, were compelled to cite and discuss old precedents before they could give judgment for their master, and even at their most subservient they would take pains not openly to break with medieval tradition.

The 18th Century.—By the mid-17th century the current of legal doctrine was so strong and voluminous that such events as the Civil War, the Restoration and the revolution of 1688 hardly deflected its course. The statute book of Charles II was enriched by leaves excerpted from the acts of the lord protector, Oliver Cromwell; and Matthew Hale (*q.v.*), who was, perhaps, the last of the great record-searching judges, sketched a map of English law which Blackstone was to colour. Then a time of self-complacency came for the law, which knew itself to be the perfection of wisdom, and any proposal for drastic legislation would have worn the garb discredited by the tyranny of Cromwell. The need after 1689 for the yearly renewal of the Mutiny act secured an annual session of parliament. The mass of the statute law made in the 18th century is enormous; but, even when such acts as are technically called private are excluded, the residuary matter bears a wonderfully empirical, partial and minutely particularizing character, and in modern times the legal practitioner is likely to know less about the statutes of the 18th century than he knows about the statutes of Edward I, Henry VIII and Elizabeth I. Parliament, it should be remembered, was endeavouring to govern the nation directly. There was little that resembled the modern permanent civil service. The choice lay between direct parliamentary government and royal prerogative; and lengthy statutes did much of that work of detail which would now be done by virtue of the powers that are delegated to ministers and governmental boards. Moreover, extreme and verbose particularity was required in statutes, for judges were loath to admit that the common law was capable of amendment. A vague doctrine, inherited from Coke, taught that statutes might be so unreasonable as to be null, and any political theory that seemed to reflect the stress laid upon sovereignty by Thomas Hobbes (*q.v.*) would have been regarded with not unjust suspicion. But the doctrine in question never took tangible shape, and enough could be done to protect the common law by a niggardly exposition of every legislating word. It is to be remembered that some main features of English public law were attracting the admiration of enlightened Europe. When Voltaire and Montesquieu applauded, the English lawyer had cause for complacency.

The common law was by no means stagnant. Many rules which came to the front in the 18th century can hardly be traced before that time. Especially is this the case in the province of mercantile law, where the earl of Mansfield's (d. 1793) long presidency over the king's bench marked an epoch. It is too often forgotten that until Elizabeth I's reign England was a thoroughly rustic kingdom and that much of its trade was in the hands of foreigners. Also, in medieval fairs the assembled merchants declared their own "law merchant" (*q.v.*), which was considered to have a supernational validity. It is late in the reports of the common-law courts before some mercantile usages appear which can be traced far back in the statutes of Italian cities. Even on the basis of the excessively elaborated land law—a basis which Coke's commentary on Littleton seemed to have settled forever—a lofty and ingenious superstructure could be reared. One after another, delicate devices were invented for the accommodation of new wants within the law; but the subtlety thus displayed was artificial and can be admired only if it is realized that the old law could not be frankly abolished. As to procedure, it had become a maze of evasive fictions, to which only a few learned men held the historical clue. By fiction the courts had stolen business from each other, and by fiction a few comparatively speedy forms of action were set to tasks for which they were not originally framed. On the other hand, that healthy and vigorous institution, the commission of the peace, with a long history behind it, was giving an important share in the administration of justice to numerous country gentlemen who were thus compelled to learn some law (*see* JUSTICE OF THE PEACE). A like beneficial work was being done among jurors, who had become "judges of fact." No one doubted that trial by jury was the palladium of English liberties, and popularity awaited those who would exalt the office of the jurors and narrowly limit the powers of the judge.

Developments in Equity.—Meanwhile, the chief addition to English jurisprudence was being made by the crystallization of the chancellor's equity. In the 17th century the chancery only narrowly escaped abolition, a fate which befell the Star Chamber, while the court of requests (*q.v.*) perished under the persistent attacks of the common lawyers. Having outlived troubles, the chancery took to orderly habits and administered under the name of "equity" a growing group of rules, which in fact were supplemental

law. Stages in this process are marked by the chancellorships of Heneage Finch (1673–82), later earl of Nottingham, and Philip Yorke (1737–56), later earl of Hardwicke. Slowly, a continuous series of equity reports began to flow, and still more slowly an "equity bar" began to form itself. The principal outlines of equity were drawn by men who were steeped in the common law. By way of ornament a Roman maxim might be borrowed from a French or Dutch expositor, or a phrase which smacked of that natural-law school which was dominating continental Europe; but the influence exercised by Roman law upon English equity has been the subject of gross exaggeration. Parliament and the old courts being what they were, perhaps it was only in a new court that the requisite new law could be evolved. The result was not altogether satisfactory. Freed from contact with the plain man in the jury box, the chancellors were tempted to forget how plain and rough good law should be, and to screw up the legal standard of reasonable conduct to a height hardly attainable except by those whose purses could command the constant advice of a family solicitor. A court which started with the idea of doing summary justice for the poor became a court which did a highly refined, but tardy justice, suitable only to the rich.

Blackstone.—About the middle of the 18th century William Blackstone (*q.v.*), then a disappointed barrister, began to give lectures on English law at Oxford (1753), was elected Vinerian professor of law (1758), and soon afterward began to publish (1765) his *Commentaries*. Accurate enough in its history and doctrine to be an invaluable guide to professional students and a useful aid to practitioners, his book set before the unprofessional public an artistic picture of the laws of England such as had never been drawn of any similar system. No nation but the English had so eminently readable a law book, and it must be doubtful whether any other lawyer ever did more important work than was done by the first professor of English law. Over and over again the *Commentaries* were edited, sometimes by distinguished men, and it is hardly too much to say that for nearly a century the English lawyer's main ideas of the organization and articulation of the body of English law were controlled by Blackstone.

The Export of English Law.—The great territorial expansion of the English-speaking world from the 17th century onward was accompanied by the export of English law into new conditions and new societies. English settlers carried the common law with them to America, at least in its general principles; in practice, however, they were free to amend and supersede it as they wished. The extent to which English law was adopted varied greatly from colony to colony. The different social and economic conditions obviously made for the growth of a different type of system, though based on some of the fundamental principles of English law. The most notable difference in American law was the strong tendency toward codification, which manifested itself at a very early stage. Codes were desirable to meet the different conditions of life and the shortage of lawyers trained in English law, parts of which were anyway irrelevant to American problems. In most states, however, the common law was accepted as a subsidiary law to whatever codes might be enacted. "The common law of England is not to be taken in all respects to be that of America. Our ancestors brought with them its general principles and claimed it as their birthright; but they brought with them and adopted only that portion which was applicable to their condition" (Justice Joseph Story, in *Van Ness* v. *Pacard*, 2 Pet. 137 [1829]). However, in the general sense that the colonists used the language and methods of English law, they retained the English heritage. As a general system of legal principles, the common law was adopted by the states after the revolution of 1688 and in most of the new states thereafter, though no specific rule of the common law became law in the United States until so determined by legislation or judicial decision. Both before and after the revolution, there was a reception of English law in the United States; after 1688, additional account was taken of the new aspects of English law. "In reality the common law of each state rests on the English common law as a foundation from which has been built up a system of common law for each state similar in most respects to the present English law and the law of the other states, but differing in many details, each having a separate existence, though developed in the same way and subject to the same general principles as the present common law of England" (W. F. Walsh, *History of Anglo-American Law*, 2nd ed. [1932]). Similarly, in English law American decisions have always been accepted as persuasive, though not of course binding, authority. (*See* AMERICAN LAW.)

The same process has been repeated with variations in nearly all the countries where the English have settled or governed. Where territories were acquired by outright conquest, however, the existing laws were often allowed to remain. Thus French Canada kept French law; and South Africa and Ceylon still retain Roman-Dutch law. On the other hand, Canada, Australia, New Zealand and India are all "common-law" countries. They have each built up their own common law upon the foundation of English legal principles and, in many cases, of specific rules of English law. The action and interaction of legal development in all these countries has greatly contributed to the subsequent development of English law.

The Age of Bentham.—By the third decade of the 19th century long wars, vast economic changes and a conservatism generated by the French Revolution had piled up monstrous arrears of work for the British legislature. Meanwhile, Jeremy Bentham (*q.v.*), by his astringent criticism and derision of the system that Blackstone had complacently expounded, provided the impulse toward a much-needed radical reform. The legal system was encumbered by a mass of noxious rubbish, the useless leavings of the middle ages; Bentham's persistent attacks upon it could not be countered, and at length a general interest in law reform was excited. The reforms of the penal code introduced by Sir Samuel Romilly and of the legal system effected by Lord Brougham were inspired by Bentham. The changes in constitutional law made at the time of the Reform act of 1832 were accompanied by many other measures, which purged the private, procedural and criminal law of much, though hardly enough, of the medieval dross. Some credit for rousing an interest in law, in definitions of legal terms and in schemes of codification is due to John Austin (*q.v.*), who was regarded as the jurist of the reforming and utilitarian group. But, though he was at times an acute dissector of confused thought, he was too ignorant of the English, the Roman and every other system of law to make any considerable addition to the sum of knowledge. The obsolescence of the map of law which Blackstone had inherited from Hale, and in which many outlines were drawn by medieval formulas, left 19th-century English lawyers without a guide, and for a while they showed interest in what seemed the voice of cosmopolitan science. Later the revived study of Germanic law in Germany gave them a fresh confidence in the value of their own traditional institutions. (F. W. M.; T. F. T. P.; T. S. Le.)

The Reform of Judicial Machinery.—From 1850 onward there has been a drastic reorganization of judicial machinery. In 1857 the jurisdiction of the ecclesiastical courts in matrimonial matters was transferred to a civil divorce court, and divorce, which, as opposed to judicial separation, had been confined to the minority rich enough to promote a private parliamentary bill, became available to persons of moderate means. In the same year the ecclesiastical courts surrendered the remnants of their jurisdiction over wills and intestacies to a newly formed probate court. In 1873–75 the Judicature acts reformed the civil courts on a pattern which has not since been basically altered. The probate and divorce courts, together with the court of admiralty, the only other English tribunal administering a system not based on native law, were merged in the probate, divorce and admiralty division of the high court of justice. The other divisions of the high court were, after 1881 (when a short-lived common pleas division and exchequer division were abolished), the queen's bench division and the chancery division. Above the high court was placed the court of appeal and, as an afterthought, the result of a change of government, the house of lords sitting in its judicial capacity as a court of final appeal. The Judicature acts were, however, much more than a regrouping and renaming of courts; they attempted a fusion of law and equity by making available legal and equitable remedies in all divisions of the high court and by providing that in case of conflict the equitable rule should prevail. Common law and equity

nevertheless preserved their separate identities, partly because of the different subject matter with which they often deal and partly because of the persistence among lawyers of a common law and a chancery outlook; judges who apply equitable notions to fields traditionally preserved for the common law are apt to be regarded as revolutionaries by the legal profession as a whole.

Except in divorce, these changes affected only those with sufficient property to risk the costs of an action in the superior courts. The poorer litigant was better served by the county courts set up in 1846, an innovation, or rather partial return to the decentralized justice of an earlier age, which has proved remarkably successful. The jurisdiction of these courts has been gradually extended; in particular, they are concerned with much legislation of great social importance, such as the Rent Restriction acts. The litigant who had a good reason but not the means to bring a case in the high court or was even too poor or ignorant to sue in the county court remained, however, at a disadvantage until the Legal Aid and Advice act of 1949. The purpose of this act was to provide a comprehensive service of legal advice and of assistance in litigation, with payment adjusted according to the means of the litigant. The full scheme envisaged by the act was postponed for reasons of economy but can be introduced under regulations to be made by the lord chancellor. (*See* LEGAL AID.)

The machinery of criminal justice was also improved in the course of the 19th century, although the basic hierarchy of courts, petty sessions, quarter sessions and assizes, remained unaffected. The most important change was in the character of the preliminary hearing by the magistrates of offenses to be tried on indictment at quarter sessions or assizes. In the 18th and early 19th centuries, when police and judicial functions had not been clearly differentiated, such hearings were often inquisitorial, but the Indictable Offenses act, 1848, confirming a growing practice, required the magistrates to be strictly judicial, to hear the evidence against the accused and on it alone to decide whether a prima facie case had been made out against him. It was not, however, until 1898 that an accused person was allowed to give evidence on his own behalf and only in 1907 that, by the institution of the court of criminal appeal, an adequate system of appeals was provided from decisions of quarter sessions and assizes. Another badly needed reform came in 1907 with the provision of legal aid for poor persons in indictable offenses, a principle which was not extended to preliminary hearings or summary trial before the magistrates until 1930. A remarkable feature of English criminal justice, as compared with most European systems, has been the continuing importance of the lay justice of the peace, in spite of the appointment of paid, legally trained magistrates in London and some of the larger cities, of barristers as recorders at borough quarter sessions, and of legally qualified chairmen at county quarter sessions. An important aspect of the magistrates' work has been their jurisdiction over young offenders, for whom special juvenile courts were first set up in 1908 (*see* CHILDREN'S COURT). The report of a royal commission on justices of the peace in 1948 strongly defended the lay justice against some public criticism; its cautious recommendations as to the appointment of justices and as to the organization of their courts were largely put into effect by the Justices of the Peace act, 1949, and the Magistrates' Courts act, 1952.

Changes in Constitutional Law.—The gradual introduction between the Reform act, 1832, and the Representation of the People (Equal Franchise) act, 1928, of a universal franchise for adult men and women has brought about great political changes but has not altered the basic legal structure of the constitution. The "queen in parliament" is the supreme sovereign authority, but the legislative powers of the house of lords have been curtailed by the Parliament act, 1911, which prevented it from vetoing money bills and left it only the power to suspend other legislation for a maximum of two years, a period reduced in 1949 to one year. The complexity of modern government and the pressure of two world wars have greatly increased the powers entrusted to the executive under broadly framed statutes, such as the Defence of the Realm acts, 1914–15, the Emergency Powers (Defence) act, 1939, and much town planning and social legislation. Acts done and regulations made under these powers have often been subject to the exclusive jurisdiction of special tribunals, with or without appeal to a minister, a challenge which the ordinary law courts have met by resort to the prerogative writs, originally available to the crown to compel royal officers to perform their duties or to prevent abuse or usurpation of the crown's authority; now called prerogative orders and with much simplified procedure, they have been widely granted at the suit of the subject.

According to A. V. Dicey's classic treatment of the rule of law, there is no administrative law in England, all officials being ultimately answerable in the ordinary courts. It is perfectly true that officials are still answerable in the ordinary courts; and the Crown Proceedings act, 1947, now allows the crown itself to be sued, so that its vicarious liability for the torts of its servants can now be fully enforced. This, however, is not enough. The enormous expansion in the scope of governmental activity which has been experienced in the 20th century has involved the creation of many new government departments with wide discretionary powers. Parliament has not usually allowed the courts to control these discretionary powers of government officials; indeed, the jurisdiction of the courts has often been expressly excluded by the enabling statute. Instead, a large number of administrative tribunals have appeared, whose decisions are made not so much under law as by government policy. Sometimes their decisions can be reviewed by the courts; more often they cannot. Within this preserve, decisions of great importance to the subject may be made with little or no publicity and on no apparent principle. Examples can be seen in the fields of housing, town and country planning, the national health service, pensions, social security, education, monopolies, restrictive practices, hire purchase (installment buying) and others.

The courts have waged a long struggle to control the exercise of governmental discretion and the decisions of administrative tribunals, and they have received help from parliament in the Tribunals and Inquiries act, 1958, and the establishment of a council on tribunals. There is still however, a need for realistic appreciation of the extent and importance of administrative acts and for an appropriate machinery to control them. The various solutions to this fundamental problem which have been adopted in continental Europe (some of which have been relatively successful) are alien to the English legal and governmental tradition, and have not found favour. But the task remains; and it poses a stern challenge to English lawyers. (*See* also CONSTITUTION AND CONSTITUTIONAL LAW.)

Codification of Criminal Law.—A considerable part of the criminal law has been reduced to statutory form (*e.g.*, Offences Against the Person act, 1861; Larceny act, 1916), but too often technical distinctions have been preserved which are only explicable as devices to avoid the savage penalties of an earlier age. On the other hand, the law relating to criminal punishment was steadily modified in response to changing conceptions of criminal responsibility, although it must be admitted that this development was inspired by prison administrators and other social reformers rather than by the legal profession as such. The culmination of this process was to be seen in the Criminal Justice act, 1948; this act abolished penal servitude, hard labour and corporal punishment (except for prison offenses); it restricted the power of the courts to order the imprisonment, and encouraged the committal to Borstal, of persons under 21; it extended the authority first given in 1908 to commit habitual criminals to long periods of preventive detention, and it widened the scope of probation as introduced by the Probation of Offenders act of 1907. Half a century later (Homicide act, 1957) the Scottish doctrine of "diminished responsibility" was introduced into English law, and the death penalty was reserved for "capital murders" as there defined.

Changes in Private Law.—No branch of the law was more ripe for reform than the land law; it carried over to the 20th century many institutions which could only be explained by reference to the peculiar position of land in feudal society. The combined effect of a series of acts in 1925 was: (1) to assimilate many of the rules affecting land and movable property; thus, a uniform system of intestate succession was applied to all kinds of property,

and primogeniture abolished, future interests whether in land or movables can only be created as equitable interests and all the property of a deceased person is administered in the same way. (2) To facilitate and cheapen the transfer of land; the purchaser for value of the legal estate (which since 1925 can only be a fee-simple absolute in possession or a term of years absolute) is no longer concerned with other estates, which are merely equitable, the owners of such equitable estates transferring their interest to the purchase money; the purchaser is still bound by third-party rights (such as a covenant not to build) but the task of discovering these is made easier by a system of registration of land charges. (3) To abolish certain feudal anachronisms, such as copyhold tenure. In spite of these changes, land law remains extremely complicated and conveyancing an expensive process, largely owing to the failure to apply throughout the country, a system of registration of title to land.

Some other branches of private law have been given statutory form. Company law, in contrast with property law, which, notwithstanding the legislation of 1925, is in no sense codified, is largely dependent on statute and has been kept up-to-date. Contract and tort remain the characteristic preserve of the common law. In contract, the Sale of Goods act, 1893, was a successful reduction to statutory form of a century of case-law. The doctrine of "frustration," which in exceptional circumstances may resolve a contract involving obligations not contemplated by the parties, is an important innovation. But, with these exceptions, the law of contract has not changed greatly since 1850, and its theories of consideration and mistake are technical and uncertain. The law of torts, however, has been greatly developed, although by comparison with European systems it presents a somewhat disorganized appearance; of cardinal importance has been the tort of negligence, which, particularly since the house of lords' decision in *Donoghue* v. *Stevenson* (1932), has absorbed much of the law relating to wrongful damage. In the law of persons, the Married Women's Property acts, 1870–93, whereby every married woman gained the right to hold her property as her separate legal estate, effected an important reform for which legal and not merely general public opinion can claim credit. In Dicey's famous phrase, "The rules of equity, framed for the daughters of the rich, have at last been extended to the daughters of the poor." (*See* HUSBAND AND WIFE, LAWS CONCERNING.)

The Role of the Judge.—Since Bentham it has been the judges rather than the writers on English law who have given it whatever common philosophy it may possess. No later writer, except perhaps F. W. Maitland, has exercised the authority of a judge, prepared to discuss legal problems in broad terms of principle. The great social and economic changes which followed World War II challenged some of the long-accepted ideals and even the traditional authority of the judiciary; the future of English law—or of that considerable part of it for which the legislature cannot or will not be responsible—depends in large measure on the understanding and courage of the judges in meeting this challenge.

See also references under "English Law" in the Index.

BIBLIOGRAPHY.—Sir W. Dugdale, *Origines Juridiciales,* 2nd ed. (1671); T. Madox, *History and Antiquities of the Exchequer,* 2nd ed. (1769); F. E. Harmer, *Anglo-Saxon Writs* (1952); E. Levy, *West Roman Vulgar Law* (1951); O. W. Holmes, *The Common Law* (1881); Sir J. F. Stephen, *History of the Criminal Law of England* (1883); H. Brunner, *The Sources of the Law of England,* trans. by W. Hastie (1888); K. E. Digby, *An Introduction to the History of the Law of Real Property,* 5th ed. (1897); H. Lévy-Ullmann, *Le Système juridique de l'Angleterre* (1928); Sir F. Pollock and F. W. Maitland, *History of English Law Before the Time of Edward I,* 2nd ed. (1898); Sir W. S. Holdsworth, *A History of English Law* (1903–52); A. V. Dicey, *Law and Public Opinion in England During the 19th Century* (1905), *Introduction to the Study of the Law of the Constitution,* 9th ed., ed. by E. C. S. Wade (1939); Morris Ginsberg (ed.), *Law and Opinion in England in the 20th Century* (1959); W. Friedmann, *Law and Social Change in Contemporary Britain* (1951); Sir W. Ivor Jennings, *The Law and the Constitution,* 4th ed. (1954); C. S. Kenny, *Outlines of Criminal Law,* ed. by J. W. C. Turner, 17th ed. (1958); R. M. Jackson, *The Machinery of Justice in England,* 3rd ed. (1960); W. M. Geldart, *Elements of English Law,* ed. by Sir W. S. Holdsworth and H. G. Hanbury, 5th ed. (1953); W. F. Walsh, *History of Anglo-American Law,* 2nd ed. (1932). (N. S. M.; T. F. T. P.; T. S. LE.)

ENGLISH LITERATURE. The literature of England, and, if it is written in English, of the whole of the British Isles, is part of the literary heritage of the English-speaking world. Its influence lies behind the literature of all the countries which have ever formed part of the commonwealth, and of the United States, and is intertwined not only with the literatures of other European countries, but of the emergent African nations, of India and Pakistan and of the middle east. This article is intended as a broad survey of the literature of England, from the Anglo-Saxon migrations in the 5th–6th centuries to the mid-20th century. It naturally deals with the work of many writers who were by nationality Irish, Scottish or Welsh, but whose writings form part of the main stream of English literature.

There are separate articles on IRISH LITERATURE; SCOTTISH LITERATURE; and WELSH LITERATURE; and on the literature in languages of minorities within the British Isles—CORNISH LITERATURE and MANX LITERATURE. In England as elsewhere in Europe the language of learning long continued to be Latin; for this literature *see* LATIN LITERATURE: *Medieval and Renaissance.*

There is also an article on ANGLO-NORMAN LITERATURE—the literature in the language of the educated classes in the period after the Norman conquest.

For other literatures in the English language *see* AMERICAN LITERATURE; AUSTRALIAN LITERATURE; CANADIAN LITERATURE (ENGLISH). The article INDIAN LITERATURE contains references to Indians writing in English, and shows the influence of English literature on the modern literature of a commonwealth country where English is not the native language. There are also articles on particular aspects of English literature, to which references will be found in the text; on such general subjects as BIOGRAPHY; DRAMA; ESSAY; NOVEL; PASTORAL; POETRY; PROSE; etc.; and on most of the writers mentioned.

This article is divided into the following main sections:

I. From the Earliest Times to Chaucer
  1. The Old English Period
  2. Early Middle English Period
II. From Chaucer to the Renaissance
  1. Poetry: The Alliterative Tradition
  2. Poetry: The Chaucerian Tradition
  3. Non-Courtly Poetry
  4. Drama
  5. Prose
  6. The New Learning
III. The Renaissance (*c.* 1510–1660)
  1. Character of the Age
  2. Renaissance Culture
  3. Poetry Before 1580
  4. Prose Before 1580
  5. The Novel, 1580–1600
  6. Other Prose, 1580–1600
  7. Spenser and Sidney
  8. The Successors of Spenser
  9. Sonnets and Mythological Poetry
  10. Lyric Poetry
  A. Elizabethan and Jacobean Drama
    1. The Transition From Medieval Drama
    2. Plays of Lyly, Peele, Greene and Kyd
    3. Marlowe
    4. Shakespeare
    5. Jonson
    6. Other Jacobean Tragedy
    7. Other Jacobean Comedy
    8. Minor Plays and Masques
  B. Early 17th Century: Prose and Poetry
    1. Character of the Prose
    2. Translations and Books of Travel
    3. The Essay and "Character"
    4. History and Biography
    5. The Influence of Religion
    6. Religion, Science and Philosophy
    7. Verse Translations
    8. Spenserian Influence
    9. Donne and the Metaphysical Poets
    10. Ben Jonson and His Influence
    11. John Milton
IV. The Restoration Period
  1. Prose
  2. Poetry
  3. Drama
  4. Summary of the Period

V. The 18th Century
1. Essayists and Pamphleteers
2. Swift
3. Criticism
4. Johnson
5. The Revolution of Taste
6. The Novel
7. The Poetry of Pope
8. Other Satiric Poets
9. Reflective Poetry
10. Blake and Burns
11. Belles-Lettres and Learned Prose
12. Drama
VI. The Romantic Period
1. The Nature of Romanticism
2. Poetry
3. The Novel
4. Miscellaneous Prose
5. Drama
VII. The 19th Century
1. The Aftermath of Romanticism: Tennyson
2. The Aftermath of the French and Industrial Revolutions
3. The Cult of the Medieval
4. Carlyle
5. Bentham and Coleridge
6. Carlyle and Newman
7. Darwinism
8. The Novel
9. Critics
10. Conclusion
VIII. The 20th Century
1. Early 20th-Century Novelists
2. The Early Development of "Modern" Poetry
3. The Impact of World War I and Its Aftermath
4. The Changing Scene After 1930
5. Poetry After 1945
6. Postwar Fiction
7. Drama

(X.)

I. FROM THE EARLIEST TIMES TO CHAUCER

Since the profound effects of the Norman conquest on literature begin to be apparent in the early 12th century, pre-Chaucerian writings are conveniently divided between two periods, with the boundary about 1100. These are the Old English (Anglo-Saxon) and the Early Middle English periods.

1. The Old English Period.—*Poetry.*—The Anglo-Saxons, who invaded Britain in the 5th and 6th centuries A.D., brought with them a vigorous oral poetic tradition. This versifying was in the alliterative accentual metre (*see* ALLITERATIVE VERSE), with an elaborate heroic diction adapted to the celebration of wars and leaders, or to the expression of folklore or pagan religious formulae. Prose writing occurred only with the later development of a more self-conscious Saxon-Christian culture and, even at its maturity, this prose remained far more secondary and derived than the poetry. Apart from annals and chronicles it consisted, indeed, very largely of translations, adaptations or expansions of Christian Latin sources.

The great event in the early story of Old English poetry was the adaptation of the aristocratic heroic warlike style and diction, with the traditional metre, to the imaginative expression of Christian subjects and thought-patterns. This major revolution may be symbolized by the work of the 7th-century poet Caedmon as described by Bede in his *Historia ecclesiastica gentis Anglorum.* Caedmon, a peasant lay-brother miraculously granted the gift of poesy through a vision, is said to have dictated a series of poems based on free adaptations of biblical material: but instead of the peasant language that might have been expected, he used all the technical diction and stylistic devices of the native aristocratic tradition. Only a hymn to the creation, of nine lines, has survived of Caedmon's poetry; but the impulse he gave set the pattern for almost the whole art of Anglo-Saxon verse—and notably for the pieces entitled *Genesis, Exodus* and *Daniel.* Of these latter works of "the School of Caedmon" *Exodus* is of real literary worth, vivid, imaginative and at times most difficult (*see* CAEDMON: *The "Caedmon Poems"*).

This early Christian poetry, beginning with Caedmon, seems to have followed an earlier tradition in being recited to the accompaniment of a harp to bring out features of the rhythm. Such an instrument was found in fragments among the treasures in the Sutton Hoo ship cenotaph and is shown, expertly reconstructed, in the British museum, London.

About 30,000 lines of a vastly larger quantity of Old English poetry have survived. It includes work originally composed from dates throughout the whole period, though mostly surviving only in copies made at the time of the 10th-century monastic reformation.

The four main collections, all made in religious houses about 1000, are as follows. The Beowulf manuscript, which besides containing the most important of all Old English heroic poems, from which it takes its name, includes a late one on *Judith* as well as prose pieces of marvels and didactic entertainment from Latin sources. The so-called Caedmon manuscript of biblical poems contains the Old Testament pieces noticed above and some others. Thirdly there is the Exeter Book (*q.v.*) in Exeter cathedral; and, lastly, the Vercelli Book (*q.v.*), found in 1822 in Vercelli cathedral library, north Italy. The Exeter Book, apart from some few prose documents, looks like a kind of anthology of lyric poems. Besides Cynewulf's *Juliana* and *Christ* (*The Ascension*), it contains a group of lyric poems of real aesthetic value, didactic but of high imaginative quality. These are *The Wanderer, The Seafarer, The Ruin, Deor, The Wife's Complaint* and *The Husband's Message.* They are elegiac in the sense of Thomas Gray's *Elegy,* treating of universally significant themes for meditation. The last two, with a fragment in the same manuscript known as *Wulf and Eadwacer,* are the only approaches to Old English love poetry that have survived.

There are also a few fragmentary heroic poems outside the four principal collections: the *Finn Fragment,* dealing with heroic feuds and fighting; *Waldere,* probably of very early date; *The Battle of Brunanburh* (937); and that on the epic fight at *Maldon* (991), from the end of the 10th century.

Apart from Caedmon, the only considerable Old English poet whose name is known is Cynewulf, who probably lived in the late 8th century and has left his name in Runic characters in the form of an epilogue to each of his four authentic poems: *Juliana* and *Christ* (in the Exeter Book), and *Elene* (on the finding of the Cross by St. Helena) and *The Fates of the Apostles,* in the Vercelli Book. He was a careful and technically skilful craftsman in verse who could make his mainly hagiographical or homiletic themes at times beautiful with lyric passages of deep feeling. Cynewulf seems also to have founded a school of poetry which produced work in his style such as *The Phoenix, Andreas* (on traditional legends of St. Andrew, written in heroic style), *St. Guthlac,* and a series of *Riddles,* all in the Exeter Book except *Andreas,* which is in Vercelli. The *Riddles,* though based on well-known Latin sources generally, include matters of folklore and rustic life that vivify the social scene, as well as some that show in their descriptive powers the Old English unsentimental but semimystical attitude to nature at its best.

Old English poetry is generally largely didactic, and sometimes entirely so, as, for instance, the *Rhyming Poem* (combining alliteration with end-rhyme) and the series of platitudes called Gnomic verses in the Exeter Book. But, to an audience that liked and expected moralizing, it is probable that the aesthetic appeal of rhythm and diction and sonorous recitation to the harp made almost any of such verses acceptable entertainment. Among all the Old English religious poetry there is one poem that stands out as of supreme excellence in itself as well as for its historical importance. This is the anonymous *Dream of the Rood,* recorded in some fragments in a Runic inscription of the early 8th century on the Ruthwell cross in Dumfriesshire and preserved in full in the Vercelli Book. This poem personifies the Cross, describes Christ in terms of a supreme Germanic hero with his disciples as the traditional band of loyal followers (*comitatus*) and puts the whole within the framework of a vision, ending with a personal elegiac epilogue. The unknown poet, who had something in common with Cynewulf but a far greater depth of feeling, used the symbolism of medieval hymns and ideas from the ritual associated with the Rood and its later legends to make an almost perfect

harmony and fusion of form and matter, of pre-Christian and Christian tradition, of individual and universal feeling. His blending of the dream with personification seems unconsciously to point a way which was to be fulfilled in the 13th century by the French *Roman de la Rose* (*q.v.*) and its progeny. *The Dream of the Rood* is one of the great religious poems of Britain.

But the greatest claim of Anglo-Saxon to a place in European literature lies in *Beowulf*, an epic and episodic poem of 3,280 lines composed probably in the mid-8th century. Written in the best traditional style and metre, it treats, with high seriousness, of combat between an ideal Germanic hero and demon monsters, against a background of legendary history and Christian symbolism, and shows the noblest qualities of Anglo-Saxon culture, with its loving connoisseurship of treasure and the craftsmanship of the smith, which have been made comprehensible by the discoveries at Sutton Hoo. *Beowulf* is a great poem in its own right, and the greatest poetic monument in Britain before the work of Chaucer and Langland. (*See* also the article BEOWULF.)

Prose.—Old English prose, which mainly follows Latin models, consists largely of translations and the like which scarcely fall within an account of literature proper. King Alfred (*c.* 848–899) made many translations of Latin philosophic and theological writings, the most important being Boethius' *De consolatione philosophiae* (*Consolation of Philosophy*) and Pope Gregory's manual *Cura pastoralis* (*Pastoral Care*), and these to some extent laid the foundations of a native prose. His own prefaces and interpolations in these works, and in particular his introduction to the *Pastoral Care*, show some individuality and throw light on his personality and times. It was, however, not till the monastic revival of the later 10th century that prose became a conscious literary art with clear stylistic effects in the homilies of Aelfric (950–1020) and Bishop Wulfstan (d. 1023). Aelfric could use his material with flexibility of style, the arts of a trained rhetorician and the efficient enlargement of the English vocabulary for the expression of theological lessons. He could also show both simple pathos and lofty emotion. His *Colloquies*, conversations on everyday topics and rural occupations intended for teaching Latin to boys, are the first known examples of really colloquial Old English. Wulfstan shows more vitality and individual force in his popular sermons, and some of them have something like the best qualities of a good revivalist sermon.

But the most outstanding performance in prose in Old English was the Anglo-Saxon Chronicle (*q.v.*), starting with the annalistic writings of the 7th century (*see* also CHRONICLE) and extending in one version to the 12th century. The early annals, as well as being the first examples of vernacular prose in any European language, also mark the beginning of non-Latin historiography. Beginning as a brief record of the year's events, they gradually developed into chronicles giving quite detailed descriptions of what had happened. By the time of the monastic revival the chroniclers had attained to a variety and excellence of style, some conscious rhetoric and something of an art of history. The entries for these years include several poems, notably the ballad on the battle of Brunanburh (937) and one on Edgar's coronation (973).

The Anglo-Saxon Chronicle illustrates four phases of development in Old English prose: first, the purely utilitarian, in the briefly recorded annals which are parallel in style to early laws and charters; then the clear narrative without any literary art, illustrated by the entries for King Alfred's period; thirdly, an apparent survival of that kind of orally tradited local history associated with the term saga, with its rapid passing from indirect to direct speech, its allusive assumption of an already fully informed audience and its stark vividness, shown in the interpolation in the Chronicle under the year 755 on the slaying of King Cynewulf and its avenging. Finally, the reign of King Aethelred II the Unready (978–1016) shows a historiography which at times approaches conscious and successful literary art.

Of the three great ages of Anglo-Saxon culture, therefore, the time of Bede produced the finest poetry, King Alfred's period the beginnings of prose, and the monastic revival, which revitalized literature, created a mature prose. While this latter was primarily didactic, it did include the first signs of interest in prose written entirely for enjoyment and serving no utilitarian purpose. This is seen in the romantic tale of *Apollonius of Tyre* (*q.v.*), freely adapted from Latin in the later 10th century, and in the inclusion in the Beowulf manuscript of the romantic marvels of a St. Christopher legend, and of Aristotle's *Letter to Alexander* and *The Wonders of the East*. *Apollonius of Tyre* is the one example of secular prose with real literary merit: for it is simple, clear, vivid and occasionally even graceful. At this time, which marks the beginnings of scholarship in the modern sense in England, some scientific works, such as *Byrhtferth's Manual*, were produced; and there was a further widening and deepening of that use of the powers of the Old English vocabulary for technical purposes already noticed in Aelfric's writing.

Cultural and Literary Achievement.—Anglo-Saxon culture reached a remarkably high level, and has left outstanding monuments in almost all the arts except painting and music. From the Sutton Hoo treasures of the mid-7th century with their exquisite examples of the craft of native goldsmiths and gem setters to the Bayeux tapestry (*q.v.*) of the 11th century, which commemorated the first conquering exploits of the Norman invaders, there flowed, at varying speeds and levels in different regions and periods in Anglo-Saxon England, streams of all kinds of aesthetic development in the arts. Sculpture in 8th-century Northumbria, church ritual and liturgy in the work of Alcuin and the monastic revival, illuminated writing—from the Irish-Latin Lindisfarne Gospels, copied at Lindisfarne, Northumberland, and the Book of Kells, made in the monastery at Kells in Ireland, both of the earlier period, to the world-famous Winchester school, which produced St. Dunstan's exquisite Benedictional—are among the many works of art of extraordinary variety. These provide a background and some explanation of the qualities of Old English literature, especially its poetry. The belief among 19th-century historians of the relative decadence of the time immediately preceding the Norman conquest and of Britain's isolation from continental Europe, has been generally discarded in favour of the view that Anglo-Saxon England, which had been a main civilizing force in western Europe in the 8th century, the period of Bede, St. Boniface and Alcuin, was in many ways at a higher cultural level than its Norman conquerors. The Norman conquest can be seen, indeed, as a severe setback to the development of literature in England, though ultimately it contributed to its broadening and enrichment.

Difficulties caused by the great changes in vocabulary and patterns of thinking make it hard to form a judgment of the qualities of Old English literature. Its prose is of great value to the historian and the student of sociology, but, with a few exceptions, cannot be said to merit study strictly as literature. Old English poetry, however, with a varied life extending over the four centuries from Caedmon to the Norman conquest, provides, amid much nonliterary didacticism and purely utilitarian matter, monuments that will repay those who seek a purely literary pleasure, as well as furnish endless riches to the specialist in comparative literature. This poetry, preserving in its earlier examples the Germanic tradition of the *scop* or poet reciting, after feasting, to the harp in the halls of heroic leaders, is a unique example of how a heroic diction could be adapted to Christian purposes without the loss of its basic qualities, and in a native rhythm that was never entirely to disappear from English verse.

Throughout the Old English period, inevitably, Latin remained the most widely used language among writers, who were generally ecclesiastics. Bede produced the pioneer work in scientific historiography; St. Aldhelm and Alcuin were Latin poets as well as learned writers; and Aelfric wrote both in Latin and English. King Alfred was a great preserver of native poetry, but most of what we know of his life is to be found only in Asser's biography of him (in Latin), which was the first secular work of the kind to be written in Britain. King Alfred, and the monastic revivalists, did much for the art of translation from Latin, however; and it may fairly be said that Aelfric left the language a varied and powerful vehicle for the expression of literary creation as well as of a large variety of technical learning.

2. Early Middle English Period.—Though oral tradition

among the peasantry and in remote religious houses where English was still spoken persisted in poetry, and to some extent in purely religious prose, the Norman conquest generally for a time obliterated the best tendencies of Old English literature and caused a period of relative vernacular barrenness: for popular native matter was scarcely likely to get itself preserved in writing under Norman and then French domination. Learning came to be again confined to Latin, thus undoing much that Aelfric's age had achieved. For entertainment and higher secular culture the new educated classes naturally had their literature in Anglo-Norman and later in French. The old poetic metre and diction ceased to be written save in remote places. There was, in fact, at least in the written literature, something like a real break in continuity during most of the 12th century. The one literary monument of high stylistic value and importance for literary history is the Peterborough Chronicle (a continuation of the Anglo-Saxon Chronicle), especially its closing portion narrating King Stephen's reign. This is prose of startling colloquial vitality, vivid and yet often of great rhetorical force. It is a direct ancestor of modern English. In verse only the occasional survival of a purely didactic poem such as the *Poema Morale* or the fragment popularly known as *A Poem on the Soul and Body* (discovered in the 19th century among the archives of Worcester cathedral) mark the 12th century; so that this period is mainly one of breach and of transition, though this breach would probably not seem so apparent if the lost writings had survived.

Latin and Anglo-Norman were the languages that provided much of permanent value in the 12th and, to a great extent, the 13th, century, though there was probably a good deal of popular and religious writing in English that has failed to survive, for there are references to what appear to have been vernacular heroic legends and traditional sagas in Latin and French writings of the period which suggest more continuity than can be demonstrated. Furthermore, the long verse chronicle, Layamon's *Brut*, of the earliest years of the 13th century, from rural Worcestershire, with its blend of traditional alliterative metre with internal rhyme and its poetic diction often recognizably derived from Anglo-Saxon practice, has clear links with the pre-Norman age, and the heroic spirit of Old English poetry sometimes becomes apparent in the long, rather cumbersome lines as a conscious inheritance. Even though he was adapting his material from the rhyming Norman chronicle of Wace, Layamon seems to look back to the more aristocratic Old English heroic poetry. But it must be assumed that alliterative accentual verse also survived as an oral tradition among the English peasantry, though no written contemporary evidence survives, if the emergence of such writings as those of the poet of *Piers Plowman* in the later 14th century is to be explained.

Layamon's *Brut*, so named from a mythical Trojan Brutus, is one of the signs of an early 13th-century renaissance of vernacular literature. This came about partly through historical causes and was greatly animated by the revival in the church, especially by the coming of the Friars in the 1220s (*see* ENGLAND, CHURCH OF), who interested themselves particularly in native literature—an interest still remembered by Chaucer in his description of his Friar as one who could "make his English sweet upon his tongue." The revival of English lyric poetry, associated with music and a highly developed technique, owed most to the Friars who collected all kinds of lyrics, both religious and secular, in omnibus manuscripts such as the famous Harley manuscript 2,253, in the British museum, London, which includes poems in Latin and French, sometimes blending the languages, as well as a wealth of purely English lyrics.

The *Brut* is also important in literary history for including the earliest surviving English romantic legends of King Arthur. Taking material of Celtic origin which, it is claimed, was used by Geoffrey of Monmouth (*q.v.*) in his 12th-century Latin account of the British kings who reigned before Christ, the *Historia Regum Britanniae*, and expanded by Wace in his poem the *Roman de Brut* soon after, Layamon added other Welsh material of Arthurian lore, notably the account of the founding of the Round Table. He is indeed, in a sense, the fountainhead of English Arthurian romance. But it was from France, especially from the dominating writings of Chrétien de Troyes in the 13th century, that the romance, which is mainly derivative from French, got its best impulses. English Arthurian work, however, both in verse and prose, did not produce most of its best till the later 14th century. *Arthur and Merlin* and *Richard Coeur de Lion* (both *c.* 1300) are among the best that fall within this period, but neither is above the level of good entertainment in verse. (*See* also ARTHURIAN LEGEND.)

The early 13th-century literary Renaissance manifested its vitality in two works of real and permanent merit: in verse *The Owl and the Nightingale* and in prose the *Ancrene Wisse* ("Guide for Anchoresses"), better known as the *Ancren Riwle* from a title invented by its first editor, which shares its good qualities to some extent with a group of other religious writings of the same date and in the same southwest Midland dialect. *The Owl and the Nightingale*, written just before 1200, apparently by a Surrey priest living in Dorset, is a blend of debate and beast-fable in vivid and often deliberately homely octosyllabic couplets, using traditional French literary machinery. It is entirely original, however, in combining a formal disputation, between two symbolic yet vividly birdlike protagonists, about everything that interested the poet, with the inheritance of the Bestiary (*q.v.*), in which animals and their traditional legends have an allegorical significance. This poet, while following French models in form and much of his technique, is the first in English to seek a language for poetry of the kind Wordsworth was later to strive after—"a selection of the language really used by men"—in conscious protest against the more artificial style suggested by incoming French models culminating in the highly literary *Roman de la Rose* of the mid-13th century. His birds are vividly human as well as real birds, and his knowledge of birds makes him the first English literary ornithologist. *The Owl and the Nightingale*, which also reminds one so remarkably of Chaucer's *Nun's Priest's Tale* and *Parliament of Birds*, shows a type of subtle yet realistic humour which is strangely anticipative of Chaucer.

The *Ancrene Wisse* is a prose treatise in vigorous and often colloquial language by a wise old priest for three young aristocratic women who needed help in trying to evolve their proper daily lives as anchoresses. It shows a powerful and attractive personality using traditional material blended with lively experience of the world, and has a rhythmic style that looks back to the Old English Wulfstan at his best in the passages of strong feelings, or to Aelfric in others of simple narrative. Its thought seems to look forward to the religious prose of the ensuing centuries as far as Sir Thomas More (1478–1535); but this continuity of thinking is not matched by continuity of language, since the *Ancrene Wisse* uses a west Midland dialect a good way removed from the ancestor of modern literary English (which looks back rather, as has been said, to the Peterborough Chronicle). A group of saints' lives in MS. Bodley 34, in the Bodleian library, Oxford, often called the Catherine Group, was written at about the same time and place (probably Herefordshire) but in a more definitely rhythmic manner, almost at times becoming free verse rather like Aelfric's *Lives*. The homily on the defense of the soul named *Sawles Warde* in the same manuscript is among the very best of medieval sermons. Its description of hell is terrifyingly stark and vivid. All these pieces of didactic prose belong to about the same period as the *Brut* and *The Owl and the Nightingale*.

A peculiarly English type of romance, more bourgeois and popular and homely than the chivalrous French products, grew up toward the close of the 13th century. *Havelok the Dane*, *King Horn* and *Floris and Blauncheflour* are tales in simple verse such as the less aristocratic minstrels were wont to provide. They illustrate a definitely English kind of life and way of thinking, even consciously rejecting the manners and attitudes of romantic *amour courtois* or "courtly love" (*q.v.*). *Havelok* has been called an "Anglo-Danish romance" because it comes from Lincolnshire, which had been invaded and settled by the Danes, uses Viking legendary history and shows something of Viking thought-patterns. *Floris and Blauncheflour* is an English love story which shows the oriental influences of the Crusades. *Athel-*

ston, composed in the middle of the 14th century, is a less literary piece of fiction intended for the entertainment of London's artisans and craftsmen.

It was toward the close of the 13th century that English literature, which had probably drawn on Welsh sources in very early times, again felt Celtic influence: this time in a creative way, though it came only indirectly from Brittany through French. This renewed influence provided a new kind of semipopular romance which is, like its French prototype, named the *Breton Lay*. These English adaptations from French are short romances in verse, embodying tales of love and the supernatural, and introducing an originally Celtic conception of *Faerie*. Marie de France, who lived in England in the mid-12th century, first invented this *Lai Breton*, which she claimed was a translation from the Celtic: and these French aristocratic courtly romances were later, often through intermediate French versions, rendered or imitated in English, but in a more popular vein and with definitely English attitudes replacing the French courtly ones. *Sir Orfeo*, based on a lost *Lai d'Orphée* of Marie de France, is the earliest English Breton lay, from the close of the 13th century: it is an attractive blend of the ancient classical tale of Orpheus and Eurydice with happenings in the Celtic land of enchantment or faërie. It was from this romance that Spenser may have got his famous "lond of Faerie," as the phrase first occurs in it. The Breton lay was popular in the 14th century, with *Sir Launfal* (going back to a French *Lanval*) as its finest exemplar, though the oldest surviving manuscript was written in the 15th century. Chaucer's *Franklin's Tale* suggests a revival of this popularity of the Breton lay in his time.

Another feature of the 13th-century renaissance was the development of lyric love poetry, both secular and also with the application of its style and diction to religion. Love poetry had had to wait for the impact of medieval Latin-French culture on Britain for its proper flowering; for, as has been observed, it scarcely existed in Old English. Songs to idealized loved ones and to the Blessed Virgin mark the 13th century; and at its close there appeared one of the greatest religious love poems of the language. This is entitled, from its refrain taken from the Song of Solomon, *Quia Amore Langueo*, "For I am sick of love." It seems to anticipate many of the qualities of Francis Thompson's "The Hound of Heaven."

Early Middle English literature is basically didactic, though its hearers and readers were not conscious of the separation of the religious from the secular which in modern times is generally assumed. Hence much, both in verse and prose, in this period seems, to the modern reader, to have very little literary value. Long religious histories and homiletic series in verse, such as *Cursor Mundi* and the *Ormulum* in the 13th century, and *The Prick of Conscience*, ascribed to Richard Rolle of Hampole, and the *Ayenbite of Inwyt* ("Remorse of Conscience") in the 14th century, have little interest except for the linguistic student. Similarly, the 13th-century *Proverbs of Alfred* and *Proverbs of Hendyng*, collected from older material some of which was attributed to King Alfred, scarcely find a place in literary history proper. Or, again, there are verse chronicles, of which the most lively is that of Robert Mannyng of Brunne. His treatise on the Seven Deadly Sins, called *Handlyng Synne* ("Manual of Sins"), however, is of real literary value in parts, especially for its vivid and vigorous illustrative anecdotes. The political and satiric topical verses of Laurence Minot, in the early 14th century, are sometimes entertaining.

Under French influence at this time the *fabliau* (*q.v.*) appeared in English dress in the ribald but entertaining tale of *Dame Siriz*, which is the only example of the *fabliau* strictly so-called in English before Chaucer. About this time, the mid-13th century, the only successful version of a traditional fable occurred, in the lively and realistic verses of *The Fox and the Wolf*. This latter is the only real echo of the French cycle of the *Roman de Renart* (*see* REYNARD THE FOX) till Chaucer's *Nun's Priest's Tale*.

Mysticism and its practical exposition for disciples, which was a marked growth in western Europe at this time, found creative exponents in England. Richard Rolle wrote often moving and attractive works in both prose and verse in English though his finest work is in Latin. The anonymous author of *The Cloud of Unknowing* and of some related smaller treatises, showing the influence of versions of Dionysius the Areopagite (*q.v.*), and Walter Hilton in his *Scale of Perfection*, wrote prose in the latter half of the 14th century which has at times a permanent and universal aesthetic appeal although it was primarily intended for those with a vocation to the contemplative life.

This brief outline of English verse and prose, from the beginning of the 12th century to the second half of the 14th, shows that Early Middle English literature is far larger and more varied in its types than Old English, and that it developed new and permanently significant genres in lyric poetry, in prose and in verse romance. It looks forward in many ways to the greater Chaucerian age.

(C. L. W.)

II. FROM CHAUCER TO THE RENAISSANCE

1. Poetry: The Alliterative Tradition.—The middle years of the 14th century saw the beginning of a remarkable new fertility in the composition of long poems in the alliterative metre (*see* ALLITERATIVE VERSE). Among the early group are very varied themes—love romances such as *William of Palerne*, a version of a French romance (written before 1361); two quasi-historical fragments on the life of Alexander (*see* ALEXANDER ROMANCES); political satire such as *Winner and Waster* (*c.* 1353); religious legend such as *Joseph of Arimathie* (*c.* 1350), which treats of the Holy Grail—and all these kinds long maintained their importance. Surviving records give the impression that this cultivation of alliterative verse had suddenly been taken up again after a long interval, and the movement is generally called "the alliterative revival." But the term is not wholly appropriate, for though there was apparently a great upsurge of interest in this kind of writing it cannot have been a return to a mode long out of use. Even in the early poems of the "revival" the technique is correct within its own conventions, and these conventions are not, except in a general way, those of Old English versification. It cannot therefore be an antiquarian attempt to return to old practices, but must be rather the continuation of a tradition. The differences from Old English technique are due largely to the developments that had taken place in the language, and evidently the alliterative mode must have been continuously practised and gradually modified. The so-called revival should therefore be seen as a striking expansion of an already existing way of writing. It was a western and northern movement, contrasting with the rhyming technique by then universal in London and the east. Its regional and archaic quality appears not only in the use of the old native metre but also in the preservation of many words and phrases characteristic of the old poetic diction. This kind of verse continued to be written in Scotland up to the early 16th century. There is a considerable volume of it, much of no great accomplishment; but the best work, all apparently of the 14th century, is equal to the finest that Chaucer wrote. The "heroic romance," *Morte Arthure*, though not of the first rank, is virile and noble and served as Sir Thomas Malory's model for part of his *Morte Darthur*. *Piers Plowman* (*see* LANGLAND, WILLIAM) expresses in a complex allegory, with unique power and richness, the author's burning faith in man's need for Christian truth and charity. In part Langland seeks a remedy in religion for the social evils that he depicts with extraordinary vigour, but he is concerned with Christian truth in all its aspects and with the salvation of the soul far more than with social reform. The poem is long and irregularly constructed; its language is often unpolished and its metre loose. It is therefore often difficult to follow, and striking felicities of detail are only occasional. But its intensity of feeling and range of thought have no parallel. It is a magnificent statement of the mind and faith of the late 14th century, and one of the greatest of English poems.

In a very different tone, and yet drawing upon essentially the same springs of behaviour, is the chivalric romance *Sir Gawain and the Green Knight*. The author tells a fabulous tale of enchantment in an Arthurian setting; but he depicts his characters, and his hero's moral dilemma, with such acute and subtle understanding that the poem has the effect not of a mere entertainment but of

a serious reflection upon human conduct: Gawain is presented as a devout Christian, humanly imperfect. The technique of this poem is greatly superior to that of *Piers Plowman*. The alliterative lines, which are broken up into irregular stanzas by short rhyming passages, are tautly constructed, and the vocabulary is extraordinarily rich—it is strongly influenced by French in the scenes of courtly life, but is coloured also by many dialectal words, often of Scandinavian origin, of the northwest of England. The blend of sophisticated atmosphere, psychological depth and regional language produces an effect unlike that of any other work of the time.

In the same manuscript as *Sir Gawain*, and with so many similarities of language and style that they are now generally attributed to the same anonymous author, are two alliterative homilies, *Patience* and *Purity*, and an ostensibly elegiac poem called *Pearl*, which is written in rhymed 12-line stanzas heavily reinforced by alliteration and linked by verbal repetition. Here the poet presents himself as seeing in a vision his daughter who had died in infancy and being instructed by her in submission to God's will. He conducts an earnest theological debate in an intricate technique, showing in descriptions an enjoyment of colour and light like that which appears in many scenes of *Sir Gawain;* yet for all this complexity he conveys a most poignant sense of personal grief. *Pearl* is on the fringe of the alliterative school. The basis of its technique is the French use of rhyme and syllabic measure, adapted to the stress of English, and alliteration, though prominent, is inessential. Similar combinations of alliteration with rhyme and stanza schemes are found in romances such as *The Aunters of Arthur*, and in the mystery play cycles, most successfully in those of Wakefield and of York.

2. Poetry: The Chaucerian Tradition.—The alliterative revival reached great heights, especially in *Piers Plowman* and *Sir Gawain;* yet it was the end of a line. The future lay mainly with the other tradition of Middle English verse, that modeled on French and represented at its best in the earlier period by *The Owl and the Nightingale* and *Sir Orfeo*. Soon after the alliterative revival began this other tradition acquired new and decisive life from the early works of Geoffrey Chaucer. His *Book of the Duchess* and *House of Fame* were written in octosyllabic couplets as *The Owl* had been, and technically were not markedly superior to it. Even in tone they did not make a revolutionary change from earlier English work on French models; Chaucer's humorous detachment in *The House of Fame*, especially, is foreshadowed in *The Owl*, though his delicacy and subtlety are superior to anything that had gone before. These early works, for all their skill and grace, still owe an enormous debt to France—the influence, both in setting and in form, of the *Roman de la Rose*, which Chaucer says he translated, is very great. It is with *The Parliament of Birds* (*c.* 1382), light though the tone is, that Chaucer establishes himself as a powerful new genius. It has some philosophical depth, and the dialogue has a range and perceptiveness hitherto unattained, looking forward to the superb conversation scenes of *Troilus and Criseyde*. In *The Parliament* Chaucer used, for the first time in a major English poem, the ten-syllable or five-stress line which has been immensely important ever since. He evidently adopted it from French, with the additional example of the hendecasyllabics used by the Italian poets whom he had begun to read and imitate. In *The Parliament* the lines are arranged in stanzas of seven, the so-called "rhyme royal," which Chaucer used again in *Troilus and Criseyde*. It was in the *Prologue* (the first version of which perhaps dates from *c.* 1386) to *The Legend of Good Women* that he first used the ten-syllable line in couplets, which he continued to use in most of the best of *The Canterbury Tales*. Chaucer's absorption of continental influences is most strongly seen in *Troilus and Criseyde*. Here his plot derived from Boccaccio, much of his philosophical background from Boethius and Dante, much of his attitude to his leading characters from the conventions of courtly love (*q.v.*). But he transformed Boccaccio's direct narrative, especially by deepening the treatment of character. He developed the story on such a scale that he had scope to illuminate the slow movements of inclination or loyalty, and to exhibit the effects of long-drawn hope or despair, with a precision that he had no room to attain even in *The Canterbury Tales*. Yet though *Troilus and Criseyde* has claims to be considered Chaucer's masterpiece, it cannot finally stand against the "God's plenty" that Dryden hailed in *The Canterbury Tales*. Though the tales vary greatly in interest, the finest of them are immensely rich in the subtle play of incident and character, and the pilgrims who tell them are revealed, in the linking passages and in the incomparable *Prologue*, with the greatest shrewdness and penetration. Chaucer never essayed a fully tragic tale or the kind of religious intensity that Langland expressed; but few other aspects of life escaped his sympathetic or satiric eye.

Chaucer was early recognized as a great master of English poetry, but for a long time he seemed less pre-eminent than he does today. John Gower, a somewhat older contemporary, was much praised up to Elizabethan times. Like Chaucer, he was a court poet, and his principal English work consists also of a series of stories, *Confessio amantis*. In addition he wrote two long didactic allegories, the *Mirour de l'omme* in French and *Vox clamantis* in Latin, which merit the address "O moral Gower" with which Chaucer dedicated to him *Troilus and Criseyde*. They are of interest mainly as showing the freedom with which a poet might then choose to write in any of three languages. The English *Confessio amantis* has real poetic quality, though of a minor kind. The style is always lucid, often graceful, the stories are neatly told, and the octosyllabic verse well managed. Gower is best in passages of plain narrative or gentle sentiment; but these, unrelieved by much wit of the Chaucerian kind, grow monotonous after some thousands of lines of the smooth but rather tinkling metre. Gower's work helped to consolidate the literary use of the cultivated language of London—his usage differs from Chaucer's only in a few minor features—and the rhymed verse characteristic of the south and east. Though the alliterative technique long survived in the north, and manuscripts of alliterative works like *Piers Plowman* were multiplied, it was as a declining mode. After the time of Chaucer, and no doubt largely because of his superb achievements, the practice of rhyme, and all that accompanies it in diction and manner, became dominant.

Chaucer's successors acknowledged their debt to him. Thomas Hoccleve or Occleve, in his excessively long *De regimine principum*, called him "the first fynder of our faire langage" and "my maister Chaucer, floure of eloquence, mirrour of fructuous entendement." Hoccleve was not a distinguished pupil. His works offer occasional interesting glimpses of London life and some unusually direct personal statements, but are for the most part naïvely didactic. The quotation above illustrates one of his faults, the affecting of Latinate vocabulary or "aureate language." From this time onward this pernicious habit was very common. It was much practised by the prolific John Lydgate, especially in some of his shorter poems. Lydgate also felt intense admiration for Chaucer and imitated him as far as he could (which was not very far) in versification—for example, he wrote *The Fall of Princes* (more than 36,000 lines), mainly in the rhyme royal that Chaucer had used in *Troilus and Criseyde*, and *The Troy Book* (about 30,000) in approximately decasyllabic couplets. Lydgate's long poems are verbose and laboured; in some of the shorter ones, such as the amusing if somewhat heavy-handed satire of *Bycorne and Chichevache*, he did much better. He was greatly admired in his own day and commissioned to write a large number of occasional pieces; and William Caxton printed seven of his works. His influence on other writers was thus much greater than his merits.

After Lydgate the "Chaucerian" or courtly tradition persisted mainly in minor works. A considerable number of short love-lyrics has been preserved, some by known authors such as William de la Pole, duke of Suffolk, and Charles, duke of Orleans, who was taken prisoner at Agincourt and spent more than 20 years in England. The best of them are highly accomplished, though they are generally skilful variations upon familiar themes rather than truly original compositions. Of longer poems perhaps the finest is the anonymous *The Flower and the Leaf*, a graceful "debate" poem in rhyme royal which purports to be written by a woman. The tradition continued into the 16th century with *The Pastime of Pleasure* (printed 1509), a laborious allegory of the seven liberal arts under the guise of a knight's suit for a lady's hand, by Stephen

Hawes. Hawes claimed Lydgate as his model; but, though his mind was not without quality, apart from a few scattered successes he lacked even Lydgate's ability in versifying.

Scotland.—The imitators of Chaucer in England were thus sadly unequal to their pretensions. It was otherwise in Scotland. Before the close of the 14th century the only noteworthy Scottish writing had been the *Bruce* (1376) by John Barbour and *The Pistill of Susan* attributed to Huchown. With *The Kingis Quair* (*c.* 1423), possibly by James I of Scotland, the "Scottish Chaucerian" school began; the poem ends with a reference to "my maisteris dere, Gowere and Chaucere . . . superlative as poetis laureate." The debt to Chaucer was real, and Lydgate also contributed by providing a model for the aureate diction which became an obtrusive mannerism in some of the work of William Dunbar and others. But the Chaucerian example was treated by the Scots with complete freedom. They showed a poetic power and an independence of mind utterly distinct from the insipidities of the southerners.

3. Non-Courtly Poetry.—The courtly tradition was not the only poetic field to be cultivated in the 15th century. There is a great body of religious poetry, mainly hymns, songs and prayers on various occasions of the Christian year, moralizing verses, and saints' lives. Of the last group the *Legends of Holy Women* by Osbern Bokenham and *The Life of St. Katherine* by John Capgrave are the most important. Some of the moral verses interestingly express familiar injunctions in pithy or proverbial form. The prayers and songs are often rather formal; some of them tend to ring the changes on a limited set of expressions, but some achieve a stately dignity. The best are the simplest, perhaps the finest of all the justly famous "I syng of a mayden that is makeles." This song is close in spirit to many Christmas carols, and it was in the 15th century that the carol—a song with a burden, not necessarily connected with Christmas—first appeared in English in great numbers. Most of those that survive are anonymous, but the names of some authors are known; for example John Audelay (*c.* 1430) collected 26 of his poems under the rubric, "I pray yow, syrus, boothe moore and las, Syng these caroles in Cristemas." Of 15th-century songs that are Christmas carols in the modern sense many combine a simple expression of rejoicing with the embellishment of Latin phrases:

> Wel mowe we glad and mery bee,
> Sith we were thralle and nowe be free;
> The fende oure foo he made to flee
> *Qui natus fuit hodie.*

Another important group is formed by the "lullaby" carols, which descend from the type of the well-known 14th-century song, "Lullay lullay, litel child, whi wepest thou so sore?"

Beside the richness of the devotional verse secular non-courtly poetry is limited in interest, but there are highlights. The outstanding poem is *The Nut-brown Maid*, a delightfully managed dialogue on women's constancy. This, with its theme of the outlaw, has affinities with the Robin Hood ballads, and these too flourished in the 15th century though they must have originated much earlier—by about 1378 Sloth in one version (the B-text) of *Piers Plowman* claimed to know "rhymes of Robin Hood and Randolph Earl of Chester." How far other ballads such as "Chevy Chase" and "Sir Patrick Spens," in the late forms in which we know them, date from this period is debatable; but there must already have been a substantial body of this quasi-historical minstrelsy, using simple language in simple stanzas, with frequent repetitions and many conventional turns of phrase. There is a wide range of ballad themes: dramatic scenes of battle, romantic elopements and pursuits, supernatural visitations. Nearly all are northern or Scottish in language and setting. Many achieve, with their austere technique, great power of feeling. (*See* also BALLAD.)

Political themes appear in a good deal of 15th-century verse, much of it workmanlike. Topics extend from a carol of thanksgiving for the victory of Agincourt (1415) to coarse triumph over a defeated opponent, as in the lines on the arrest of the duke of Suffolk in 1450—"Now is the fox drevin to hole!" The most important political poem is the anonymous *The Libel of English Policy* (*c.* 1436–37), rough in its metre but a competent piece of propaganda urging, "Cheryshe marchandyse, kepe thamyralté/That we bee maysteres of the narowe see."

There was still a public for chivalric romances throughout the 15th century. Several were translated or adapted from French, as *Sir Launfal* and *Landavall* from Marie de France, *Ywain and Gawain* from Chrétien de Troyes, *Ipomedon* (in three English versions) from Hue de Rotelande (*c.* 1175), an Anglo-Norman poet. Others were more or less independent compositions, though heavily derivative in incident and language; such were *Sir Eglamour of Artois* (printed 1500), *Sir Degrevant* (*c.* 1410) and *The Squire of Low Degree* (*c.* 1450). Some of them show a fair control of plot and some capacity for lively description, but in the main they are journeyman work, shapeless and unoriginal, and the "tail-rhyme" stanzas in which many of them are composed move at a tedious jog-trot. All too much of this later romance fully merited the ridicule that Chaucer poured upon the genre in *Sir Thopas*.

Toward the end of the century comes the unique figure of John Skelton, who fits into no tradition, yet is rather a reactionary than a man of the Renaissance. In his boisterous vitality he has affinities with William Dunbar. His most characteristic "Skeltonic" verse, a succession of short rhyming couplets, is all his own. It is most effective in passages of violent action or invective, but can also be strangely successful in tenderer moods—the lyrics addressed to the ladies of the court in *The Garland of Laurel* (1523) have a freshness and delicacy which look forward to the great flowering of Elizabethan song. In the keenness of his mind, the depth of his human sympathy and his mastery of words, Skelton is the greatest English poet of his time and his importance is much more than merely historical. The same cannot be said for Alexander Barclay, who has something to contribute in two ways—his adaptation of *The Ship of Fools* (printed 1509) from Sebastian Brant's *Narrenschiff*, and his introduction, in the *Eclogues* which he imitated from the Italian Mantuan, of a pastoral convention of writing which was to prosper in the more accomplished hands of Spenser and others.

4. Drama.—Probably to the 14th century belong the beginnings of English religious drama. The most important texts are the four cycles from Chester, York, Wakefield and the so-called *Ludus Coventriae* or "Hegge plays," which are often said to be from Lincoln but on linguistic grounds are more probably from Norwich. The Chester cycle, for instance, covers in 25 plays the Bible stories of the history of mankind from the beginnings—the Fall of Lucifer, the Creation, the Fall, Noah, etc.—until the Day of Judgment. It appears to be the earliest of the cycles in origin, though the extant manuscripts date from the late 16th century. The texts of the other three series are of the 15th century, and were certainly revised during that century—most notably in the work of the "Wakefield Master" who contributed five plays and parts of several others to the Wakefield cycle. His work is distinguished by a characteristic stanza form, combining rhyme and alliteration, and still more by its humorously realistic treatment of character and original command of colloquial language, seen best in his *Noah* and his second *Shepherds* play. The Wakefield plays are the finest of the cycles in variety of interest and in literary quality but there is some impressive dramatic writing in the York cycle also. All these cycles are Corpus Christi plays, acted by craft guilds. There were also a number of individual plays on similar subjects (such as *Abraham and Isaac*) not associated with cycles, and some on the lives of saints—the latter often called "miracle" plays in distinction from the "mysteries" based essentially on Scripture. These were all mainly narrative treatments of their subjects. Beside them arose, perhaps about 1400, a different type of play in which the characters are abstract qualities or types and the plot is allegory; these are called "moralities." The most important of the earlier ones is *The Castle of Perseverance* (before 1450), and of the later, *Everyman*, printed by Richard Pynson early in the 16th century. By their nature these plays never achieve the vigorous realism of the best of the mysteries, but they do rise to beautifully simple solemnity in treating the theme of death and the fate of man's soul.

5. Prose.—In the later medieval period English prose takes on a new importance. The devotional tradition of Richard Rolle and Walter Hilton was carried on in *The Cloud of Unknowing* and other anonymous treatises, probably of the late 14th century, in *The Sixteen Revelations of Divine Love* by Dame Julian of Norwich (*c.* 1342–after 1413) and in such work as that of Nicholas Love, prior of the Carthusian house of Mount Grace at Ingleby, Yorkshire, early in the 15th, especially his translation of the *Meditationes Vitae Christi* as *The Mirror of the Blessed Life of Jesu Christ* (printed 1486). Much of this prose is beautifully lucid, simple and easy in movement; some of it affects a strongly rhythmical cadence which removes it from the field of normal prose. But well before the 14th century ended good plain prose had begun to be used for a much greater range of subject. John Wyclif wrote many controversial works in which vigour of thought is matched by strong clarity of style; and in addition he inspired the great translation of the Bible which is known by his name, though his share in the work is a matter of doubt. There are two versions of the translation, the earlier often forcible but crude, the later—attributed to John Purvey—for the most part competently written and in places excellent. The later version won great popularity, and maintained it until William Tyndale brought out his New Testament in 1526. It had some influence on Tyndale, and through him on the Authorized (King James) Version of 1611 (*see* BIBLE, TRANSLATIONS OF). Chaucer also wrote prose of an edifying kind, rather stiffly, in *The Tale of Melibee*, *The Parson's Tale* and his rendering of Boethius. More important as a symptom of the trend of writing was his use of prose for a scientific purpose in *A Treatise on the Astrolabe* and, if it is his, *The Equatory of the Planets*.

From the later 14th century onward English prose came to be used for all kinds of secular purposes—chronicles, romances, documents, letters. John of Trevisa (d. 1402) translated Ralph Higden's *Polychronicon*, a universal history from the Creation to 1352, and the great 13th-century encyclopaedia *De Proprietatibus Rerum* by the Franciscan Bartholomaeus Anglicus. There are three English versions, about 1400, of *Mandeville's Travels*, a compilation of mainly fictional traveler's tales of the east, originally written in French (*see* MANDEVILLE, SIR JOHN). In the 15th century Capgrave wrote a *Chronicle of England;* an anonymous chronicle of England called *The Brut* survives in more than 120 manuscripts and there are several lesser chronicles of national and local affairs. Three different translations of the 13th-century collection of tales, the *Gesta Romanorum*, met the popular taste for stories. The style of such works as these is for the most part unpretentious, sometimes awkward but at its best lucid and with some sense of form and rhythm. It is close to the language of conversation, and this is what is naturally found again in many of the private letters of which the collections of the Paston family of Norfolk (*see* PASTON LETTERS), the Stonors of Oxfordshire and the Celys of London are the best known. These are the first documents of their kind in English history. They are not literature in the ordinary sense, for they were written with no thought of publication; but, besides giving an intimate picture of the life of the time unmatched in any earlier century or in any literary work, the best of them show a remarkable command of direct and serviceable English. A comparable plainness and vigour characterize the remarkable autobiography of Margery Kempe, a mystic of King's Lynn in Norfolk (*c.* 1373–*c.* 1440). Apparently illiterate herself, she dictated, late in life, an account of many religious experiences she had undergone and pilgrimages she had made. More formal prose was also written in various styles. The most congenial to modern taste is perhaps that of Sir John Fortescue, whose work *On the Governance of England* is clear, straightforward and shapely. Reginald Pecock in many controversial writings shows a logical mind and a highly individual style, too much given to elaborate sentences and eccentric word-order to suit English idiom. Skelton in his youth translated the *Bibliotheca Historica* of Diodorus Siculus (from a Latin version) into an astonishingly aureate English which anticipates the excesses of the "inkhornists" of the following century. By far the greatest writer of the century is Sir Thomas Malory, whose collection of Arthurian stories was finished in 1469–70 and printed by William Caxton, after a good deal of editing, in 1485, under the title *Morte Darthur*. From sources in English verse and French prose he composed a masterpiece of dignified romance and nostalgic tragedy, couched in a style conspicuous for its sonority of rhythm which yet depends greatly on the simple language of his day. No other work of the time is so deeply felt.

The introduction of printing into England in 1476 by Caxton revolutionized the dissemination of literary works. Caxton made accessible to a wider public than ever before most of the important English writings of the preceding century, and also published many translations, mostly from French, both by others (*e.g.*, Earl Rivers, John Tiptoft earl of Worcester, William Worcester, Stephen Scrope) and by himself. Some of these have value rather historically than as literature, and Caxton's prologues and epilogues commenting on his problems are of livelier interest than some of the books they accompany. He was keenly concerned to write English that would be readily intelligible; and though he was often led by his French originals into affectation and verbosity he could on occasion produce a simple if rather naïve style which still reads pleasantly. By circulating large numbers of books in a fairly homogeneous type of English he contributed much to the final stabilization of the language of London as the general literary medium.

6. The New Learning.—For most of the 15th century English literature was closely linked, in both form and content, to what had gone before, and was little affected by the new "humanism" which had arisen in Italy in the 14th century. Cultivated Englishmen were well aware of the movement, but as patrons rather than practitioners. Even early in the century men of scholarly interests such as Humfrey, duke of Gloucester, knew and sometimes employed Italian scholars and collected manuscripts of classical authors. But it was not until the beginning of the 16th century that the full tide of Renaissance humanism reached England. Even then, the full effects on English literature were slow to be felt. The epoch-making work of English humanism, the *Utopia* (printed 1516) of Sir Thomas More (*see* also below), was written in Latin, still the international language of scholarship. More did write English prose clearly, especially in *Richard III*, and movingly, especially in *A Dialogue of Comfort* (1534); but it continues the informality of earlier prose rather than imitating classical models. Good English prose was written, too, by men concerned with education—Sir Thomas Elyot, Sir John Cheke, Sir Thomas Wilson, Roger Ascham and by the Bible translators Tyndale and Miles Coverdale. Practice in the classics did much to give form to the English sentence; yet many of its most potent rhythms derive from the simplicities of English speech. The older drama survived beside the new. Spenser inherited his fable from Malory and much of his language from Chaucer. Though great masterpieces at the end of the medieval period are few, its importance for the understanding of the wealth that followed is immense. (N. D.)

III. THE RENAISSANCE (*c.* 1510–1660)

1. Character of the Age.—The reign of Elizabeth I (1558–1603) was the most splendid age of that period of literature which extended from about 1510 to 1660. Within this period there was a sudden and distinct darkening of tone, a watershed after which the prevailing temper, and in particular its literary expression, became less sanguine. If a date is to be set to this marked shift from optimism to introspective pessimism, from the spontaneous ease and "copie" of the earlier writers to the tension and "pressness" of their successors, the year of the queen's death will perhaps serve. This is not to say that before 1603 all drama was romantic comedy, stirring history or straightforward melodrama, or all poems simple and lyrical, or all prose euphuistic or rhetorical; nor is it to say that after 1603 all dramatists were as negative as Cyril Tourneur, all poets as complex as John Donne, or all prose writers as "strong-lined" as Lancelot Andrewes. But, on a general view, the more sombre colouring of the Jacobean period is apparent, and the reign of Elizabeth stands in contrast as a radiant period of expansion and hopeful experiment in literature.

The positive achievement of the Elizabethans was the more remarkable since they lived in an age which had seen nearly all the traditional beliefs and fundamental assumptions of the middle ages shaken, if not overthrown, by the physical and intellectual discoveries of the Renaissance, and the religious, political and social changes following the Reformation. First, the theories of Copernicus suggested that the earth itself, "this little darke starre wherein wee live," stood no longer in a special position in the centre of the universe but was a planet circling in space like any other. Further, the explorers of the 15th and 16th centuries had established that Europe was but a small and crowded corner of this planet. Secondly, the Reformation had called in question the absolute nature of the greatest single creation of the middle ages—the Church of Rome. The mutability of religious institutions was particularly apparent in England since the English reformation had demonstrated, especially in the 1550s, the strength of secular power over the church. Thirdly, Machiavelli in *Il Principe* (*The Prince;* written 1513, published 1532) was held to have superseded the naïve political morality of the middle ages and to have recommended in its place the practice of government by bad faith, terror, hypocrisy, exploitation of the weakness and viciousness of men, and forthright but calculating pride in action. Although no published translation of *Il Principe* existed in English until 1640, and the version of its teaching current in England misrepresented the author's arguments, it was clear that, if only by implication, he questioned the reality of all ethical principles in human conduct. No defense of the traditional morality in the 16th century was as forceful or as influential as Machiavelli's attack. Lastly, despite the wide currency given to favourite medieval authors by the printing press, it was plain that in Europe a major effect of the Renaissance had been an intense interest in classical literature and in the civilizations of Greece and Rome. From this followed the attempt to transcend at once the limits of medieval thought and the divinity of the middle ages, and to shape European culture and civilization by the secular example of the past. Thus humanism assumed that the potentialities of man, especially as an artist, were boundless.

Although these movements, so powerful and so close in time, disturbed the old orthodoxies, leaving Europe in general and England in particular in ferment, from the welter of new visions and new doubts came not a literature of disillusionment but, at least to begin with, a literature of hope and abounding confidence. For, although the world of the medieval Englishman had been disrupted by the great discoveries, by the growing national strength, by the Reformation, and, especially, by the humanism of Europe, these same influences determined the nature and fostered the growth of the Renaissance in England.

The explorers and voyagers of England symbolized the nation's feeling of expansion. Until the time of Richard Hakluyt, however, they did not provide direct literary inspiration, although the Elizabethan poets readily introduced proud images from the exotic new geographies. Similarly the politics of the age, were not directly material for the imaginative writer. Nor, apart from the Great Bible of William Tyndale and Miles Coverdale and Archbishop Cranmer's Book of Common Prayer (1549), did the fierce stir of theological disputation produce enduring works of literature. Although the "flyting" pamphlets, Cranmer's homilies, the plain style of sermons such as those of Hugh Latimer, and the *Book of Martyrs* (1563) by John Foxe, all showed the potentialities of English as a literary language, the importance of the Reformation, like that of the great voyages or the policies of the Tudors, is not to be measured by its immediate effect on individual authors, but by its contribution to national pride. Gradually during the 16th century the English became conscious of their freedom, of their power to expand and to devise their own future in religion and in foreign affairs, and of their own growing eminence among the nations of Europe. On such a people, newly liberated (as it seemed) from all superstitions, endowed with a vernacular capable of sustaining the verse of Chaucer and the prose of Wycliffe, and eager, especially after 1580, to assert its jubilant patriotism, the culture of the Renaissance worked with power.

2. Renaissance Culture.—Elizabethan literature is ultimately distinguished from its medieval and Gothic predecessors by the characteristic leavening presence of Renaissance culture. Although its most important effects followed indirectly from the Renaissance in Europe, it is necessary to note first the direct impact of humanist studies in England.

From 1500 John Colet and his friends, following Erasmus, diffused a sounder knowledge of Greek and Latin, of classical texts, and so of ancient life and thought. This period of humanism in the stricter sense was less brilliant than the humanistic scholarship of Italy or France, but the classical disciplines were instituted and established, and one work at least, Sir Thomas More's *Utopia* (Latin 1516; trans. 1551), exemplified the fundamental process which appropriates the classical mind, not only for purposes of schooling but as a source of new and independent thinking. Colet's work was continued during the middle of the century by Sir John Cheke and Sir Thomas Elyot, by Thomas Wilson in his *Art of Rhetoric* (1553), and by Roger Ascham, especially in his *The Schoolmaster* (1570). These men and many like them put forward the classics as the true staple of a liberal education, and the pattern for a simple yet lettered vernacular prose, an ideal furthered by a series of classical translations which, at first plain and pedestrian, slowly grew in power and elegance and were at last, like the translations from modern languages, written by a succession of masters of English. The rapid growth in size and importance of the two English universities encouraged this direct dissemination of the classics.

An even more powerful force, however, was the indirect impact of the Renaissance, working on English authors through Italian or French humanists. Thus in England the Renaissance was more complex than in Europe, since the knowledge of the philosophy, art and interests of Rome and Athens was as a rule transmitted through France or Italy, tinted and charged in the passage with something characteristic of those countries. The Englishman's knowledge of, for example, Seneca, the Sicilian pastoral poets and classical literary theory was more than half derived from his reading of continental Senecan tragedies, of Jacopo Sannazaro and Clement Marot, and of Italian critics such as G. G. Trissino, A. Minturno and G. Mazzoni. For despite the great influence of France, Spain and Germany, it was especially through Italy that the Renaissance made itself a force in English letters. Not only did the reading and imitation of Italian authors reveal new prosodic and melodic principles and fresh resources of literary form, but from the writings of men such as Pandolfini, from Machiavelli's *Prince* and his memorial to Leo X, and from the example of many a "complete man" such as L. B. Alberti or Leonardo da Vinci, the Englishman derived the ideals of the objective consideration of the state as a deliberate and notable creation of man, and of the full recognition of man as an individual. Thus it was not simply as an assortment of derivative techniques or borrowed theories that the Renaissance showed itself in England, but as a habit of analytic but intensely imaginative perception and thought —the intellectual temper of men who, deprived of their traditional buttresses and windbreaks, found instead in the mind of man infinite promise of achievement and stability.

Older ways of thought were not, of course, immediately and universally abandoned; essentially medieval concepts such as that of the "great chain of being" (*q.v.*) or the correspondences between man, the state and the universe had a flourishing literary and popular life for centuries after the Renaissance. Throughout the reign of Elizabeth strong native traditions of thought and expression worked with and modified the newer modes, to produce a fruitful and vigorous literature, receptive to new ideas but preserving many valuable elements of the English medieval inheritance. Until 1580, however, this blending of English and Renaissance influences could show few works in prose or verse of any permanent value.

3. Poetry Before 1580.—The sudden flourish of lyric poetry at the court of Henry VIII was the only achievement of the first half of the century worthy to be considered part of a true Renaissance. The leaders of the movement were Sir Thomas Wyat and Henry Howard, earl of Surrey. Wyat's short lyrics, mostly on conventional subjects, and sonnets, generally in the familiar

Renaissance character of the despairing lover, show a genuine originality of handling and a dogged pursuing of a line of poetic argument which are most attractive. In his sonnets, as in his *ottava rima*, rondeaus and his satires in *terza rima* (all forms which he introduced to England), Wyat obviously found English scarcely flexible enough to express complex reactions to a situation in the exacting forms which he had borrowed. The measure of his success is the very large number of poems in which, despite the intractability of his material, a single line reveals at once his own poetic genius and the potentialities of the language. Surrey shows in contrast greater technical assurance, but his pages rarely glow with Wyat's sudden splendour. His lyrics seldom bear in their accomplished cadences the marks of poetic travail: his strongest effects are achieved, as in his well-known "The Happy Life," by simple affirmation in emphatic and disciplined verse. His great metrical achievement was his use of blank verse in his translation of books II and IV of the *Aeneid;* the metre is well handled, except that the frequent end-stopping makes the line and not the paragraph the unit, and what variations of stress occur seem due rather to carelessness than to a desire for rhythmical variation. Nevertheless Surrey's importation of the metre from Italy marked the introduction of the cardinal measure of English verse.

The poems of Wyat and Surrey were not printed until 1557, when many of them appeared with poems by Nicholas Grimald, Lord Vaux and others in Richard Tottel's *Miscellany,* the first of a series of collections of lyric verses. Its successors included *A Handful of Pleasant Delights* (1566), *The Paradise of Dainty Devices* (1576) and *A Gorgeous Gallery of Gallant Inventions* (1578). The publication of miscellanies continued throughout the century; one of the best, Francis and Walter Davison's *Poetical Rhapsody,* appeared in 1602. Although the earlier anthologies are agreeable reading, none of the authors, apart from Wyat and Surrey, has more to show than an ear for melody and a rough talent for verse. Nor was the one large literary project of the age, *A Mirror for Magistrates* (1559), successful. It originally contained 19 legends from English history from Richard II to Edward IV "wherein may be seen by example of Other, with bowe grevous plages vices are punished." It was issued with enlargements several times until 1621: the edition of 1563 contained the famous *Induction* by Thomas Sackville, an accomplished piece of writing in rhyme royal. Otherwise the years between Surrey and Sir Philip Sidney can show little that recalls Wyat or anticipates Spenser except delightful single poems by generally undistinguished writers such as R. Edwards, G. Turberville, G. Gascoigne (best-known for his blank-verse satire *The Steel Glass,* 1576), G. Whetstone, T. Churchyard, Sir John Harington, Barnabe Googe, J. Pickering, R. Wisdome, T. Howell, J. Sandford and T. Proctor. Partiality may perhaps excuse the mention of Thomas Tusser, a Suffolk farmer, whose *Points of Husbandry* (1557; often reprinted) convey their thrifty advice in doggerel couplets with an entirely medieval charm, as in the following, for example:

> Who goeth a borrowing
> Goeth a sorrowing.
> Few lend (but fools)
> Their working tools.

4. Prose Before 1580.—Spoken English changed slowly in the 16th century. From the prose scenes of the early plays, from state documents, from the sermons of Hugh Latimer, from letters and memoirs and from works such as William Bullein's *Dialogue . . . Against the Fever Pestilence* (1564) we can learn how the language was spoken, and how gradually it assimilated classical influences.

Those who used language as a literary medium, however, more rapidly adopted Latin ideals of prose writing, but always at first with the intention of strengthening English. Humanists such as Thomas Wilson, Cheke and Ascham were constantly vigilant against pedantic Latinisms and Italian affectations. Cheke, for example, wrote to Sir Thomas Hoby "I am of this opinion that our own tongue should be written clean and pure, unmixt and unmingled with borrowing of other tongues." Some of the best prose of this period is found in the writers on rhetoric, and much also in the works of churchmen and translators such as Sir Thomas Hoby, whose version (1561) of Castiglione's *Cortegiano* (*The Courtier*) is in style and matter one of the most notable imported books between Berners' Froissart (1523–25) and Sir Thomas North's Plutarch (1579). The finest example of a prose rich in Ciceronian cadence and yet unmistakably native is the English prayer-book (1549, 1552, 1559), while the successive versions (1525, 1535, 1537, 1539, 1560, 1568) of the Scriptures provided a noble and enduring pattern for vast heroic narrative and sublime lyrical prose. A fourth important group was that of the chroniclers, antiquarians and voyagers: the chronicle of Edward Hall (1542, 1548) and Raphael Holinshed's *Chronicles of England, Scotland and Ireland* (1577), though rough, make good use of the English plain style, especially the topographical sections of Holinshed, for which William Harrison was responsible. Imaginative prose did not flourish: the collections by William Paynter, Sir Geoffrey Fenton and George Pettie of stories from classical and continental authors were more valuable as sources of stage plots than as models for English novelists. Gascoigne's *The Adventures of Master F. J.* (1573) is a very distant ancestor of the work of the early 19th-century novelist, Thomas Love Peacock, little influenced by the Italian *novelle* and undeservedly neglected. The one work which not only shows the desire to exploit the vernacular but demonstrates the truly humane quality of the writer's mind is Roger Ascham's *The Schoolmaster* (1570), which describes in simple lucid prose the ideal education of a young man of quality.

5. The Novel, 1580–1600.—The first English novels, *Euphues* (1578) and its sequel *Euphues and His England* (1580) by John Lyly, although their plots were improbable and awkwardly handled, were far livelier than the *novelle* translated from Italian, and met with immediate success. They were the fruit of a shrewd and topical observation, they appealed to gentlewomen by presenting an idealized picture of polite society (being derived in part from the Italian books of courtesy), they embodied many sensational or sententious fragments of philosophy, natural history or sociology, they were of a faultless, almost priggish, moral tone, and they were written in a circuitous but rapid style which displayed the most elaborate artifices of the rhetorician. *Arcadia* (1st version 1577–80; 2nd version 1580–84; published, 1590) by Sir Philip Sidney, was also a picaresque pastoral romance, written in a similarly adorned prose, with grandiose phrases used, as in *Euphues*, to dignify a simple act. The plot is extravagant and many of the characters are poorly drawn, but, under the cover of fiction, the book offers a guide to conduct written by a man of unmistakable ardour and integrity. *Arcadia* and *Euphues* were in style and matter extremely influential, although their successors, such as Anthony Munday's *Zelauto* (1580), Robert Greene's *Card of Fancy* (1584) and *Pandosto* (1588), Thomas Lodge's *Rosalynde* (1590), John Dickenson's *Arisbas* (1594) and Emanuel Forde's *Parismus* (1598) and *Parismenos* (1599), were much below them.

Of more realistic novels the best is *The Unfortunate Traveller* (1594) by Thomas Nashe, a brutal and realistic tale of adventure, narrated with speed and economy. Nashe's only rival as a storyteller is Thomas Deloney, who, though he lacks Nashe's fire and gift for the biting phrase, in his *Jacke of Newberie* (1597) and *Thomas of Reading* (1600), wrote subtle and well-constructed novels of everyday bourgeois life. After this brief burst of activity from 1578 to 1600 the English novel had to wait a century before it achieved again even minor importance.

6. Other Prose, 1580–1600.—Although, apart from the novels mentioned above, there was little work during this period for English prose in the service of imaginative literature, it was indisputably the recognized medium for every other form of writing. For example, it was naturally used for works of literary theory and accounts of voyages of discovery, and in each kind notable books are to be found. Of the critical writers Sidney is the most important. His *Defence of Poesie* (1595) was the first work to put forward in English the ideas of the neo-Platonic theorists of the Renaissance. He writes in a lofty but clear style, expressing noble ideas of the nature and function of poetry, with many acute observations on the state of contemporary English

literature, especially the drama. The other writers on poetry or rhetoric, such as George Puttenham, Abraham Fraunce and Francis Meres, cannot rival Sidney in clarity or grace, and are read chiefly for the incidental light they throw on the literary tastes of their age.

The discoverers are well-represented in *Principal Navigations* (1589; expanded 1598–1600) of Richard Hakluyt, a wide-ranging collection of travelers' accounts by various authors, including Sir Walter Raleigh. The narrations are of uneven quality, but the best of them are in a well-turned literary prose, and many of the others in an agreeably plain seamanlike style. A similar use of language is found in the *Survey of London* (1598) by John Stow.

The capabilities of English prose were better displayed in two provinces where the writers were attempting to do something more than convey information. The great theological controversies did not cease with the Reformation; indeed the quarrel with Rome gave way to even more bitter dissension. The two branches of the new Protestant church—the Puritans on the one hand and the more moderate Episcopalians on the other—were in violent conflict, particularly about the relation of civil and ecclesiastical powers in the state. They conducted their disputes in a series of long and learned works, culminating in the work of an Episcopalian, Bishop Richard Hooker, whose *Of the Laws of Ecclesiastical Polity* (books I–IV, 1593; V, 1597; VI and VIII, 1648; VII, 1662) expresses a liberal view of the nature of the Church of England. His style is Latinized but limpid, admirably suited to the persuasive expression of his comparatively tolerant and moderate opinions.

The religious pamphleteers seem almost to burlesque the erudite disputations of the great divines, by conducting the same arguments in a series of scurrilous pamphlets, full of general denunciation and personal abuse. On the Puritan side the most celebrated and the most effective were the seven anonymous "Marprelate" tracts (1589; *see* MARPRELATE CONTROVERSY), which attacked the bishops with great rancour and energy, employing a sharp colloquial prose. The replies to these were also made anonymously, but their authors probably included Lyly and Nashe. Nashe was also engaged in a more personal battle, which grew indirectly from the Marprelate controversy. He (with Greene until his death) exchanged insulting pamphlets at intervals from 1590 to 1596 with the brothers Gabriel and Richard Harvey, each side employing all its resources of sarcasm, ridicule and invective. Many lesser disputants joined issue in a similar way. In addition to his shafts against Harvey, Nashe wrote a number of pamphlets the contents of which are only loosely related to the nominal subject, but his style is so spirited and engaging that the reader does not soon lose interest. Other writers produced pamphlets on an extraordinary variety of topics. The most lively are those which describe in an appropriately informal style the low life of London, partly in admonition, partly in admiration; those of Robert Greene are justly the most celebrated.

Even among such diverse writers a general development may be traced from an ornamental Latinate rhetorical style, often florid, but often eloquent, to a simpler, more flexible prose, much more closely related to English speech and capable of expressing with equal ease the commonplace and the sublime. The prose passages in Shakespeare's plays provide a good specimen of the later style.

7. Spenser and Sidney.—Edmund Spenser and Sir Philip Sidney were the first considerable poets of the queen's reign. Sidney was the greatest of all those Elizabethans who regarded poetry as one accomplishment of the complete courtier rather than as an all-demanding art. In his sonnet-sequence *Astrophel and Stella* (*c.* 1581; published 1591) he attempted and achieved more than any of his predecessors. His sonnets, many of which comment on contemporary fashions in poetry, are compressed, but rarely contorted; his sentences, especially when he writes in Alexandrines, flow naturally, metre and diction unobtrusively accommodated to the rueful irony of his dejection.

> Some lovers speak, when they their Muses entertain,
> Of hopes begot by fear, of wot not what desires,
> Of force of heavenly beams infusing hellish pain,
> Of living deaths, dear wounds, fair storms and freezing fires:
> Some one his song in Jove and Jove's strange tales attires,
> Bord'red with bulls and swans, powdred with golden rain:
> Another, humbler wit, to shepherd's pipe retires,
> Yet hiding royal blood full oft in rural vein.
> To some a sweetest plaint a sweetest style affords:
> While tears pour out his ink, and sighs breathe out his words,
> His paper pale despair, and pain his pen doth move.
> I can speak what I feel, and feel as much as they,
> But think that all the map of my state I display
> When trembling voice brings forth that I do Stella love.
> (*Astrophel and Stella,* VI)

A few years earlier Spenser had published *The Shepherd's Calendar* (1579) "conteyning twelve Aeglogues proportionable to the twelve monethes," a scheme which allowed him considerable room for experiment in form and matter. The work, which was often reprinted in Spenser's lifetime, was notable more for the poet's deliberately archaic and dialectal language than for its poetic merits. Spenser is most successful when his form or his matter impose limits on his fluent genius, as in the April and November eclogues, *Mother Hubberd's Tale,* a satire on contemporary court and clerical life, *Epithalamion* and *Prothalamion,* two poems of intricate and sustained melody, and perhaps the best of his minor poems, *Colin Clout's Come Home Again.*

For his unfinished masterpiece *The Faerie Queene* (books I–III, published 1590; books IV–VI, 1596; part of book VII, 1609) Spenser chose an infinitely looser form, the long poetic romance made popular by writers like Ariosto, and further broadened it to accommodate an extended allegory in the medieval manner. Spenser's explanatory letter to Raleigh, which does not entirely fit the poem as it stands, declares that he designed 12 books, portraying "the twelve private moral virtues, as Aristotle hath devised." The title virtues of the extant books are Holiness, Temperance, Chastity, Friendship, Justice and Courtesy. Each virtue is exemplified by a knight, whose quests form the matter of the book. The heroic adventures are of a kind with which English readers would be familiar from Arthurian legend, as retold by writers such as Sir Thomas Malory—the rescue of maidens from enchanters, the capture of the strongholds of the wicked, tournaments and single combats and the use of magic weapons. King Arthur appears, himself in quest of Gloriana, the "Faerie Queene"; he is to bear some part in the fortunes of each book and thus to acquire experience of all the virtues, standing finally for the "complete man" or Magnificence. Blended with this medieval matter and with the characteristically Renaissance aim of fashioning "a gentleman or noble person in virtuous and gentle discipline," was a third level of meaning, contained in an unsystematic equation of certain incidents and characters in the poem with contemporary events, such as the subjugation of Ireland, and personages, such as, inevitably, the queen. Modern neglect of the poem is a result of its failure as narrative. The adventures of the knights are not sufficiently vivid or coherent to recommend the deeper meaning of the poem, few quests are pursued to a dramatically satisfying conclusion and the final picture is of virtues and vices straying at random over an undefined and ill-lit countryside. The reader has constantly to change focus, seeing the poem first as a struggle between abstract qualities, then as a parable of man's struggle against sin, occasionally almost as an account of the world of Platonic ideas. Although this is not an unrewarding exercise, since passages of effortless description and plaintive reflection abound, the total impact of the poem is divided.

Spenser's handling of rhymed verse is unsurpassed in English, but, like his story, his verse moves slowly and without excitement. His nine-line stanzas give him room to amplify every description and reflection; to this he is so often tempted that the current of his story, never fast, sometimes checks and stands motionless.

> The waies, through which my weary steps I guyde,
> In this delightfull land of Faery,
> Are so exceeding spacious and wyde,
> And sprinckled with such sweet variety,
> Of all that pleasant is to ear or eye
> That I nigh ravisht with rare thoughts delight,
> My tedious travell do forget thereby;
> And when I gin to feele decay of might,
> Its strength to me supplies, and chears my dulled spright.
> (*The Faerie Queene,* Book VI, lines 1–9)

The Faerie Queene then is a region for lovers of poetry to wander in, rather than an uncompleted but entirely conceived poem.

The two surviving cantos of book VII stand by contrast in high esteem, for in them Spenser considers the "pittious worke of Mutabilitie," free from even the loose discipline of his romantic narrative, finding instead an organic unity which gives the two cantos great power and control. He gives voice to the tranquil melancholy which is but the undersong of the earlier books and finds in this fragment of a fragment his greatest stature as a poet.

8. The Successors of Spenser.—The most original of those who succeeded in the manner of Spenser were the brothers Giles and Phineas Fletcher. Giles Fletcher in *Christ's Victory, and Triumph* (1610) retains Spenser's control over a long stanza terminated by an Alexandrine but anticipates Milton in the burning solemnity of his vision of heaven. Phineas Fletcher's *The Purple Island* (1633) is a long and detailed allegory of the human body and the soul. It is in the main more descriptive than narrative and closer to the emblem books (*q.v.*) than to such a work as John Bunyan's *The Holy War*. Fletcher's fine metaphysical intuitions and his command of a condensed version of the *Faerie Queene* stanza make *The Purple Island* the most successful of the poems inspired by Spenser. If Sir John Davies' *Orchestra* (1596) owes to Spenser its graceful rhythms and Platonic radiance, he shows in his *Nosce teipsum* (1599) a deeper concern than Spenser with the spiritual state of man:

> We that acquaint ourselves with every Zone
> And pass both Tropics and behold the Poles,
> When we come home are to ourselves unknown,
> And unacquainted still with our own Souls.

It is significant that each of these poets writes with more fervour than "grave morall *Spencer*" and seeks to win the reader not to the classical virtues but to an active Christian belief. In this they look forward to the great religious poems of the 17th century, for before 1600 Christianity and the Reformed Church had inspired notably few poets.

Michael Drayton may be taken as the type of the Elizabethan poet. He followed Spenser in pastoral, the fashion of the age in the sonnet, Samuel Daniel in chronicle and legend, and Marlowe in mythological story, and yet remained himself. His *Endimion and Phoebe* (1595) in passages rivals Marlowe's *Hero and Leander;* his *Nimphidia* (1627) was the most popular of burlesque fairy poems; and his pastorals, especially *The Muses' Elizium* (1630), are full of graces and felicities. He is best remembered for his odes, "The Battle of Agincourt" and the "Virginian Voyage" which, like his *England's Heroical Epistles,* epitomize the triumphant pride of the age. The work of Drayton that is most often mentioned and least read is his *Polyolbion* (1612–22), a vast and pious effort, now and then nobly repaid, to versify the scenery, legends, customs and particularities of every English county. In his "Epistle to Henry Reynolds" Drayton gives a brief survey in easy-paced verse of the state of English poetry until his own day. In it he says of Samuel Daniel "His rimes were smooth, his meeters well did close, But yet his manner better fitted prose." This is a just account of Daniel's long chronicle poems, such as his *Civil Wars* (1595–1609), but in his sonnets and shorter poems he displays a cultivated and equable artistry. All this patriotic and historical verse, like that of William Warner, Charles Fitzgeoffrey and Gervase Markham, was written during or inspired by the last 20 years of the queen's reign; the same is true of Shakespeare's and most other history plays of the period. Outside the theatre, however, patriotic verse, although full of enthusiasm, started at a disadvantage compared with the love sonnet, song or mythological narrative, because it had no models before it in other lands and therefore always remained less disciplined.

9. Sonnets and Mythological Poetry.—The English love sonnet, introduced by Wyat and revived by Sidney, enjoyed its greatest popularity between 1591 and 1598. In sequences such as Samuel Daniel's *Delia* (1592), Thomas Watson's *Tears of Fancy,* Barnabe Barnes's *Parthenophil and Parthenophe* and Thomas Lodge's *Phillis* (all 1593), Henry Constable's *Diana,* William Percy's *Caelia* and Drayton's *Idea's Mirror* (all 1594), Spenser's *Amoretti* and Richard Barnfield's *Cynthia* (1595), Bartholomew Griffin's *Fidessa,* Richard Lynch's *Diella* and William Smith's *Chloris* (all 1596), Robert Tofte's *Laura* (1597) and *Alba* (1598), especially in the earlier ones, lovely lines, passages and complete sonnets abound. This holds true also for the irregular "sonnets" in Fulke Greville's *Caelica,* the work of an unhappy genius, uneasily conscious that the complete man of the Renaissance was too often fit only for a complete life in this world and neglected his own salvation. In most of these series the form is that of three separate quatrains with the closing couplet for emotional and dramatic climax, a scheme slowly evolved from the stricter Petrarchan sonnet, in which the sestet is at once the complement and the resolution of the octave. Although the sense of personal revelation in these cycles is strong, it is often misleading. While for example the sequences of Sidney, Drayton and Spenser and Shakespeare's *Sonnets* (printed 1609, but probably written in this period) are addressed to definite persons and are charged with sincere poetic emotion, ecstatic or downcast, desperate or aspiring, the three strands of imitation, invention and reminiscence are inseparably intertwined, and now one, now another is most prominent. It is thus extremely dangerous to treat any of these collections as autobiographical records.

The amorous mythological tale in verse derives immediately from contemporary Italy and ultimately from Ovid, whose *Metamorphoses,* familiar in Arthur Golding's version (1565–67), furnished descriptions, decorations and many tales, while his *Heroides,* of which George Turberville's translation went through five editions between 1567 and 1600, provided a model of tragic or pathetic sentiment. Within ten years, between 1589 and 1598, appeared Lodge's *Scillaes Metamorphosis,* Shakespeare's *Venus and Adonis* and *The Rape of Lucrece,* Thomas Edwards' *Cephalus and Procris,* Drayton's *Endimion and Phoebe* and Marlowe's *Hero and Leander.* Marlowe's unfinished poem promised to be the major achievement of the period in nondramatic verse; the common fault of the others is that the love situations of which they treat are not deeply imagined, but used, even by Shakespeare, as convenient occasions for the display of virtuosity. As a result they are in general marked by extreme technical brilliance and by a rhetorical hollowness of sentiment.

10. Lyric Poetry.—As has been said above, the first English lyric poet of the 16th century was Sir Thomas Wyat. His translations and adaptations from the Italian embodied one feature which became almost a convention in all subsequent lyrics. On the one hand the favourite subject was the poet's sorrow—at the cruelty or inaccessibility of a mistress, the pain of parting, the death of a lover, the elusiveness of sleep or the transience of earthly delights. Yet the poems are light and exuberant and the poets obviously enjoyed writing them. This does not point to a lack of sincerity: indeed the strength of many of the lyrics lies in this very tension (seldom openly ironical) between the subject and the manner. Further characteristics, again inherited from Wyat, are the immediacy of the subject and the concreteness of the imagery in the lyrics. They reflect every side of Elizabethan life, from war, exploration and trade to the life of the town and home, feasting and dancing, gardens, books and tobacco. The poets clearly enjoy their world and are not deterred by fear of seeming unpoetical from using the stuff of their everyday affairs as material for their lyrics, which are thereby invested with extraordinary solidity.

> The Andalusian merchant, that returns
> Laden with cochineal and china dishes,
> Reports in Spain how strangely Fogo burns
> Amidst an ocean full of flying fishes:
> These things seem wondrous, yet more wondrous I,
> Whose heart with fear doth freeze, with love doth fry.

(Anonymous: published in T. Weelkes's *Madrigals of 6 Parts;* 1600)

The introduction of Italian music, with its favourite metrical schemes, powerfully schooled and coloured lyric. The Italian poems were often translated in their own measure, line by line, and the musical setting retained: these tunes were then coupled with new and original poems. Both kinds of lyric appeared together in the songbooks of John Dowland the lutenist, William Byrd, Thomas Morley and especially those of Thomas Campion. Campion, who was also a musician, published most of his poems in

songbooks such as *A Book of Ayres* (1601). The poems themselves are usually of simple structure with little of the "turn" that is characteristic of, for example, the sonnet. Moreover, the words are chosen to be sung and are therefore clear and melodious. Yet the lyrics, far from being insipid, find in Campion's predominantly elegiac tone the perfect foil to their own imperishable grace and harmony.

> All that I sung still to her praise did tend,
> Still she was first, still she my songs did end;
> Yet she my love and music doth both fly,
> The music that her echo is and beauty's sympathy.
> Then let my notes pursue her scornful flight:
> It shall suffice that they were breathed and died for her delight.
> (*A Book of Ayres*, X; "Follow Your Saint," lines 7–12)

Much of the best lyric of the period appeared in songbooks, or in the miscellanies which, after Tottel, were published regularly. *The Phoenix Nest* (1593), *England's Helicon* (1600) and Davidson's *Poetical Rhapsody* (1602) contain poems of a surprisingly high and even quality, which serve to demonstrate the wide distribution of the lyric gift, the assured command of most poets over the English vernacular and a sensuous appreciation of the pleasures of Elizabethan life.

A third source of lyric is in the plays of the period. From the time of Lyly it had been customary for playwrights to diversify their action with a song, which, while providing a pause in the action, maintained the dramatic atmosphere. For their full effect the songs thus require their context, yet even in isolation they are of singular power and beauty. Shakespeare is here, of course, pre-eminent, but Lyly, Peele, Greene and Dekker, to name only four, had an art as fine, though slighter.

Everywhere from 1580 to 1600, and intermittently thereafter, the nation, already notable for its love of music, found words for its lovely airs and madrigals. The poems of men such as Lodge, Sir Walter Raleigh, Nicholas Breton, Thomas Watson, Thomas Nashe, John Donne (in his earlier poems) and Henry Constable are among the happiest and most durable monuments of the Renaissance in England. Their realization of the difficulties inherent in living only for this world was only deep enough to tinge their lyrics with an appropriate sense of the urgency of seizing the world's delights. It was not until after 1600 that disillusionment with the material prosperity of England, with the slack restraints of the Reformed Church, and, finally, with certain humanistic ideals influenced poetry deeply. Just as there was scarcely a poet, except William Baldwin (one of the poets who planned and contributed to the *Mirror for Magistrates*), the Jesuit Robert Southwell and occasionally Campion, concerned in his poetry with the state of his soul, so there were few satirists concerned with the state of society. Wyat in his admirable *terza rima* had, as so often, shown the way and established the manner, but his successors in the 16th century were heavy-handed, preferring a general complaint to a precise attack.

Despite the large number of writers (for every man seemed to turn his hand to lyric and the chance of immortality) the general impression of Elizabethan verse of the great period is uniformly high. No doubt, as admirers of the later Metaphysicals claimed, the lyric became gradually overconventional, its diction lost immediate force and its rhythms were in the end too often determined by musical necessity and not demanded by the poem itself. Yet at its matchless best it fittingly expresses the temper of Elizabethan England, half springtime, half golden age.

A. Elizabethan and Jacobean Drama

1. The Transition From Medieval Drama.—The development of the drama in the 16th century was prolific but confused. Out of the morality play, a reverent enactment of religious beliefs, had grown up the shorter interlude, which was usually a social or ethical debate in an approximately realistic setting between characters representing different types or trades, and was often lightened by "divers toys and gestis," such as the comic play of a "vice" or incarnation of grotesque roguery. Though the interlude as such continued throughout the century, later examples were little more than pedestrian dramatizations of economic or political arguments. The value of the interlude was most clearly seen when its separate elements broke away to form new dramatic compounds. For instance, when the interlude dealt with issues such as man's duties to his king it clearly involved affairs of state, and suggested a possible way of writing a historical play. When it drew a moral lesson from the fall of a king or great man through ill-fortune or as retribution for ill deeds it clearly approached tragedy, yet if it showed the good rewarded and the wicked confounded it was closer to comedy. The "vice" himself is a distant ancestor of the Machiavellian prince (such as Richard III), the ogrelike villain (such as Marlowe's Jew of Malta) and the scheming servant whose intrigues spin the plot of many later comedies.

The simple medieval idea of tragedy as the rise and fall of a great man was complicated by the Senecan plays of Italy and by those of Seneca himself (all translated by 1581). All Seneca's plays were moral melodramas of violent passion; they always featured revenge, normally bloodthirsty, often ghosts, and the philosophy was stoic. Direct English imitations, such as *Gorboduc* (acted *c.* 1561) by Sackville and Norton were poor things, decorating a chain of slaughters and revenges with a scattering of philosophical tags. The Senecan form was not barren in itself, as its sequel in France, where it culminated in the heroic tragedies of Corneille and Racine, was to show; it was barren only for England. The Senecan tradition was, however, a vital influence on the later playwrights, for it set the pattern of a high tragical action, often motivated by revenge, swayed by large ideas of fate and retribution and told in blank verse; and it bequeathed, besides many moral sentences, such points of stage mechanics as the ghost, the chorus and the play within the play. In short it enriched, rather than purified, the older native traditions of tragedy.

The high comedy of the first purely secular play in English, Henry Medwall's *Fulgens and Lucrece* (*c.* 1496), had inspired no immediate successors, but from about 1550 various academic writers produced comic plays in English. They accepted from the Latin playwrights, especially Plautus, such disciplines as the five-act structure, the unity of action and consistency of characterization, but, like the writers of the interludes, they chose native themes and treated them with a fair degree of realism. The most celebrated comedies of the middle of the century were *Ralph Roister Doister* (probably before 1553) by Nicholas Udall and the broad rustic farce *Gammer Gurton's Needle* (1559). These rely in the main on English lower-class humour, but introduce many classical elements. Other comedies of note in the period are George Gascoigne's *Supposes* (1566), which employed a light flexible prose to convey a Terentian story of intrigue, and the romances *Damon and Pythias* (1564) and *Palamon and Arcite* (1566) by Richard Edwards, which set the fashion for extravaganzas of chivalrous adventure.

By 1580 a number of chronicle plays and historical interludes such as John Bale's *Kynge Johan* had already appeared, and thus all the major dramatic forms had been foreshadowed. But as Sidney pointed out in his *Defence of Poesie*, they were not clearly distinguished and were often ruinously entangled. While it was to be a characteristic of the English drama to prefer in general mixed modes to a classical purity of tone, the clearer differentiation of the various forms which took place after 1580 was wholly beneficial.

2. Plays of Lyly, Peele, Greene and Kyd.—Of the playwrights who flourished between 1580 and 1595 many were distinguished by their elegance of expression and many by their ability to write powerful pieces for the theatre, but few by both. A notable group of dramatists was that known as the "university wits," of whom the most polished writer was John Lyly, the author of the novel *Euphues*. His plays are usually slight, even anecdotal, renderings of classical legends or fanciful superstitions, such as *Campaspe* (1583) and *Endymion* (1586–87), written with a close regard for the unity of action, their delicate structure distorted only by ingenious deferential allusions to the queen. Lyly was the first to employ with assurance many devices which later became conventional—the ready introduction of light but poised lyric, the handling of a comic subplot as a foil to romance, the exploitation of the ambiguity of the sexes as a pivot of the action, and above all the use for high comedy of a highly wrought

prose, mannered but flexible and equaled in its kind only by the best of Shakespeare and Congreve. George Peele and Robert Greene, though less precise craftsmen, were instrumental in the development of dramatic blank verse, and each in his way contributed to the tone of idyllic romance which characterized later comedy, Peele in *The Arraignment of Paris* (1584) and *Old Wives' Tale* (1595), Greene in the *Honourable History of Frier Bacon and Frier Bongay* (1594).

The work of Thomas Kyd was well suited to the taste of the time. His one important play was *The Spanish Tragedy* (acted *c.* 1587 or 1588), in which, drawing freely on Seneca and Machiavelli, he established the theme of revenge which ran like a crimson thread through later tragedy. The play renounces subtlety for a series of vivid and terrifying *coups de théâtre,* until at last the wronged Hieronimo takes a spectacular revenge and dies in his moment of triumph. *The Spanish Tragedy* was as popular as it was influential; Kyd's other work is of less consequence, although it has been maintained that he wrote a play of *Hamlet,* now lost, on which Shakespeare worked.

3. Marlowe.—More important than any of these writers was Christopher Marlowe, author of the two parts of *Tamburlaine* (acted 1587–88), *Dr. Faustus, The Jew of Malta, Edward the Second* and probably of *Dido, Queen of Carthage* (published 1594). They are in the widest sense morality plays, in which the problems of right conduct in an expanding society are illuminated, if not solved, by their embodiment in figures gigantic rather than heroic. Marlowe, often himself thought of as a typical product of the Renaissance, saw more clearly than most the strength and the weakness of Elizabethan ambitions:

> Nature that framed us of foure Elements,
> Warring within our breasts for regiment,
> Doth teach us all to have aspiring minds.
> (*Tamburlaine,* part I, act II, scene 7, lines 18–20)

The minds of Marlowe's heroes aspire to boundless dominion, forbidden knowledge, ideal beauty or "infinite riches in a little room" and their aspirations are expressed in "great and thundering speech." The verse is of unprecedented force, occasionally bombastic or rhetorical, but as often rising to a classical serenity and inevitability. Although Goethe praised the plan of *Faustus,* Marlowe never reveals a great talent for dramatic construction, yet his plays are sustained and borne irresistibly forward in the theatre by the unfailing resource of his magnificent verse and the splendour of his imagery. Marlowe established blank verse beyond question as the medium of the heroic drama for the next 50 years; he established also the nature of most plays in the same period by exalting spontaneity above careful construction, the exploitation of convention above realism, and power above subtlety. Above all he established the Elizabethan theatre as the true home of the poet.

4. Shakespeare.—*The Histories.*—Perhaps the first poet, certainly the greatest, to combine a keen appreciation of the exhilarating freedom of the Renaissance with a profound intuition of the spiritual needs and potentialities of man was William Shakespeare. At first there was little sign of this. Apart from *Titus Andronicus,* a tragedy of blood, strongly influenced by Kyd, his first plays were all histories or comedies, little different from those that already held the London stage. *Henry VI* (three parts, 1590–91) broke no new ground; it shows a command of rhetorical verse and an eye for a dramatic or pathetic situation, but it has not the vigour of its successor, *Richard III* (1592–93). These four plays form Shakespeare's first historical cycle. His later histories *Richard II, Henry IV* (two parts) and *Henry V* (all written by 1599) show an advance not only in the mechanics of the theatre and in the handling of dramatic blank verse, but in a far more important direction—the power to transcend the mere subject of a play, reaching beyond it to a larger view of the world. In the eight plays, which cover a period of nearly a century, Shakespeare used the well-known tragic histories of the Plantagenet kings in much the same way as the Greek dramatists had used the stories of the house of Atreus, to show the transmission of evil from one generation to the next. He thus offers not only a chronicle of a medieval dynasty but a contribution to the popular Renaissance debate about the right government of nations. The plays differ sharply in dramatic technique. While *Henry V* is more of a patriotic pageant than a play and *Richard II* may be regarded as a morality demonstrating the irony of kingship, *Henry IV* is something more. The creation of Falstaff (the touchstone of all comic characters) and the way his life in Eastcheap is set against the king's life in the court, mark a new complexity in the drama, and show that Shakespeare was exercising upon the past of England not simply the historian's power of reconstruction but the creative powers of a poet, compact with imagination and invention. He wrote two more history plays: *King John* (?1596–97), a mixture of sentiment and rhetoric in which a few scenes of great power stand isolated, and *Henry VIII* (1612–13), his last play, a formal and complimentary attachment of the Plantagenet cycle to the reign of Elizabeth.

The Comedies.—Shakespeare's early comedies showed his willingness to follow any fashion which seemed profitable. Certainly in *Love's Labour Lost* (?1594) he showed his admiration for the courtly comedy of Lyly, with its delight in bandying words and its lyric tenderness. Yet he excels Lyly in human interest, in the lively pace of his play, and in his more diversified diction. Similarly, since Greene had in his earlier comedy *Frier Bacon and Frier Bongay* successfully exploited a mixture of sentiment, comedy, sensation and mystery, Shakespeare in *Two Gentlemen of Verona* (?1594–95) was happy to continue in the same style; but he developed Greene's occasional interludes of horseplay into a subplot closely connected with the main action. *The Comedy of Errors* (?1592–93) is likewise of the same genre as *Ralph Roister Doister,* by Interlude out of Plautus, yet it is unmistakably the work of a poet, however inexperienced. The first play which Shakespeare wrote in a form new to the English stage was *A Midsummer Night's Dream* (1595–96). In this he developed the pattern of graciousness, gentleness and magnanimity in the main plot set against the humbler virtues of honesty and conviviality in the subplot which he followed in *The Merchant of Venice* (1596–97) and in his three mature comedies *Much Ado About Nothing, Twelfth Night* and *As You Like It* (1598–1600). These represent at once the greatest achievement of the English stage in romantic comedy and the first incontestable proofs of Shakespeare's poetic and dramatic genius. Each comedy, however loose the action, has a unifying homogeneity of tone: the threat of tragedy, or a hint of the uncertainty of fate, especially in love, is usually in the background to infuse a strain of elegiac tenderness or an ironic contemplation of the transience of gaiety and happiness, while the more robust subplots effectively prevent the plays from becoming mere exercises in sentiment. The idea of comedy which animates them is tolerant rather than corrective. Above all, what distinguishes these plays from earlier and later English comedy is that they are written mainly in blank verse. What Surrey had introduced, Peele and Greene had developed and Marlowe had all too briefly mastered, Shakespeare used as the medium of his greatest work. He varies it with expert prose, courtly or vulgar as the action requires, but his staple is verse of unprecedented flexibility, in which constellations of linked images form coherent and significant patterns.

It has been suggested that by 1600, as certain deficiencies of Renaissance humanism became apparent, there was a general darkening of the tone of literature: verse became less conventional and less romantic, prose less Ciceronian and "copious," comedy less sentimental. Shakespeare's later comedies, *All's Well That Ends Well* (1602–03) and *Measure for Measure* (1604–05), strain continuously against the conventionally happy ending of their plots. The motives of the characters are less elevated, the verse moves more haltingly, the images are more sombre, the impact is wholly ambiguous. These plays, like *Troilus and Cressida* (1601–02), are closely related, in the predominantly cynical view they take of fate and human nature, to *Timon of Athens* (1607); they have neither the radiance of Shakespeare's other comedies nor the sublimity of his tragedies.

The Tragedies.—Shakespeare's second tragedy, *Romeo and Juliet* (1594–95), written at about the same time as *A Midsummer Night's Dream,* shares that play's tone of youth and hope. Shake-

speare's conception of the ends of tragedy had clearly deepened before the later tragedies, *Hamlet* (1600–01), *Othello* (1604–05), *Macbeth* and *King Lear* (1605–06), *Antony and Cleopatra* (1606–07) and *Coriolanus* (1607–08). They are written in the blank verse of which Shakespeare is the master. Relieved and heightened as before with passages of prose, it binds each play into a unity which, while always dramatic, is close to the unity of a poem in which thought and its expression are one. His comedies move with greater ease, even, as Dr. Johnson said, with greater naturalness, but only in the tragedies can be seen the words white-hot from the poet's brain and glowing with the hammer-strokes of his imagination. At times, as in *Antony and Cleopatra*, his language appears to break the bounds of sense, and metaphor to have run mad. Yet Shakespeare never loses sight of the "wooden O" of the theatre, and even in *Antony* the poem and the drama are coterminous. There are no standards by which to measure the tragedies except the work of the Greek dramatists, nor can the plays be reduced to a formula. From Romeo's

> O here
> Will I set up my everlasting rest:
> And shake the yoke of inauspicious starres
> From this world-wearied flesh

to Lear's

> You do me wrong to take me out o' th' grave,
> Thou art a Soule in blisse, but I am bound
> Upon a wheele of fire, that mine own teares
> Do scal'd, like molten lead,

the tragic heroes are more than studies in flawed virtue: they are stations in the life and death of Everyman, man as the star-crossed lover, as the cause of pity, horror or terror, as the victim of anxiety or pride. The tragedies explore the resources of humanity, reflecting not only the burning curiosity of the Renaissance but, in their compassionate, almost mystic, conjectures about the whole business of mankind, a completeness of vision and a charity which transcend all categories and periods.

The Last Plays.—From 1608 Shakespeare showed an increasing interest in plays in which the action contained all the potentialities of tragedy and many tragic emotions, but in which tragedy was at the last averted, usually through supernatural powers. *Pericles, Cymbeline, The Winter's Tale* and *The Tempest* (1608–12) abound in violent emotions, crimes and disasters, yet in each play characters thought to be dead are found to be still living, parents and children are reunited, "Jack hath his Jill," and the piece ends to the strain of heavenly music, so that the final emotion is akin to Miranda's wonder: "How beauteous mankind is! O brave new world . . ." It seems probable that Shakespeare wrote this kind of play because he wished after his exploration of the fatal passions of men in the tragedies to explore the complementary themes of restoration and reconciliation. The plots of the plays are unrealistic and even inconsistent, widely deployed in space and time, suggesting a comprehensive vision of man's place in the world.

Yet although Shakespeare wrote on cosmic themes and saw more deeply than his contemporaries, his plays are of their time. He used whatever came to hand—other men's plays, the explorers' tales, dislike of Jews, fear of the pope, the state of the theatres, the presence in the company of a talented boy-actor or of an accomplished theatrical bear—so that his plays are in the widest sense topical. Occasionally we could wish him to be less concerned with the theatre; a haste in some passages, an impatience with the heaviness of language in others, a love of a well-chased quibble, a desire for momentary surprise—all these faults, which we should willingly see "blotted" (to use Ben Jonson's word), are the price Shakespeare paid for his audience, an audience whose receptivity and subtlety were the vital challenge to his genius. It is his power to please in the theatre no less than his power to move in the study that places him on his solitary height of unquestioned greatness.

5. Jonson.—Shakespeare's only rival among his contemporaries was his friend Ben Jonson, whose sardonic comedies were a lasting influence on English drama. Jonson, looking back to the classical playwrights, condemned romantic Elizabethan comedy for its fairy-tale plots, its amorality and its mingling of dancing, music and drama. He learned dramatic practice from Plautus, doctrine from Aristotle's *Poetics* and from Horace, and consistency of characterization from the theory of humours (*see* HUMOUR). This theory, essentially a caricature of human motives, ignoring all but the most obvious, is the mainspring of his early comedies *Every Man in his Humour* (1598) and *Every Man out of His Humour* (*c.* 1599). The plays were successful and the method admired, but Jonson modified the theory in his later comedies, thus allowing more scope for development of character. *Volpone* (1606), *Epicene, or The Silent Woman* (1609) and *The Alchemist* (?1609), Jonson's major works, were designed to anatomize and correct the vices of his own time, especially vanity and greed. Unlike many Elizabethan dramatists, who happily borrowed and adapted whatever story was at hand, Jonson carefully created a series of plots of great ingenuity, moving to a calculated climax of exposure and denunciation. He divides men into cheats and gulls. The gulls have a false view of themselves, the cheats a low view of mankind. Jonson constructs his plots to demonstrate the villainy of the cheats, and then, using their confidence tricks as his instruments, as it were, he lays bare with a deadly expertness the true nature and motives of the hypocrites, misers and braggarts who are caught by the bait. So contemptuous is his satire of the self-deluding folly of the dupes that his attack on the straightforward depravity of the cozeners is by comparison sympathetic. His Roman plays, *Sejanus His Fall* (1603) and *Catiline His Conspiracy* (1611) show the same painstaking application, but, lacking the "rage and spleen" of the satires, they appear formal and contrived. Although his later comedies, such as *The Devil Is an Ass* (1616) and *The Staple of News* (1626), are less neatly constructed, they have abundant life in the theatre. The unflagging prose of his broader comedies such as *Bartholomew Fair* (1614) has been rightly praised, yet his finest technical accomplishment is the blank verse of his high comedies, *Volpone* and *The Alchemist*. Jonson was very much an Elizabethan in the delight he took in language, and having something of Marlowe's power over the "drumming decasyllabon," he achieved in the metre a distinctive metallic splendour, admirably fitted for his hard and energetic comedy.

6. Other Jacobean Tragedy.—Apart from Jonson, Jacobean drama is represented by its tragedies, or rather by plays which induce all the terror of tragedy but little of its pity, plays of which one line from John Webster may stand as the epigraph: "A dead man's skull beneath the roots of flowers." It has been maintained above that the traditional contrast between Elizabethan confidence and expansiveness on the one hand and Jacobean pessimism and doubt on the other is to be observed in particular in matters of literary style and taste. The more sombre tone is evident in the satiric comedy of Jonson and in the tragedies and dark comedies of Shakespeare, although he resists it in his romances: it is even plainer in the work of the other Jacobean dramatists, many of whom were fascinated by the doctrines of Machiavelli in the form in which they were then current in England. Such a view of humanity, while it makes for powerful dramatic situations, sets its own problems. It is when their characters are at the point of death that the Jacobean tragedians are confronted with a dilemma which was implicit in all Renaissance literature. They search continuously for a compromise between humanism and the desire for immortality, for a refuge from the fear of death, but they find none, nor any solution which does not require the creation of a world beyond the visible world. Their heroes thus die either in tragic defiance or in bewildered doubt or in stoic resignation. This last derives directly from Seneca whose influence throughout this period is strong and continuous not only, as in Elizabethan times, for his repertoire of theatrical devices, but as a tutor in the language of disillusionment.

The plays are fairly termed "revenge tragedies," since the dominant motive in all of them is the desire to exact violent and ingenious vengeance for a real or imagined injury which the conventional forces of order are unable or unwilling to punish. Crime is thus heaped on crime, each avenger becomes "the deed's creature," and the last act sees the spectacular death of all those who have been contaminated by sin. All the plays emphasize man's

viciousness or helplessness: they are explorations of the mental climate of societies in which moral standards have no relevance, virtue is weakness, pride is the only source of strength, and individual ambition or caprice the only valid motives for action.

Only John Webster in *The Duchess of Malfi*, (1613–14) and *The White Devil* (1611–12) offers more than his material of murder, madness, torture and treachery. Yet each of his plays is a night piece, ending in despair, with only the vaguest hints that men are anything but "the stars' tennis balls, struck and bandied| Which way please them." The most positive thing in the plays is Webster's poetry, which even when its subject and theme are most despairing glows with an iridescent light, illuminating the dark places of his mind and offering some promise that beauty can exist even in a charnel house.

> I am puzzled in a question about hell;
> He says, in hell there's one material fire,
> And yet it shall not burn all men alike.
> Lay him by. How tedious is a guilty conscience!
> When I look into the fish-ponds in my garden,
> Methinks I see a thing arm'd with a rake,
> That seems to strike at me.
> (*The Duchess of Malfi*, act V, scene 5, lines 1–7)

Although Cyril Tourneur's *The Revenger's Tragedy* (published 1607) is concerned with the same theme as Kyd's *Spanish Tragedy*, that of the wronged man, Tourneur, from the opening scene in which Vendice apostrophizes his dead mistress' skull, gives revenge a much more sinister interpretation. Everything in the play, even the seat of justice, is vicious: Vendice, who is to cleanse this foul world, is however himself sinning in taking his murderous revenge, and the carnage ceases only with the death of the avenger and all his victims. By contrast *The Atheist's Tragedy* (published 1611) is pervaded by the idea that private revenge defeats its own ends. The injured man therefore forbears to attack the Machiavellian villain, who in the end dies by a grotesque accident, leaving again the impression that morality asserts itself only by chance if at all. In both plays Tourneur writes a "strong-lined" verse, rich in macabre and shocking imagery; in neither does he offer a view of man which is more elevated than that of his corrupt villains.

Thomas Middleton, similarly, in his great play *The Changeling* (1621–22) saw the dramatic possibilities of a series of murders and the horror that strikes to the soul at the prospect of further crimes and degradation. ("Vengeance begins;|Murder, I see, is followed by more sins.") His heroine and her hireling lover are portrayed in all the complexity of their relationship with an insight and psychological thoroughness unique in the literature of the period. In the subplot, which is unrelated to the rest of the play, Middleton displays the gifts of satire and irony which he used to powerful effect in his comic plays and in his tragedy *Women Beware Women* (?1621).

The work of John Marston is patently that of a less gifted poet. His most famous work, *The Malcontent* (1604), although it uses all the apparatus of Jacobean revenge tragedy, is singular since the wronged man kills none of his victims but preserves them all in suspense for the denouement, when the minor villains reform and the archvillain is expelled with ridicule. The play, with its characters from the comedy of humours, is thus almost a satire on the tragedy of blood, yet Marston's odd, cramped verse has the unmistakable accents of seriousness. The works of George Chapman are crammed with poetry of the spacious majesty to be expected from the translator of Homer. Many of his dramatic speeches are indeed only imperfectly intelligible in the theatre, and in the study are seen to be a riot of fine figures and involved syntax. His historical tragedies on French themes, full of sudden action and stretches of supremely undramatic verse, are the highest expression in English drama of the Stoic philosophy. Philip Massinger, on the other hand, was a craftsman of the theatre. His tragedy, *The Roman Actor* (1626), stands second only to Webster's two plays as a study in the corruption of princes, a demonstration of what happens when the Machiavellian has attained supreme dominion. John Ford maintained until the Puritan ordinance closed the theatres in 1642 the tradition of tragedy wrung from the extremes of human emotion and the depths of human depravity. *The Broken Heart* and *'Tis Pity She's a Whore* (both written 1627–38) demonstrate the inevitability of fate and the pettiness of man's defiance, even as, like Ford's historical drama *Perkin Warbeck*, they recall Webster's ability to extract beauty from despair.

In all these plays the authors faced the problem of "placing" the acts of their characters at a time when no code seemed strong enough to sustain the action. Yet, if we deplore the moral confusion of Ford and Webster, the nihilism of Tourneur, Middleton's lack of cohesion and Marston's lameness, Chapman's lack of drama and Massinger's lack of fire, we must not forget that, for all their bewilderment, these men wrote poetry of a high and difficult order, piercing, like Marlowe and Shakespeare, in one line of verse to the springs of human conduct.

7. Other Jacobean Comedy.—A similar change of tone may be traced in comedy, where *The Shoemaker's Holiday* (1600) by Thomas Dekker is, apart from the romances, perhaps the last play without overt social criticism: like his other comedies, and those of Thomas Heywood, it is sentimental in plot and realistic in setting. With Dekker and Heywood may be named the troop of professional journeymen, who did not lack power but wrote usually in collaboration, like Anthony Munday, Henry Chettle and Michael Drayton, or supplied, like William Rowley, underplots of rough, lively comedy or tragedy. A more considerable comic writer was Middleton, who owed much to the use by Jonson and Chapman of comedy as a foil to satiric commentary. Of his strong comedies *A Chaste Maid in Cheapside* (1611) offers a particularly penetrating examination of matrimony by bringing together in an ingenious plot couples with different opinions about the purposes of marriage. It is, like the bulk of Elizabethan and Jacobean comedy, not for the polite. Most of the humour is obscene, a taste of the time, as may be gathered from, for instance, the scurrilous pamphleteers. Since, however, many of the jokes are plays upon words of which one meaning has become unfamiliar, most of the more licentious passages are sheltered for modern readers and audiences in the decent obscurity of a virtually obsolete language. The shifts of vulgar speech have scarcely affected Massinger's famous comedy *A New Way to Pay Old Debts* (1622) which is as firmly built as any play of Jonson's. If Massinger lacks the final gift of sustained poetry, he shows here and in *The City Madam* (1632) his command of an easy and natural blank verse, and the ability denied to his greater contemporaries to imply moral values through the action of his plays.

The work of Francis Beaumont and John Fletcher runs through all the available forms from riotous topical burlesque, as in *The Knight of the Burning Pestle*, to pathetic melodrama, as in *The Maid's Tragedy*. Most of the many plays with which they are jointly credited are by Fletcher. The plots are generally romantic in the extreme, while the over-flexible verse is a fit medium for characters who change their natures so conveniently and dramas which veer so readily away from tragedy to the imposed happy ending. The tradition of tragicomedy persisted in the fluent writings of James Shirley, in whose work the decadence remarked in Fletcher is even more apparent. Like Richard Brome, Henry Glapthorne and Nathan Field, he is happiest in the realm of light comedy, where his best work, such as *The Gamester* (1633) and *The Lady of Pleasure* (1635) looks forward to the comedy of the Restoration. In general it may be said that the strongest comedies in the period were Jonsonian and satirical, while, where the playwrights favoured the romance, their work became progressively more sentimental and diluted.

8. Minor Plays and Masques.—A last group defies classification; powerful domestic tragedies, such as *Arden of Feversham* (printed 1592), chronicle plays, melodramas, citizen comedies, academic entertainments such as the Parnassus plays, pastoral plays in the Italian style and Senecan dramas intended only for the study. Many of these plays are anonymous and many unprinted; many are crude and many insipid; yet they show that the great vitality and esteem of the drama in this period was not confined to the work of the major playwrights. A final word must be given to the masque (*q.v.*), a genre which, in its search for pure plastic beauty, called for the most exquisite combination

of lyric, music and setting. From Jonson and Inigo Jones onward the country was enriched by a series of masques which, against a background of tragedies suffused with blood and comedies of sardonic and angry laughter, stand as a graceful memorial to the Renaissance. (P. Dw.)

B. Early 17th Century: Prose and Poetry

The inward and outward climate of the 17th century, a climate of unceasing disturbance and violent storms, was the result of many converging and conflicting forces. The great question of the century, which confronted serious writers from Donne to Dryden, was Montaigne's "What do I know?" or, in expanded terms, "What are the grounds and relations of knowledge and faith and reason and authority in religion, metaphysics, ethics, politics, economics and natural science?" Indeed, few ordinary citizens were untouched by conflicts that were fundamental in the most diverse spheres of thought and action, from the shop and tavern to the house of commons, from the pulpit of St. Paul's to the hillside conventicle, from the several parts of Great Britain to the growing communities in America, from the scholar's study to the battlefields of the Civil War, from Henry Vaughan's mystical raptures over cloud or flower to Thomas Hobbes's materialism, from Bacon's vision of aspiring man's conquest of nature to Milton's vision of Satan's conquest of aspiring man.

1. Character of the Prose.—Although so much of the century's prose forms a major portion of the treasury of great English, very little of it was written as "literature." Most of the many books that remain alive—as well as the many more that do not—were works of instruction and edification. There was, to be sure, a fairly abundant and heterogeneous literature of entertainment, such as printed plays, books of characters, pictures of London life (like those of the humorous or macabre Dekker), and so on (fiction, which would bulk so large in later times, was meagre in quantity and quality and chiefly in the vein of romance); yet the great mass of books and pamphlets, whatever else they proved to be, were utilitarian in origin, as the categories to be followed here will indicate.

In the theory and practice of prose style the century witnessed a signal evolution. During the 16th century prose had moved uncertainly toward such a disciplined maturity as appears in the stately but flexible Ciceronian periods of Richard Hooker. The distinctive movement of the 17th was the anti-Ciceronian, inaugurated, appropriately, by the scientific and realistic Bacon. Whereas the rotund, balanced Ciceronian style was a natural vehicle for accepted truth, the skeptical spirit of the new age found more congenial models in the pregnant irregularity of Seneca and Tacitus. Such prose, whether succinct or loose, was a more suggestive, informal medium for the actualities and questionings of individual experience and reflection. This general impulse received a special stimulus from science. While Elizabethan writers had thought and felt in metaphor, in prose no less instinctively than in poetry (the supreme example in both is Shakespeare), science promoted "the dissociation of sensibility"; the scientific mind tended to outlaw metaphor and image (unless as illustration) and to seek precisely denotative language for things and ideas. Thus the varieties of figurative and emotive eloquence exemplified by the prose of Donne and Milton, Sir Thomas Browne and Jeremy Taylor gave way, in the Restoration period, to a clear, urbane efficiency. But the evolution from the Ciceronian march to "the Senecan amble," from half-poetic to "civilized" prose, cannot be seen as a neat chronological or even logical pattern; varieties of temper and subject matter prevented that. Besides, at all times plain prose was being produced by travelers, men of affairs and pamphleteers, who, innocent of theories, wrote in the pungent language they spoke.

2. Translations and Books of Travel.—This age was, among other things, the great age of translation, and "translators' English" was as yet unknown. Philemon Holland, translating chiefly from the classics, combined scholarship with colour. John Florio's translation of Montaigne's *Essays* (1603), Thomas Shelton's *Don Quixote* (1612–20), and Sir Thomas Urquhart's *Rabelais* (1653–93; completed by P. A. Motteux) were done, if not with strict fidelity, with a racy gusto that has made them English classics. The greatest of all English translations, the King James (Authorized) Version of the Bible, was in a large degree the work of a succession of 16th-century translators (and hence was somewhat archaic even in 1611), yet the committees of Jacobean translators must share the glory through their discriminating taste in phrase and rhythm. Modern versions, though they embody great advances in learning and accuracy, all suffer immeasurably by comparison. To leave religion out of account, it is impossible to begin to estimate the cultural effects of the century's saturation in the English Bible.

Serious observation of foreign countries was a conscious duty of the Renaissance humanist, and that motive was not lost in the growing aristocratic obligation of the Grand Tour. People who stayed at home found diversion and information in multiplying books of travel. The liveliest reading is provided by such humble and invincibly curious globetrotters as the eccentric Thomas Coryate and William Lithgow and the more sober Fynes Moryson and Peter Mundy. Of the value of such narratives to imaginative writers *The Tempest* is one example; and George Sandys' *Relation of a Journey* left traces in *Paradise Lost*. Accounts of Virginia and New England were likely to be in the way of "promotion literature"; among the many absorbing pictures of American explorations and settlements are those of the amazing Capt. John Smith. For travels of all kinds, Samuel Purchas' collection, *Purchas His Pilgrims* (1625), was a not altogether unworthy successor to Hakluyt's.

3. The Essay and "Character."—Increasing observation and analysis of both one's self and society was manifested in the development of the essay and the "character." Early essayists, such as the pioneer Sir William Cornwallis (who inclined to the personal, discursive manner of Montaigne), Ben Jonson, in the pithy entries (mainly translated) in his commonplace book *Timber, or Discoveries*, and Owen Feltham, in his attractive *Resolves*, carried on in their several ways the ethical preoccupations and ideals of Renaissance humanism. Francis Bacon's essays (the ten aphoristic paragraphs of 1597 were much enlarged and multiplied in 1612 and 1625) were an integral part of their author's general scheme for the advancement of knowledge, and his mundane scale of values, which alienates modern readers, is that of the objective social psychologist examining men's actual motives and behaviour. Whether personal or impersonal, essayists mostly inclined to the serious and solid, though James Howell's *Epistolae Ho-Elianae* (1645–55) are lively reminders of the kinship between the letter and the familiar essay. To stretch the term essay out of all compass (though not without reason), Robert Burton may be included here. *The Anatomy of Melancholy* (1621; enlarged later) was a long and immensely learned treatise on the related maladies of body and mind, but no other psychological treatise has been, no modern one could be, a treasured bedside book. This cloistered scholar was an essayist in temper and the master of a frisky colloquial style. He was both an ironic observer of the human comedy and a wise, compassionate counsellor of tormented souls.

The "character," modeled on Theophrastus' brief, concrete sketches of ethical types of mankind, was especially cultivated in this period. The best-known books are Joseph Hall's *Characters of Virtues and Vices* (1608), the lighter, wittier and more popular collection (printed 1614–22) of Sir Thomas Overbury and others, and—the most inviting of all such works—the *Microcosmographie* (1628–33) of the young Oxford don John Earle, who combined wit with penetrating irony and warm humanity. The Theophrastian pattern was not much regarded in such things as Thomas Fuller's edifying and amusing conduct-book *The Holy State and the Profane State* (1642).

4. History and Biography.—In the writing of history, classical and modern foreign models emancipated English historians from the more or less naïve outlook and methods of the older chroniclers. The first really critical historian and antiquary was William Camden, whose *Britannia* was translated by Philemon Holland (1610). Sir Walter Raleigh's *History of the World* (1614) might seem to proclaim itself an exception, since, not to

BY COURTESY OF (1, 3, 4, 6–10, 12–15, 17–21, 23–25) THE NATIONAL PORTRAIT GALLERY, LONDON; (11) NATIONAL MARITIME MUSEUM, LONDON; PHOTOGRAPHS (5) WIDE WORLD, (16, 22) "PICTURE POST" LIBRARY

BRITISH MEN OF LETTERS, 14TH–18TH CENTURIES

1. Geoffrey Chaucer (1340?–1400). 2. Edmund Spenser (1552?–99). 3. Sir Philip Sidney (1554–86). 4. Francis Bacon (1561–1626). 5. William Shakespeare (1564–1616). 6. John Donne (1573–1631). 7. Ben Jonson (1573–1637). 8. John Milton (1608–74). 9. John Bunyan (1628–88). 10. John Dryden (1631–1700). 11. Samuel Pepys (1633–1703). 12. Jonathan Swift (1667–1745). 13. William Congreve (1670–1729). 14. Joseph Addison (1672–1719). 15. Alexander Pope (1688–1744). 16. Henry Fielding (1707–54). 17. Samuel Johnson (1709–84). 18. Robert Burns (1759–96). 19. William Wordsworth (1770–1850). 20. Sir Walter Scott (1771–1832). 21. Samuel Taylor Coleridge (1772–1834). 22. Jane Austen (1775–1817). 23. Charles Lamb (1775–1834). 24. Lord Byron (1788–1824). 25. Percy Bysshe Shelley (1792–1822)

ENGLISH LITERATURE

BY COURTESY OF (1–3, 5–11, 14) THE NATIONAL PORTRAIT GALLERY, LONDON, (4) GARRICK CLUB (LONDON) (13) OXFORD UNIVERSITY PRESS; PHOTOGRAPHS (12, 19) THE PRESS PORTRAIT BUREAU, (15) KARSH OF OTTAWA, CAMERA PRESS, LONDON, (16) 'PICTURE POST' LIBRARY, (17, 18) WIDE WORLD, (20) CAMERA PRESS, LONDON

BRITISH AUTHORS, 18TH–20TH CENTURIES

1. John Keats (1795–1821). 2. Thomas Carlyle (1795–1881). 3. Alfred Lord Tennyson (1809–92). 4. William Makepeace Thackeray (1811–63). 5. Charles Dickens (1812–70). 6. Robert Browning (1812–89). 7. Anthony Trollope (1815–82). 8. Charlotte Brontë (1816–55). 9. George Eliot (1819–80). 10. John Ruskin (1819–1900). 11. Matthew Arnold (1822–88). 12. Thomas Hardy (1840–1928). 13. Gerard Manley Hopkins (1844–89). 14. Robert Louis Stevenson (1850–94). 15. George Bernard Shaw (1856–1950). 16. William Butler Yeats (1865–1939). 17. H. G. Wells (1866–1946). 18. James Joyce (1882–1941). 19. Virginia Woolf (1882–1941). 20. T. S. Eliot (1888–1965)

mention some of his topics, he saw history as the working out of God's providential will in human affairs. Raleigh, a passionate poet in verse, showed his humane and poetic vision in his brooding on the mortality of men and empires, and his military sagacity in his discussion of logistics. The newer tradition of Machiavelli and Tacitus was behind Bacon's study of statecraft, *History of Henry VII* (1622). Very different, of course, was Fuller's *Church History of Britain* (1655) which, especially as it approaches its end in 1649, is a moving story of the church's trials.

Individual personalities were then attracting biographers, or telling their own story, as they had rarely done before. The leading biographical writers were the clerical Fuller and the businessman Izaak Walton, the friend and eulogist of clerics. (The best portraits of many writers and other men of this period were to come from the unclerical John Aubrey and the analytical earl of Clarendon.) Fuller was no Clarendon, but a born essayist, humourist and antiquary, and his love of anecdote and oddity showed itself in *The Holy State and the Profane State* and in the *Worthies of England* (printed 1662), the first attempt at a dictionary of national biography. Izaak Walton turned by degrees into a semiprofessional biographer. His lives of Donne, Sir Henry Wotton, Hooker, George Herbert and Robert Sanderson were printed during 1640–78; whatever their limitations, they remain both delightful and valuable, as a result of their author's sweetness of temper and his concern for truth and style. Fulke Greville's monument (printed in 1652) to the friend and idol of his youth, Sir Philip Sidney, is less of a biography than a political treatise.

It is axiomatic that no autobiography can be dull, and the period yielded a new wealth of memoirs and letters of both men and women—many not printed until later times—which are nearly all highly readable. Perhaps the most famous books are the poet-philosopher Lord Herbert of Cherbury's account of himself in a mainly swashbuckling role; *Eikon Basilike* (1649), the stained-glass image of the royal martyr—the joint work, it is now thought, of King Charles and John Gauden—which at once became as potent a weapon of propaganda as ever appeared in English; and that idyllic "picture of my own disposition," Walton's *Compleat Angler* (1653). But a multitude of more ordinary records reveal intimately the actual life and feeling of the times. Among collections of letters, those of Dorothy Osborne (1627–95) stand first, for the rare fragrance of their humour, pathos and charm.

5. The Influence of Religion.—Religion, religion of profound and passionate conviction, so permeated life and thought and literature that a few paragraphs cannot begin to suggest its range and power. It may be observed in its normal sphere, in the national conflict and in its reactions to the new science.

In this century the sermon was a major branch of literature. It was indeed the only literary form familiar to all classes of people, and educated hearers and readers were connoisseurs in the homiletic arts. Although preaching—and plain preaching—was much more stressed by Puritans than by Anglicans, the names that chiefly survive in literature are Anglican. Some readers have been led by T. S. Eliot to the clipped, "witty" and minute exegesis, and austere exaltations, of Lancelot Andrewes. Many have read in the large body of Donne's sermons, and know Walton's picture of his angelic fervour in the pulpit. In ringing the changes on his great theme, the sinfulness of man and the redeeming grace of Christ, the Jacobean Augustine, with all his intellectual resources, is haunted and inspired by his own guilt and fear and hope, by the terrors and joys of death and judgment and eternity. Jeremy Taylor is known mainly, from anthologies, for the richly pictorial images from nature which remind the modern reader of Spenser or Ruskin, and these, though characteristic, may obscure the earnest religious and moral counsel that make up the bulk of his sermons. No less characteristic of the Christian humanist are the frequent echoes of the ancient moralists and poets. Despite the note of ornate excess which Restoration preachers (and Matthew Arnold) censured, Taylor is in the classical tradition; he has nothing of Donne's agonized involvement and little of his rhetorical violence.

The best-sellers of the century were such manuals of practical piety as those of Arthur Dent (1601) and Lewis Bayly (1612?)—which were the dowry of Bunyan's first wife—and there were large publics for books of meditation and moral guidance, both Catholic and Protestant, original and translated. Under this general heading may be put Andrewes' *Private Devotions*, Donne's very individual *Devotions* (1624), Taylor's *The Rule and Exercises of Holy Living* and *The Rule and Exercises of Holy Dying* (1650–51), Richard Baxter's popular *Saints' Everlasting Rest* (1650), the cool but also popular *Whole Duty of Man* (1658) by the shadowy Richard Allestree, and Thomas Traherne's radiant visions of felicity, *Centuries of Meditations* (first printed in 1908).

In the national arena, religious and ecclesiastical issues were interwoven with politics and economics, and a host of writers—and preachers—engaged in the controversies that preceded and accompanied war and revolution. In the Puritan appeal to the Bible and apostolic simplicity, the Anglican appeal to tradition, order and expediency, King James's conception of divine right, Roger Williams' ideal of toleration, John Lilburne's advanced Leveller ideas, Gerrard Winstanley's Christian communism, Milton's dynamic gospel of Christian liberty and his Platonic-Puritan republicanism, the anticlerical, antidemocratic Thomas Hobbes's deification of civil power—in these and other arguments religion, or the relation of church and state, was a more or less central force. Hobbes, who had lived through the Civil War (in Paris), sacrificed traditional rights and liberties to the supreme necessity of order; he might be said to have turned Calvin's absolute God into an absolute sovereign (as John Locke, the philosopher of the 1688 Revolution, was to enlarge and solidify Hooker's theory of government by consent).

6. Religion, Science and Philosophy.—The conflict between religion and science first took modern shape in the 17th century. The main tradition of western belief and thought, which the century inherited, was the Christian humanism of the middle ages and the Renaissance. This fusion of the natural light of ancient reason with the supernatural light of revelation had become an all-embracing philosophy of order which combined metaphysical, religious, ethical, political, social, economic and "scientific" knowledge and thought. The great Elizabethan exposition of divine and human reason and law was the first book of Richard Hooker's *Laws of Ecclesiastical Polity* (1593). This classical-Christian view of God, nature and man encountered three chief agents of disintegration: Calvinism, with its emphasis on the arbitrary, irrational omnipotence of God and the irrational, impotent depravity of man; the Baconian doctrine of scientific progress, which, however necessary, tended—as the modern world knows better than Bacon could—to enthrone "things" and to bestow power upon man's appetites and passions; and traditional skepticism and naturalism, now fortified by scientific discovery and thought, which issued in the progressively mechanistic philosophies of Descartes and Hobbes.

Francis Bacon, an eminent lawyer, judge and statesman, was, if not fully abreast of scientific activity, the prophet and philosopher of a new and dazzling conception of progress. In *The Advancement of Learning* (1605), the broadest statement in English of his grand scheme, and a monument of his massive prose, he arraigned traditional and obstructive modes of education and inquiry, outlined the current state of knowledge in all fields, and proposed new lines of exploration. But his prime interest was science, the mastery of nature for the use and benefit of man. His inspiring vision and his inductive method were set forth in the *Novum Organum* (1620); after long wanderings in the desert of Aristotelian logic, man might enter the promised land through experimental science. Bacon's aims and method (a method admittedly incomplete) were praised by many scientists of his century; from many modern writers his ideas have received more criticism and less credit than they deserve. His dream of co-operative research, illustrated in *The New Atlantis* (1627), was fulfilled in the Royal society.

The disruption of the traditional religious synthesis proceeded rapidly in some advanced intellects. The dualism of Descartes—God and mind on one side, matter on the other—opened the way to either idealism or materialism (or pantheism). For Hobbes,

already a thoroughgoing materialist, mind was, like the cosmos and everything else, a system of bodies moving in space and time; and his bare, trenchant style (in *Leviathan*, 1651, and lesser works) was a fitting instrument for his thought. The deterministic philosopher was attacked on all sides. In the face of both irrational Calvinism and irreligious materialism the Cambridge Platonists (*q.v.*) reasserted traditional Christian humanism in partly new terms. Benjamin Whichcote, John Smith and Nathanael Culverwel may be said to belong to literature. Henry More and Ralph Cudworth hardly do (unless More's darkly Platonic verse be excepted); but in their effort to meet Descartes and Hobbes, to accommodate the new science to liberal Christianity, to vindicate metaphysical and ethical idealism, they were philosophically important. More in particular influenced Newton and Yeats.

While the Cambridge Platonists have a place in histories of philosophy but are not much read, Sir Thomas Browne, who has no such place, is cherished as one of the great artists. *Religio Medici* (written in 1635), though pre-Cartesian and pre-Hobbesian, was the work of a young scientist who was also a Christian Platonist conscious of a rationalistic climate. A devout Anglican in the essentials of faith, Browne could let his "sceptical" intellect and Platonic idealism play with reverent wonder about the paradoxical mysteries of God and nature ("the art of God") and man. In this early but central book his Senecan style gives mainly incidental foretastes of his later splendours. *Pseudodoxia Epidemica* (1646) was a soberly scientific (and often diverting) attempt to fulfil Bacon's desire for a sifting of popular errors, especially in natural history. Browne's chief later works were the sombre *Letter to a Friend* (written perhaps in 1656; published in 1690), *The Garden of Cyrus* and *Hydriotaphia, Urne-Buriall* (1658). The last two, antiquarian in their initial impulse, open out into complementary meditations, richly imaginative and religious, on life and death, time and eternity, and in them, especially in the fifth chapter of *Urne-Buriall* and the conclusion of the *Garden*, Browne is the supreme metaphysical poet of prose. His blend of monosyllabic Saxon homespun and polysyllabic Latinate purple, of Christian and pagan ideas of the greatness and littleness of man —all this is pure yet substantial incantation.

7. Verse Translations.—The 17th century opened with some famous translations including Chapman's *Homer* (1598–1615), Edward Fairfax's *Godfrey of Bulloigne* (1600), Joshua Sylvester's version of Du Bartas (1605) and George Sandys' translation of Ovid's *Metamorphoses* (1626)—this last partly done in Virginia, the first American literary work. If Chapman's conceits ruffled the clear stream of Homer, and if his ethical creed was a bit obtrusive, his translations, the work of fervent devotion, were truly heroic poems. Fairfax's *Tasso*, in ottava rima and Spenserian style, remains the standard English version. Sylvester's *Du Bartas* reigned as the Protestant epic of creation until displaced by *Paradise Lost*. Some of these translations, notably Sandys', contributed to the development of the closed, balanced Augustan couplet, which in this genre and age culminated in Dryden's *Virgil* (1697) and *Fables, Ancient and Modern* (1700). Sylvester and Sandys illustrate also the beginnings of Augustan "poetic diction." And Sandys' allegorical commentary on Ovid (1632), one of the last flowerings of a potent tradition, while soon blighted by Augustan rationalism, was to revive in the mythological imagination of Keats.

8. Spenserian Influence.—The poetry of the first half of the 17th century has been commonly divided into three main currents represented by the names of Edmund Spenser, John Donne and Ben Jonson; the division has not much meaning but remains convenient. In an age whose young poets inclined to the personal and intellectual concentration and complexity of Donne, the followers of Spenser, notably Giles and Phineas Fletcher and Drayton, carried on old fashions in themes and manner and are therefore treated above, among the Elizabethans with whom their work belongs (*see* above: *The Successors of Spenser*).

The Scottish William Drummond, like the Spenserians, found the new intellectualism alien to his conservative, Italianate and especially Petrarchan taste. Among his somewhat florid and lowpulsed pieces on love, death and religion, a few are sustained poetry; some of the best share the Neoplatonism of his prose meditation on death, *A Cypresse Grove* (1623). Samuel Daniel, George Chapman and Fulke Greville were concerned with the cultural, ethical and religious predicament of modern man. Daniel in *Musophilus* (1599) and some weighty epistles, and Chapman in *The Tears of Peace* and elsewhere, upheld humane culture and Stoic integrity and wisdom against worldly, sensual or merely intellectual distractions. But while Daniel writes with a smooth half-prosaic clarity, the texture of Chapman's thought, feeling and expression is "metaphysical" in its rugged intellectual and figurative complexity and density. Greville's treatises in verse, and his uniquely philosophical sonnet-sequence, have a sinewy toughness that is nearer Chapman than Daniel; but the poet-statesman, with his dark religious view of man's incurable self-will and corruption, could not share the partial optimism of their humanistic creed and still less the scientific optimism of his fellow statesman Bacon.

9. Donne and the Metaphysical Poets.—The 20th century's revival of John Donne and other metaphysical poets, and their effects upon modern poetry and critical canons, have altered the present perspective. In defining the term Samuel Johnson went far beyond Dryden, and Johnson's definition has been refined. Some elements are: "wit," the yoking of apparently unrelated ideas and things; organic rather than merely illustrative images; concrete particulars rather than abstract generalities; the concentrated fusion of thought and feeling, or the intellectualization of feeling; inward tension between opposed ideas and emotions, and hence complexity, obliquity, irony, paradox; predominantly realistic and colloquial language and rhythms; the fully expressive rendering of individual experience, of, in Donne's phrase, the "naked thinking heart." Some of these headings suggest the parallel between the poetry of wit and the anti-Ciceronian movement in prose. Yet if Donne and his fellows are thought of as rebels against Elizabethan or Spenserian orthodoxy, as seeking revolutionary novelty, qualifications are needed. In theory and even in practice there were not two completely opposed schools. Many poets, from Shakespeare down, were more or less metaphysical; and Donne might be said to have carried into the poetry of love and religion the technique that Elizabethan decorum reserved for satire. A more general and important fact is that wit seems more discordant now than it did in an age when all ideas, and all creatures and things in the divine hierarchy of the chain of being, were linked together by correspondence and analogy; the essence of wit is the allegorical and symbolic mode of apprehension, which was a common heritage.

What has just been said holds for Donne and, with varying reservations, for other metaphysical poets. The circumstances of Donne's life did not predispose him to share the conventionalities of either Petrarchan or religious devotion. In his love poetry he can be cynical or sensual or outsoar Petrarchan hyperbole, but he is concerned with the relations of body and soul; in some of his best poems he finds in love the only reality. His explorations, negations and affirmations are presented with explosive force and originality; he is witty and serious, dramatic and personal, dialectical and passionate, abstrusely learned and concretely realistic, harshly staccato and richly resonant. The same man and artist appears in the religious poems; the bulk of these, including the sonnets, were written in the years 1607–12 (he was not ordained until 1615). Although Donne had always been a Christian in belief, he enacted in his own quivering soul the whole drama of human guilt and divine love. The torments of the sonnets give way in the long *Anniversaries* (1611–12) to less personal meditations on the sinful decaying world and the perfections of heaven; and in the several later "Hymns" a note of something like personal assurance is attained.

Donne's poems were for the most part not published until after his death, but they circulated in manuscript and exerted a strong influence on both secular and religious poetry; and of course the climate that affected Donne affected other poets. Of the chief religious writers, George Herbert, Richard Crashaw, Henry Vaughan and Thomas Traherne, only the first fits into a strict definition of metaphysical poetry. Herbert's richly symbolic wit op-

erates in two spheres (though his conception of Love makes them really one): a poet-priest's celebrations of Christ and his church, and the more personal poems in which he strives to subdue his rebellious will or to quicken what seems to be his flickering zeal. Such tensions make the stronger impact through his quiet, neat and subtle craftsmanship, his organic structure and unified imagery, his habit of understatement, his colloquial language and his expressive delicacy (and rich variety) of rhythm. Crashaw, the Catholic convert, is in English the very definition of baroque poetry, and the Italianate and Spanish spirit and style of his flamboyant devotional ecstasies have only a tincture of the metaphysical. The ecstasies of Vaughan are those of a Christian Neoplatonist who, but for the intimations of God in nature, feels himself an exile on earth and longs for the pure white radiance of eternity. Vaughan has less paradoxical wit than his themes might seem to invite; he is a poet—an uneven poet—of contemplative vision, and he can, in colloquial speech, touch sublimity.

It is impossible even to glance at the many poets of the secular line, such as Lord Herbert, Henry King (the clergyman author of the splendid *Exequy* on his young wife), the fantastic and satirical John Cleveland (who was outdone in fantasticality by the religious Edward Benlowes), and the one man of great fame, Abraham Cowley, author of clever amatory verse, the unfinished epic *Davideis*, and the "Pindaric" odes on public themes; but a word must be said about the finest of them, Andrew Marvell. To pass by the Restoration satires, Marvell lives in the poems written, apparently, during 1650–53, when both metaphysical and Cavalier strains might have seemed to be dying of excess or exhaustion. "To His Coy Mistress," "The Definition of Love," "Bermudas," "The Garden," "The Picture of Little T.C.," "A Dialogue Between the Soul and Body," "The Coronet," "Horatian Ode Upon Cromwell's Return From Ireland" are examples of his range and quality. Marvell combines the complex metaphysical intelligence and sensibility with the easy grace of a Cavalier lyrist and a structural and stylistic symmetry, economy and clarity that may be called classical. The combination is unique.

10. Ben Jonson and His Influence.—In the golden age of English music, before and after 1600, numberless musicians and lyrists created a wealth of song. Thomas Campion is the most noted example of the composer-poet. Along with the elaborate lyricism of art went the simpler ditties of popular tradition; the songs of Jonson and Shakespeare illustrate both. Ben Jonson was of course more than a lyrist. While he and Donne have stood as progenitors respectively of the cavalier and metaphysical schools, they had affinities in their matter and "unpoetical" manner, and their followers were affected by both; in the first half of the century the religious and philosophical climate made the metaphysical sensibility almost inevitable. The classicism that distinguishes Jonson extends from the ethical and especially Stoic ideals of Renaissance humanism (embodied in reflective poems) to the artistic ideals that moulded his songs and epigrams. Whether as lover of the living or elegist of the dead, Jonson cultivated disciplined impersonality, symmetry and clarity; his normal style is rational and urbane, without shocks of surprise or extravagance. His poems are "made objects" rather than jets of self-expression.

Among the many writers, "sons of Ben" and others, who felt Jonson's influence, perhaps the purest disciple and the most original poet was Robert Herrick. He stands, as it were, between Jonson and Marvell. His truly metaphysical wit, his instinct for the concrete particular, for ambiguity, paradox and surprise, is half-concealed by his smooth classical elegance which includes a feeling for word and rhythm more exquisite than his master's. Herrick can live in the happy pagan world of his fancy or in the Christian world of his belief, or he can be poignantly conscious of both. The courtier poets, Thomas Carew, Sir John Suckling and Richard Lovelace, are also heirs of Jonson who have more or less wit. Carew, the most studied artist, praised Donne's masculine originality but he himself shows little of Donne's mingled passion and intellectualism; and while Donne's wit may shoot in all directions, Carew is given to a classical pattern, to one controlling idea and image. If Carew elevates the sensual game of love into a courtly ritual, "natural, easy Suckling" speaks with the casual, humorous nonchalance of a man about town. Lovelace lives in the few songs that represent the ideal Cavalier, but even in these simplicity is touched with complexity; and he can strike quite different notes.

11. John Milton.—A poet of the stature of John Milton cannot be surveyed briefly. His career, while one of continuous development, falls into three periods: that of the early poems in Latin and English; the years 1641–60, which were given almost wholly to public affairs (and occasional sonnets); and the last phase, which saw the publication of his three major poems. The volume of 1645 included the ode of baroque beauty "On the Morning of Christ's Nativity," the charming "L'Allegro" and "Il Penseroso," which suggest and surpass the Jonsonian manner; the masque "Comus," diversified in its stylistic texture, unified in its glowing Christian Platonism; and "Lycidas," one of the most complex and reverberating poems in the language, a first effort to justify God's ways to men, in which profound tensions are half-veiled by the pastoral mask. The metaphysical current left the young classicist almost untouched; if he began in the Spenserian line, Milton was from the start a master of original power.

Like other Renaissance poets, Milton dreamed of writing the great modern heroic poem, but his sense of religious and civic responsibility impelled him to give 20 years, and his eyesight, to the Puritan struggle for liberty. He wrote five tracts against episcopacy; four on divorce; *Of Education; Areopagitica* (1644), a plea for freedom of the press; and, on the eve of the Restoration, a defiant and despairing plea for a republic. Milton's pamphlets help to explain the motives of the Puritan revolution. They record his increasingly radical thought, his early hopes of a grand reformation and his successive disillusionments; and they are the work of an impassioned and individual master of loosely Ciceronian prose who runs the gamut from satirical snorts to sublimity.

When he came at last to compose *Paradise Lost* (published 1667), Milton was no longer the militant revolutionary; his three major works deal with the temptations, defeats and victories of individual man. His early Baconian ardour has likewise given way to the need for reasserting in the face of science the Christian-humanist view of God, man and nature. Eve, in her fatal desire for the godlike knowledge Satan promises, Adam, in his scientific curiosity and his following of Eve, re-enact the sin of Satan, irreligious pride and rebellion. The poem links itself with modern pictures of an inner and outer wasteland created by egoistic naturalism. Satan, one of the great figures of all literature, is from the first a personality of both grandeur and corruption; in the end only the corruption survives. In presenting his heroic "villain" Milton uses the method of Elizabethan drama (along with interpretative comment); like Shakespeare, he can—or in his own age could—rely upon a right-minded audience's reaction to evil. Further, a universally known story permitted an enveloping dramatic irony: Satan (except in candid soliloquies) and his followers cannot comprehend the true nature of the conflict between evil and good; and Adam and Eve, whose idyllic life the reader witnesses only as he accompanies Satan into the garden, do not know their destiny.

The supreme classical artist in English (and modern European) poetry wrote as the conscious heir of the ancients. He recreates the large conventions and endless details of the classical epics; his images blend the general with the particular, the vague with the concrete; and his rich bold style is, like that of Homer and Virgil, elevated above common speech, though ornate stylization includes both simplicity and complex density of suggestion and overtone. The use of blank verse for a long poem was a radical novelty, and Milton's handling of it added new worlds to English prosody. Some features of his rhythmic patterns had appeared in the heroic sonnets. It gives only a bald hint of Milton's orchestration and subtle expressiveness to say that there is continual variation within the ten-syllable line and within the verse paragraphs, notably in the secondary system of rhythmic units that flow from a pause in one line to a pause in another.

Paradise Regained and *Samson Agonistes* (both 1671) show still further artistic developments. The former is in a style of almost biblical simplicity; the latter, the only English drama on the Greek model that can stand with those of the ancients, is in part in ruggedly irregular (but not free) verse that comes close to

speech. The hero of *Paradise Regained*, in a conflict of mounting tension, displays the humble obedience and integrity that Adam and Eve had lacked. In the drama a sinful human hero—isolated, like the half-human Christ, even from the sympathetic—achieves inward regeneration and inward as well as outward victory.

These three poems of faith and fortitude were not in the spirit of most Restoration literature. Milton's answer to the question "What do I know?" was doubtless, like that of the Cambridge Platonists, already an anachronism—though it may seem less so to many readers in the middle of the 20th century. At any rate the literature of the earlier 17th century is a large and living part of our best knowledge. (D. Bh.)

IV. THE RESTORATION PERIOD

1. Prose.—The year 1660 marks a change in English prose and poetry. This change was not so great as has sometimes been stated. For one thing the lives of writers overlap eras, and for another a plain and straightforward manner of expression had always been available to those who wanted to use it. Sir Thomas Browne and Jeremy Taylor (*see* above) were exceptions; and Browne's latinized and rhetorical style was essentially artificial, as can easily be seen if his ornate prose is compared with his private letters. It is true that men became increasingly interested in things rather than in words as science came to play a greater part in human affairs, and for this reason perhaps the custom of writing emotional prose largely disappeared and a more familiar diction took its place. The Royal society (*q.v.*) played some part: its origins were much earlier than the year 1662 when it received its first charter. The society was interested in prose style and soon after its foundation appointed a committee "for improving the English language." Thomas Sprat in his *History of the Royal Society of London* (1667) (really a work of propaganda) laid down a series of rules for a plain style and, in a famous passage, advocated "a close, naked natural way of speaking, positive expressions, clear senses, a native easiness." John Wilkins had done much the same in his *Ecclesiastes* (1646), often reprinted, and in his *Essay Towards a Real Character and a Philosophical Language* (1668) he made a gallant, if futile, attempt to substitute ciphers for words so that their old associations should be lost. Probably the direct influence of the society has been overemphasized. The only popular book which bore the arms of the society on its title page was John Evelyn's *Sylva* (1664), and Evelyn's style was not particularly modern. The most frequently quoted instance of an alteration in style is Joseph Glanvill's rewriting of his own *The Vanity of Dogmatizing* (1661) as *Scepsis Scientifica* (1665), but the two books are not really in fact the same. Many of the really scientific works such as those by the great naturalist John Ray were in Latin. Such works of natural history as Robert Plot's *The Natural History of Oxfordshire* (1677) were written in a simple style. It is significant that Plot still thought it necessary to write: "I intend to deliver as succinctly as may be, in a plain, easie and unartificial Stile, studiously avoiding all ornaments of language, it being my purpose to treat of things, and therefore would have the Reader expect nothing less than words." No generalization about the effects on scientific writers is possible, however, for, as a single example, Robert Boyle's *The Sceptical Chemist* (1661) is almost unreadable because of his diffuseness. There were many influences at work. Hobbes in his *Leviathan* (1651) had forbidden a counselor to use "obscure, confused and ambiguous Expressions, also all metaphorical Speeches, tending to the stirring-up of Passion." But probably the rise of journalism was as influential as any other cause (*see* Newspaper: *United Kingdom*). During the Civil War and the Commonwealth pamphlets had been abundant. On the whole they were plain and simple enough, as a political pamphlet must be if it is to be effective. Of journalists Sir Roger L'Estrange was as direct a writer (if a noisy one) as it is possible for anyone to be. As licenser and surveyor of the press (1653–88) he had considerable influence in another direction. He is chiefly remembered, however, by his *Fables of Aesop* (1692).

The sermon, too, became simpler after 1660 than it had been in the days of the Caroline divines. Robert South (*q.v.*) in particular preached in a manner which was much less ornate, if far less attractive, than Taylor's. Essentially a masculine preacher, South attacked his great predecessor: "Nothing here of the fringes of the north-star; nothing of nature's becoming unnatural; nothing of the down of angels' wings or the beautiful locks of cherubims: no starched similitudes introduced with a Thus have I seen a cloud rolling in its airy mansion and the like." This well-known passage is coarse; but nothing shows better the dislike entertained by the newer preachers for the older. Isaac Barrow, the mathematician, cannot readily be placed. His eloquence is somewhat between the new and the old styles, though, as he was apt to give practical advice, he must on the whole be reckoned among the more modern preachers. Thomas Burnet's *Sacred Theory of the Earth* (1684) can be mentioned as an example of a book containing, here and there, passages of mild splendour, which had become unusual by the date. But it was probably the influence of a few individual writers, rather than vague tendencies of the time, that made a simpler and more workaday prose the usual method of expression. A lucid and logical, if rather empty, preacher was John Tillotson (*q.v.*). He almost belongs to the 18th century in manner and his influence was for a time very great. Dryden made the strange remark reported by Congreve in 1717 "that if he had any Talent for *English* Prose, it was owing to his having often read the writings of the great Archbishop *Tillotson*." Because of the dates of their maturity as authors this can have little meaning. But it at least shows that Dryden considered Tillotson a writer akin to himself.

Apart from Tillotson, who in any case was not in a prominent position until Dryden was well established, John Dryden was the most influential prose writer in the years between 1660 and the end of the century. His first fairly long piece of prose was his "Letter to Sir Robert Howard" prefixed to his poem *Annus Mirabilis* (printed 1667). This was followed in 1668 by his *Essay of Dramatic Poesy* and by a number of dedications and prefaces. In 1700 his masterpiece, the preface to his *Fables, Ancient and Modern* was published.

Dryden mastered a style both lucid and conversational which has few equals in the English language. Although it could not be imitated few authors would have wished to go back to many of the pre-Restoration writers. With Dryden modern prose became established. Under all his seeming ease he took great care, and actually revised the syntax of his *Essay of Dramatic Poesy* in the second edition (1684). There is no doubt that the subject matter with which he and his contemporaries were dealing led to a change in style. No longer were problems such as the transience of life or death itself, which lent themselves to rhetorical treatment, discussed, but rather matters of immediate concern, such as criticism or contemporary politics. Many other authors whose writings are pleasant enough were publishing essays and memoirs, notably Sir William Temple (*q.v.*). There were plenty of pedantic writers, especially among the clergy, but innumerable books were written showing disapproval of pedantry. The most amusing is John Eachard's *The Grounds and Occasions of the Contempt of the Clergy* (1670). It was so popular that it reached a tenth edition by 1696. In 1672 Andrew Marvell introduced a vein in prose called "drolling" or, in modern language, "banter." In *The Rehearsal Transpros'd* (1672) and in *The Rehearsal Transpros'd: the Second Part* (1673) he ridiculed Samuel Parker, who later became bishop of Oxford, for his bigotry. It was an age of controversy, and in 1675 another dispute arose over toleration. Herbert Croft, bishop of Hereford, argued that Protestants differed about nothing essential. This view was furiously attacked by Francis Turner, master of St. John's college, Cambridge. Marvell took up the cudgels on Croft's side in *Mr. Smirke, or the Divine in Mode* (1676), and again produced an amusing satire. Marvell, though in a different way from Dryden, tended to make prose lighter and more familiar.

In 1678–79 there was a great outpouring of pamphlets and what were called narratives, over the "Popish Plot," supposedly a large-scale conspiracy to recover England for the Roman Catholic Church, but in fact a fabrication of the perjurer Titus Oates (*q.v.*). One of these was, or professed to be, by Oates himself. Antagonism was bitter and much of the writing of the time was polemical

and so has lost its interest. One writer stands out far above all others except Dryden. This was George Savile, marquess of Halifax (1633–95). He did not begin writing until 1681 and it was some years before he produced his *Letter to a Dissenter* (1687), *Character of a Trimmer* (1688), inspired by L'Estrange's *Humour of a Trimmer* (1684), and other pieces. The first was written to try to heal differences in the state: the second to defend his own position. Probably neither achieved their object to any degree, but they are packed with wisdom. *The Lady's New Year's Gift, or Advice to a Daughter* (1688), if it displays a sombre view of the possibility of happiness, is as valuable today as when it was written. His masterpieces were his *Maxims of State* (1693) which contain pregnant reflections and, if not so witty, are more profound than La Rochefoucauld's; and his posthumously published *Character of King Charles II* (1750).

John Bunyan belongs to a different world from that of any of the writers who have been mentioned: The world of the poor, the artisans, wage earners, yeomen and shopkeepers. It had a culture of its own founded on Bible reading and the tradition of the Puritan sermon, as well as on ballads, songs, newsbooks and romances, which Bunyan tells us he read in his unregenerate days. A tinker by trade, he was one of many "mechanick preachers" of the period who printed sermons and religious discourses, but he was the only one whose work has high literary value.

Bunyan's first notable book, the great spiritual autobiography, *Grace Abounding*, appeared in 1666 and was subsequently enlarged. *The Pilgrim's Progress*, his masterpiece, the first part of which appeared in 1678, grows directly out of *Grace Abounding*. It is a translation into allegorical terms of the Puritan soul's journey to salvation, but it is far more than this; it is a universal symbol of the human condition and the aspiration of man to divine grace, as well as a wonderfully vivid picture of the life of the English high road and the small country towns of the reign of Charles II. In the second part (1684), the realistic and humorous elements are more prominent, while the famous climax where Mr. Valiant-for-Truth and his fellow-pilgrims cross the river is one of the most resonant and moving prose passages in the English language.

Among Bunyan's other works the most memorable are *The Life and Death of Mr. Badman* (1680), a stark and powerful piece of realism, and *The Holy War* (1682), an allegory on a grand scale, less human than *The Pilgrim's Progress*, but with qualities of epic spaciousness and grandeur. Bunyan is a writer who stands between two worlds. He can be regarded at once as the last medieval allegorist and as the first great master of realistic and psychological fiction.

Religious works are apt to fall out of date, but the writings of the Cambridge Platonists (*q.v.*) are still of importance to theologians and philosophers (*see* above, *Religion, Science and Philosophy*). They must also be reckoned among the influences which led to the simplification of prose style. The most notable product of the group was probably *The Select Discourses* of John Smith, published in 1660, the expression of a singularly beautiful and enlightened mind with a surprisingly modern outlook on religious questions. The posthumously published *Select Sermons* (1698) and *Moral and Religious Aphorisms* (1703) of Smith's master, Benjamin Whichcote, are written in a terse, pregnant style and are full of lofty wisdom. Peter Sterry, for a time one of Cromwell's chaplains, is the most imaginative and poetical writer of the group. His opinions as expressed in his *Discourse of the Freedom of the Will* (1675) look back to the neoplatonism of the Italian renaissance, and forward to Coleridge's doctrine of the creative imagination as expressed in the *Biographia Literaria* (1817). Among other religious writers, Richard Baxter, the great Presbyterian leader, is remembered for his beautiful *Breviate of the Life of Mrs. Margaret Baxter* (1681) and his ill-arranged but extremely interesting posthumously published autobiography, *Reliquiae Baxterianae* (1696).

One of the most important developments in prose literature in the period is to be found in the great interest that was taken in biography; it is significant that the word itself seems to have been first used in English about 1660. With the possible exception of William Roper's *Life of Sir Thomas More* and George Cavendish's *Life and Death of Thomas Wolsey*, both written in the 16th century, earlier biographies had tended to be either eulogies descending from the classical funeral oration or hagiography deriving from the lives of saints. Sprat's *Life of Cowley* (1668) is a later example of the old kind of biography that contained little beyond praise of the deceased. The famous series of *Lives* by Izaak Walton, though begun before the Civil War with the biography of John Donne (1640), were not published in a collected edition till 1670. Though still containing traces of hagiography, they are a landmark in the development of English biography. They are not only beautiful expressions of Walton's charming personality but are based on careful observation and research and contain invaluable records of personal traits and conversations and examples of correspondence. Almost equally important are Roger North's *Lives of the Norths*. The life of John North is of peculiar interest as it successfully describes an extremely neurotic man. As the *Lives* were not published until 1742–44 they can have had no direct influence on the biography of the time, but they show the direction biography was taking. No real distinction can be drawn between biography and history, for both the earl of Clarendon and Bishop Gilbert Burnet were skilful in different ways in drawing characters. Their histories were not published until the 18th century, but they belong to the Restoration period. To confuse matters more, the earl of Clarendon's own *Life* of himself, which forms an essential part of his *History of the Rebellion*, was first published in 1759.

Athenae Oxonienses (1691–92), by Anthony à Wood (*q.v.*), was the first serious attempt at an English biographical dictionary. It was partly the work of John Aubrey (*q.v.*), who was able to supply the Oxford antiquary with much information. Aubrey's own *Brief Lives*, as they are now called, are vivid and entertaining jottings about people. He traveled much and visited a number of the people whom he describes—including, frequently, their appearance—with a minute eye for detail. It is hardly necessary to mention the diaries of John Evelyn and Samuel Pepys except to remind the reader that whereas Evelyn's *Diary* extends from 1631 to 1706, Pepys's ends in 1669. As Evelyn often wrote up his diary some days after the events recorded his dates cannot always be trusted. Neither was published until the early 19th century, Evelyn's in 1818 and Pepys's in 1825. The *Memoirs* (1734) of Sir John Reresby give an admirable account of his travels in England on various missions. Among short biographies John Fell's *Life of Dr. Henry Hammond* (1661), the duchess of Newcastle's *Life* of her husband (1667) and Gilbert Burnet's *Life and Death of the Earl of Rochester* (1680) are outstanding. Lastly Anthony à Wood's autobiography, *Life and Times of Anthony à Wood*, published in 1891–1900 from his manuscripts, is essential to the student as well as being eminently readable.

The glory of the Elizabethan and early Jacobean translations had passed away. In their place were many rather commonplace renderings of classical and French authors. The problem of translation troubled many. No doubt a more exact scholarship than that of the Elizabethans raised in an acute form the question of what a translation should be. Was it to be an exact reproduction of the original or was it to be a paraphrase? The problem was not and never can be solved. Dryden at least produced admirable versions of portions of Lucretius, while his Virgil (1697), published in his old age, if not Virgil, has splendours of its own. Scholars such as Thomas Creech undertook Ovid and Lucretius and even translated Manilius. Hobbes made a complete translation of the *Iliad* and *Odyssey*. Dryden in his preface to his *Fables* calls them "bald" and adds, truthfully enough, that Hobbes began studying poetry as he did mathematics, when it was too late. The taste for translations must have been keen as they are so often found in the miscellanies of verse of the time.

Critical essays in verse and prose ranged over literature in general. Thomas Rymer's *A Short View of Tragedy* (1693) had been preceded by his *Tragedies of the Last Age* (1678). These books are not without interest. But the most outstanding work was Robert Wolseley's preface to the earl of Rochester's version of Beaumont and Fletcher's tragedy *Valentinian* (1685). Its importance

lies in an attempt to solve the problem of the relation of art to morals and, with the exception of Dryden's prefaces, etc., it is the most significant piece of Restoration criticism. Better remembered than Rochester's *An Allusion To Horace* or the earl of Roscommon's *Essay on Translated Verse* (1684) and several other criticisms written in verse, is *A Short View of the Immorality and Profaneness of the English Stage* (1698) by Jeremy Collier. Although written without much sense of humour it is vigorous, and effected its purpose to some extent in making plays more decorous than they had been. Collier's violence gave him an advantage over William Congreve, Sir John Vanbrugh and others who attempted replies.

Criticism is nearly always somewhat ephemeral, but genuine scholarship is not. The *Monasticon Anglicanum* of William Dugdale was published in 1655–73; the real originator of the work was Roger Dodsworth. With this publication the basis of medieval scholarship was laid. Anglo-Saxon studies were much advanced, but the subject is so specialized that the student must turn to David Douglas's *English Scholars, 1660–1703*, second edition (1951), for detailed knowledge.

The age produced a number of somewhat eccentric men who wrote interesting books that fit into no classification. Nathaniel Wanley's *The Wonders of the Little World* (1678) is full of curious information. Gerard Langbaine's *An Account of the English Dramatic Poets* (1691) is useful. It is packed with charges, especially against Dryden, of plagiarism. A revised edition pruned of irrelevancies was published posthumously as *The Lives and Characters of the English Dramatic Poets* (1699). Dictionaries of various sorts were available. Probably the most popular was Edward Phillips' *The New World of English Words: or a General Dictionary*. First published in 1658, it reached a sixth edition in 1706.

2. Poetry.—The impulse to the writing of ceremonial verse was given by the Restoration. Robert Wild's *Iter Boreale* (1660) greeted the king's return. But its popularity soon faded and so far as he is remembered it is by "Alas poor Scholar! whither wilt thou go?" Dryden was soon in the field with *Astraea Redux*, published in the same year. In fact Charles was greeted with more verse than he can have found congenial. In 1661 Dryden wrote a panegyric on the coronation and in the next year addressed a poem to Clarendon. These poems probably helped to establish the use of the heroic couplet, as they were, at least, technical improvements on the poems in couplets which had preceded them. But all these thin folios could not compare in popularity with Samuel Butler's mock-heroic satire *Hudibras*. Much of this remarkable work had been written a good while before the Restoration. The first two parts were published in 1663 and 1664. A revised edition appeared in 1674 and a third part was added in 1678. The poem delighted all royalists from the king downward. It is less an attack on the Puritans than on the folly and pedantry of mankind in general. It is best understood if Butler's miscellaneous pieces are read with it. Butler's *Remains* were not published until 1759. They include his *Characters* which return in manner to the generalized form of Sir Thomas Overbury (*see* above). *Hudibras* was constantly imitated, but its octosyllabic couplets with their extraordinarily witty double rhymes were, as John Dennis wrote in 1717, "admirable in a manner in which no one else has been tolerable." The poem is not a consecutive narrative and was never finished. But it is a great work and kept its popularity until the early 19th century. No English writer has been able to say so much in such few words. Its satire is without comparison and is continuous although the episodes are often unconnected. Lines such as these constantly occur:

> The oyster women picked their fish up
> And trudged away to cry No Bishop.

Occasionally some verse in Hudibrastics, such as Tom D'Urfey's *Collin's Walk through London and Westminster* (1690), has a mild interest of its own. Butler was entirely original and stands apart from the writers of travesties, a kind of writing which was one of the worst literary vices of the age. Travesty is exemplified in John Phillips' *Don Quixote* (1687), which is neither a translation nor a paraphrase, but a rather vulgar burlesque of a great book. Virgil was a subject for many travesties and even Charles Cotton, who could do much better things, produced a mock Virgil following a custom which had begun soon after 1660. This is quite different from the far happier custom of "imitating" classical works by introducing modern names and adapting them to the contemporary scene. John Wilmot, earl of Rochester, seems to have originated this kind of writing in his "Allusion to Horace the 10th Satyr of the 1st Book" (written *c.* 1675; published 1680) and he was followed by John Oldham and others, forming a precedent for the classic works in this manner by Pope and Johnson in the next century.

Dryden dominated the age. Apart from his prose writings, already mentioned, he wrote nearly 40 plays as well as a great deal of verse. His most famous poems are nearly all satires and were produced within a short period. *MacFlecknoe*, his effective attack on Thomas Shadwell, was published surreptitiously in 1682, though written as early as 1678. His best-known poem, *Absalom and Achitophel*, was published in 1681. It has the supreme merit of satire in being reasonably fair, and was immensely popular, although it is not very well constructed. *The Medall*, also a political poem, followed in 1682, the same year as his *Religio Laici*, which indicated the evolution of Dryden's religious opinions. *The Second Part of Absalom and Achitophel* (1682) was by Nahum Tate, but contains about 200 of Dryden's most telling lines, such as

> They lost a villain and we gained a fool.

Shadwell, at whom the line is directed, had published his poem *The Medal of John Bayes*, attacking Dryden, earlier in the year. Dryden rarely wrote moving verse, but he did so in his lines to the memory of John Oldham published in Oldham's *Remains* (1684). In 1687 Dryden published *The Hind and the Panther*. Although this is an able argument in verse for Roman Catholicism, the best lines (apart from those with which the poem begins) are those which satirize Gilbert Burnet.

Among Dryden's greatest achievements are his "Pindarique" odes in irregular rhyming verse. He took this form from Abraham Cowley, who claimed to be its inventor, and who had used it effectively in his fine Odes to Hobbes and to the Royal society. Dryden gave it a new strength and dignity in his superb elegiac ode *To the Pious Memory of Mrs. Anne Killigrew* (1685) and the two memorable poems written for a London musical society, *A Song for St. Cecilia's Day* (1687) and the more famous *Alexander's Feast* (1697), masterpieces of dazzling virtuosity, in which colourful rhythm and magnificence of imagery combine with touches of wit and purple comedy to produce effects unique in English poetry. The songs scattered through Dryden's plays show great virtuosity in their different measures, but too often the substance is poor. His prologues and epilogues are perhaps the best in the language.

At the Revolution of 1688, Dryden lost his posts of poet laureate and historiographer royal. With unflagging vigour, however, in the last years of his life, he poured out a series of works including some fine plays, his famous translation of Virgil and finally the collection called *The Fables*, published in 1700, the year of his death, containing some of his most mature writing, including the three splendid verse tales founded on stories of Boccaccio, which were admired and imitated by Byron and Keats. His alliance with the publisher and anthologist Jacob Tonson was a literary event of considerable importance. The translation of Virgil published by Tonson was the first large-scale publishing venture in England and also probably the first which brought considerable profits to an author.

Older poets surviving from the reign of Charles I such as Abraham Cowley, Sir John Denham and Edmund Waller continued to produce competent verse after the Restoration but their best work had been done before 1660. More typical of the age was the poetry of the group of Restoration "court wits," which included Rochester, Sir Charles Sedley, Charles Sackville, earl of Dorset, and Sir George Etherege, nicknamed by Andrew Marvell "the merry gang." Sedley, Dorset and Etherege wrote admirable light verse continuing the courtly tradition of Thomas Carew and Sir John Suckling and also, at times, catching the note of the contemporary popular street ballad. Rochester was the one considerable poet among the group. His best lyrics have an intensity of

feeling and a purity of melodic form, which justify Sir Herbert Grierson's comment that they "might have been written by Burns with some difference." His verse satires are among the most notable of the century, ranging from "The Maimed Debauchee," well described by Charles Whibley as "that masterpiece of heroic irony," to the devastating political comments of *The History of Insipids* (1676), the brilliant social satire of *A Letter From Artemisa in the Town to Cloe in the Country* and his masterpiece, the powerful attack on facile optimism, *A Satyr Against Mankind* (1676). It is notable that all these works were apparently written before Dryden had published any of his satires and there is little doubt that he owed a considerable debt to Rochester's example. John Oldham, the young schoolmaster, who died at the age of 30 in 1683, was a *protégé* and imitator of Rochester. His early *Satyrs Upon the Jesuits*, the first of which was printed in 1679, the whole series appearing with some other poems in 1681, are marred by melodramatic ferocity, but his more mature poems have a fine vigour and freedom, and he must certainly be reckoned among the pioneers of the Augustan school.

Not all the good writing of the period came from London. Some of the great country houses were genuine centres of culture and excellent verse was written by a number of country gentlemen such as Charles Cotton, Thomas Shipman and Richard Leigh.

Much anonymous political satire was produced in the two decades preceding the revolution of 1688 and some of it, including the vigorous and colourful *Last Instructions to a Painter* (1667) was almost certainly by Andrew Marvell. The authorship of many of these lampoons, however, is uncertain as the law of seditious libel frequently necessitated their being circulated in manuscript until they could be collected (in 1689) into a series of pamphlets, later expanded into volumes known collectively as *Poems on Affairs of State*.

3. Drama.—The Restoration and revolution period is chiefly remembered for its plays. These were so many that only a few can be mentioned. The king and the duke of York gave their patronage to the drama. Drury Lane, known as the king's house, and Dorset Garden, known as the duke's, were founded soon after Charles's return. There were few real tragedies, their place being taken by heroic plays. These offsprings of the heroic romance and the French tragedy were generally a mixture of bombast with occasional fine passages; but their intention is never ignoble if often impossible. Dryden's *The Conquest of Granada* (1670) is the best known, but it had been preceded by the heroic plays of the earl of Orrery. Dryden was ridiculed in the amusing *The Rehearsal* (pub. 1672), by the duke of Buckingham and his collaborators. Of Dryden's adaptations of Shakespeare *All for Love* (1678) can be read with enjoyment if it is remembered that it is not an imitation of Shakespeare's *Antony and Cleopatra*. Dryden intimated that he was tired of rhymed verse in plays in the prologue to *Aurengzebe* (1676) and produced an excellent blank-verse play, *Don Sebastian*, in 1690 when the lapse of his posts had compelled him to return to the writing for the stage. Two of his comedies, *Marriage à la Mode* (published 1673) and *Amphitryon* (1690) are excellent. Nathaniel Lee mixed fustian with fine passages. Thomas Otway constructed his plays reasonably well and *The Soldier's Fortune* (1681) is still amusing. His masterpiece, which is still enjoyable though essentially a political melodrama, is *Venice Preserved* (1682), which has a wonderful tide of passionate language and a bold delineation of character. Sir Robert Howard's *The Duke of Lerma* (1668) has been underestimated. Real comedy began with Sir George Etherege, who mirrored in his fops the gaiety and insolence of the world he knew. His first play was produced as early as 1664 although he did not reach his zenith until *The Man of Mode* (1676). William Wycherley was a writer of comedy with a serious purpose: his *The Country Wife* (published 1675) in particular has great power while *The Plain Dealer* (published 1677) is a skilful transplantation, as well as a depravation, of Molière's *Le Misanthrope*. Wycherley scavenges in the most callous good spirits and with careful cynicism, and his prose is excellent, modern and lifelike. But the apex was reached by William Congreve, who, had his plots been coherent, would have been one of the world's greatest dramatists. As it is, his dialogue at its best, as seen in *Love for Love* (1695) and *The Way of the World* (1700), has hardly been surpassed, and never for sheer wit. Sir John Vanbrugh, the architect of Blenheim palace, comes within the Restoration period with *The Provok'd Wife* and *The Relapse* (both published 1697). Vanbrugh was a much less stylish writer than Congreve; Colley Cibber said his dialogue was "common conversation committed to paper." But his plays are full of vigour and humour. The coarse, powerful comedies of Thomas Shadwell, Dryden's butt, also deserve to be mentioned. They form a kind of link between the art of Ben Jonson, whom Shadwell greatly admired, and the realistic fiction of the age of Fielding.

4. Summary of the Period.—The Restoration period was quite unfairly regarded in the 19th century as sterile and uncreative. Actually it was a seminal period, a time of vigorous and fruitful experiment. Dryden, Rochester, Butler and Oldham can be regarded as founders of the great Augustan school and their works lead directly to those of Pope, Swift, Gay, Johnson and Goldsmith. The typical literary forms of the modern world—the novel, the biography, history and journalism—can all be traced to their beginnings in this period, when English literature can be seen as remaking itself in the face of new scientific and philosophic concepts and new social conditions. (H. M.; V. DE S. P.)

V. THE 18TH CENTURY

1. Essayists and Pamphleteers.—The remarkable development of English prose of all kinds in the 18th century owes much to the spread of popular journalism. In an age rational and argumentative by temper, the lapse of the Licensing act (1694), with its restrictions, whetted the appetite for controversy, while the increasing risk of prosecution for treason or sedition turned writers' minds to the literary modes of fiction, allegory and irony as loopholes for escaping punishment. Nearly a score of newspapers were in publication during the first decade, along with countless polemical tracts issued either singly or in series; and although little of the writing in organs such as the Tory *Postboy* (1695–1736) or the Whig *Flying Post* (1695–1731) deserves the name of literature, much of that in Lord Bolingbroke's *Craftsman* (1726–49?)—a periodical devoted to attacks on Sir Robert Walpole—or Defoe's *Review of the Affairs of France* (1704–13) assuredly does. Defoe, Swift, Fielding, Smollett, Johnson and Goldsmith all had firsthand experience of Grub Street, and the experience left its beneficial mark on their various styles. The clarity, cogency and sense of contact with the reader demanded by argument in print are everywhere to be found in the prose writing of the period.

The proximity of journalism and literature is to be seen clearly in the work of Daniel Defoe. In pamphlets such as *An Essay Upon Projects* (1698) and *The Shortest Way With the Dissenters* (1702) his clear, sensitive prose lifts the work above the level of immediate commercial or political occasions; while in the novels his shrewdness and devotion to facts leave the reader wondering whether these may be justly called works of imagination. Probably the central figures of his novels are all, in some measure, presentations of his own personality. In his *Journal of the Plague Year* (1722) he showed an astonishing power of projecting himself into a situation of which he had only read or been told; and when, at the age of nearly 60, he turned to novels, the same power of self-projection lent verisimilitude to his fictions.

If Defoe's method may be called imaginative reporting the fictional method of *The Tatler* (1709–1711) and *The Spectator* (1711–12) is the reverse. The members of the Spectator club are imaginary figures set in everyday surroundings as the mouthpieces of their authors' ideas about society. Nevertheless, though the conception of the Spectator club was a stroke of genius, one may doubt whether it is entitled to be called, as it sometimes is, the herald of the English novel. Once the fictitious framework had come into being it ceased to be of primary importance. Its function was to serve as a social microcosm within which to establish a tone at once grave, good-humoured and flexible. Having contrived this, the authors were free to consider what topics they pleased, either with reference to the fictitious framework, as in Richard Steele's account of Sir Roger de Coverley's views on marriage (No. 113), or without it, as in Joseph Addison's papers on

Paradise Lost (no. 267, 273, etc.). The achievement of *The Tatler* and *The Spectator* was to lift serious discussion out of the realms of religious and political partisanship and make it the normal pastime of the leisured middle class. Addison and Steele established both the pattern and the vogue of the periodical essay for the rest of the century; and although Parson Adams, of Fielding's *Joseph Andrews*, and Corporal Trim, of Sterne's *Tristram Shandy*, may owe nothing to Sir Roger or Captain Sentry, the influence of the papers as a whole in creating the public for which the novelists wrote, and in ensuring that the new kinds of prose writing, however entertaining, should be essentially serious, is considerable. Addison's reputation has far surpassed that of Steele, and perhaps with justification; for it was Addison who brought not only philosophy but literature out of closets and libraries to dwell at tea tables and coffee houses. But it was Steele, in *The Tatler*, who provided the first moral impetus, and his writing has a friendliness which affords needful emotional support for Addison's more dispassionate style. To assess their individual contributions to the success of *The Spectator* is, however, impossible. Their collaboration was more than the sum of its parts; separated neither was again at his best. Imitations of the two journals (*e.g., Tit for Tat*, 1710; *The Female Tatler*, 1709–10; *The Female Spectator*, 1744–66) indicate the far-reaching effect of the originals but bear no comparison with them.

The genuine Spectator tradition persists, with important differences, in Samuel Johnson's *The Rambler* (1750–52) and in Oliver Goldsmith's *The Citizen of the World* (1762). Intellectually Johnson's essays are more impressive than Addison's, but they lack the tact which enabled *The Spectator* to meet the reader on his own level without condescension. Johnson must be met on his own terms, namely the "inculcation of wisdom or piety." Goldsmith, though an inferior moralist, is freer in depicting humanity in its odd and touching aspects. As an essayist his is a midway position; he is more the exponent of personal sensitivity than Addison, but less so than Lamb.

During the second half of the century the periodical essay was gradually absorbed by the newspaper. Johnson's *Idler* essays appeared in *The Universal Chronicle*, Goldsmith's *Citizen* (as "Chinese Letters") in *The Public Ledger*. This period also saw the emergence of the first of the literary digests, Edward Cave's *The Gentleman's Magazine* (1731); and of the review devoted solely to criticism, of which kind *The Monthly Review* (1749–1845) and *The Critical Review* (1756–1817) were outstanding instances.

2. Swift.—Of the many writers whose work shows traces of the journalism of the day the most enigmatic and the ablest is Jonathan Swift. Swift was a man of subtle wit and wide reading, yet the greater part of his output was devoted to deriding the claims of the intellect. As secretary to Sir William Temple he came under the spell of the great Renaissance humanist and ironist Erasmus, through his share in Temple's quarrel with Richard Bentley on modern scholarship *versus* classical learning, and thenceforward assumed the task of ridiculing whatever he thought pretentious in religion, philosophy or science by the device of solemnly pretending to agree with his opponents, and then assuming that *they* will agree with *him* in his perversely logical reduction of their case to absurdity. The title of one of his most brilliant tracts, *An Argument to prove that the Abolishing of Christianity in England may, as things now stand, be attended with some inconveniences and perhaps not produce those many good effects proposed thereby* demonstrates the method. Here the irony is at its most deft. In *A Modest Proposal for Preventing the Children of Poor People From being a Burden to their Parents or the Country*, a pamphlet devoted to shaming the public into awareness of Irish poverty, it is at its most trenchant. Swift had a flair for adapting the forms of popular writing to his own purposes. Both these pamphlets are in the form of the projects or proposals which issued copiously from 18th-century presses. *Gulliver's Travels* (1726) is a "true narration" of the kind Defoe had used for *Robinson Crusoe;* in *The Battle of the Books* (printed 1704), a pamphlet ridiculing Richard Bentley's scholarship, use is made of the mock-heroic allegory, a form popular in political pamphleteering; in his most varied and ingenious work, *A Tale of a Tub* (1704), all these forms are exploited as well as a number of others, including the flattering literary preface made fashionable by Dryden. The effect of this device is to give Swift's writing an air of familiarity which renders his irony the more telling. Disappointed ambition may have made Swift's temper somewhat morose, but to call him the "master of hatred" on the strength of his claim (in a letter to Pope) to hate "all nations, professions and communities" is absurd, for in this he is only expressing his dislike of mass hysteria. To accuse him of "savage disgust" because he borrows man's less agreeable physical functions to symbolize man's moral shortcomings is to forget that this has been the allowed method of satire in all ages. Nobody loves an ironist, least of all one who makes the reader so acutely aware of his failings. The ironist shelters behind his irony; he cannot be pinned down. Swift is aware of this. That he does not see himself as a Houyhnhnm and the rest of mankind as Yahoos is evident from the fourth book of *Gulliver's Travels*, where Gulliver's ambiguous position is in some measure a representation of Swift's own. Swift's words flow direct from Swift's mind. His invention is inexhaustible. He could write brilliantly on any subject. Whether he is defending the Irish against Wood's halfpence (*The Drapier's Letters*, 1724) or exposing the claims of a charlatan (*Bickerstaff Papers*, or *Predictions for the Ensuing Year*, 1708) he is equally readable.

3. Criticism.—Closely related to developments in prose writing was the establishment of criticism, not as an occasional but as a continuous and widespread literary activity. As a result of the work of many hands the reputations of the major English poets—Shakespeare, Milton and, to a lesser degree, Chaucer and Spenser—were placed on a firm critical basis for the first time. The foundations of modern texts of Shakespeare were laid by the editions of Nicholas Rowe (1709; rev. 1714, the first standard edition), Pope (1723–25), Johnson (1765) and most notably by Lewis Theobald whose *Shakespeare Restored* (1726) (an attack on Pope's edition which gained him a place in *The Dunciad*) and edition of the plays (1733) represent pioneer work in textual criticism. Important editions of other poets were those of Spenser by John Hughes (1715) and Chaucer by John Urry (1721) and by Thomas Tyrwhitt (1775–78). Collections of earlier poetry were numerous, the best known being Bishop Thomas Percy's *Reliques of Ancient English Poetry* (1765). Antiquarian study is represented by the work of Thomas Hearne and by the *Thesaurus* (1705) of George Hickes, which preserves the only copy of the Anglo-Saxon *Finnsburg Fragment*. Johnson's *Dictionary of the English Language*, though not the first dictionary, introduced methods of illustration of the use of words which have changed little since 1755, when it was published.

The same signs of widespread interest and individual approach appear in the literature of appreciation. The critical tenets of Nicolas Boileau, René Rapin and Dominique Bouhours, though they are sometimes in evidence, are no longer all-powerful; for although critics such as John Dennis, Addison, Pope and even Johnson at times parade the neoclassical rules, they ignore them whenever their interest is fully engaged. In abstract criticism the principal subjects of speculation are the nature of poetic inspiration and aesthetic taste. Here the shaping spirit is that of Lord Shaftesbury, the influence of whose *Characteristics of Men, Manners, Opinions, Times* (1711, rev. 1713) is to be traced in most critics of the first half of the century. While deprecating enthusiasm or imagination in religion, Shaftesbury insists on its importance in literature as that which exalts the mind to serene contemplation. For Dennis imagination is that which produces passionate sublimity; for Addison it is the means of apprehending what is "great, uncommon or beautiful." The common element in these views is the preoccupation with sublimity; a concept vaguely apprehended but strongly felt not only in the criticism of the century but in the poetry also. As to taste, Shaftesbury posits a "moral sentiment" which draws man naturally to what is good and beautiful, and which it is the duty of the critic to cultivate and refine. Shaftesbury's deprecation of violence and partisanship, though it has been derided, did much to lift criticism above the level of religious or political squabbling in which it was too often involved. In this his even abler successor

was Addison, a sensitive rather than a perceptive critic, but one whose balanced approach to literature set an example which made easy the way for later critics. Of these the greatest is Samuel Johnson.

4. Johnson.—That Samuel Johnson once "rejoiced to concur with the common reader" should not blind us to the individuality of his criticism. He condemns those who judge by principle and precept rather than by perception and insists that literary judgment must be related to experience ("books cannot teach the study of books"). He was interested in "the biographical part of literature" because the life of the man of letters was the life he knew and because he believed that literature was only valuable if it clarified human experience. His greatness as a critic lies in his wide, humane reading and in the fact that his grounds for criticism are always apparent and never capricious. These grounds are, in a broad sense, moral. Gray's *Bard* "promotes no truth, either political or moral"; Shakespeare "keeps the high road of human feeling"; Addison's "instructions were such as the character of his readers made proper." In comments such as these one recognizes the consistency of Johnson's mind, and where his judgment seems to be at fault the failure is rather from inadequacy of experience than want of critical perception. For Johnson, as for most critics of his day, singularity implied deformity; nature was that which was common to the experience of all men, and nature was the business of literature. But whereas in many writers (*e.g.*, Pope and Reynolds) this insistence on generality sometimes arises from subservience to current aesthetic fashion, in Johnson it is the product of a settled outlook.

Johnson's position in English criticism may be inferred from the words of his *Preface to Shakespeare* (1765): "Let the fact be first stated, and then examined." In these words may be found the germ of the best English criticism since his time. That Johnson was able to judge in this way was due to the work of previous critics, Dryden, Pope (whose *Essay on Criticism*, 1711, sometimes anticipates the Johnsonian approach) and above all Addison, to whom Johnson pays a soberly glowing tribute in the *Lives of the English Poets*. But Johnson's predecessors, largely because of the critical temper of their age, had been concerned to fit literature to a theory; they could only cautiously and occasionally hint at a less a priori method. Johnson is the first English critic to insist consistently that a work is to be judged by its writer's purpose.

What made Johnson a great critic made him also the great conversationalist of Boswell's *Life* (*see* also below) and the great prose stylist of *Rasselas* (1759) and *The Rambler*. To call his prose heavy and Latinate is to ignore his aims. As a lexicographer he was always concerned with accuracy and, like Sir Thomas Browne on whom he modeled his style, he found in Latin derivatives the most precise and compact form of denotation. As philosopher and critic he combated "the inattention with which known truths are suffered to lie neglected" and his pointed, antithetical cadences are one of his weapons.

5. The Revolution of Taste.—During the second half of the century a change of taste and interest is evident, though it is neither so revolutionary nor so emphatic as is sometimes claimed. Sir Joshua Reynolds' *Discourses* (1769–91) uphold the neoclassical tradition of imitation and the analogy of poetry and painting, but Edward Young's *Conjectures on Original Composition* (1759) asserts the claims of original genius over learning. His position agrees with that of Thomas Warton whose *History of English Poetry* (1774–81) and *Observations on the Faerie Queene* (1754), along with *Letters on Chivalry and Romance* (1762) by Bishop Richard Hurd, demonstrate the growing interest in medieval literature, an interest to be seen elsewhere in the excitement aroused by the publication of James MacPherson's *Ossian* (1762) and the "medieval" poems of Thomas Chatterton. All these writings either directly or by implication stress the importance of fancy, but their originality lies rather in their emphasis than in any revolutionary critical approach. The importance of fancy and imagination had already been recognized by Shaftesbury and Dennis; and Johnson, when shown Young's *Conjectures*, was surprised to find that "he received as novelties, what he thought very common maxims." Perhaps the most genuinely revolutionary work was that of Thomas Tyrwhitt whose *Essay on the Language and Versification of Chaucer* (1775) destroyed for all time the illusion of Chaucer as a clumsy metrist. Revolutionary in another sense are William Blake's marginal comments on the *Discourses* of Sir Joshua Reynolds, but these are more in the nature of indignant outbursts than consistent criticism.

6. The Novel.—The growth of the English novel is no less indigenous than that of other 18th-century literary forms. Daniel Defoe's picaresque heroes have prototypes in the Elizabethan Thomas Nashe and Robert Greene, but *Moll Flanders* (1722), *The History of Colonel Jack* (1722) and *Captain Singleton* (1720) are not literary in conception; they are imaginative heightenings of actual rogue histories of the time and his most famous novel, *Robinson Crusoe* (1719), is based on fact. The novels of Samuel Richardson are no less indigenous. The originals of *Pamela* (1740) are not P. de Marivaux's *Vie de Marianne* or the *Abbé* Prévost's *Manon Lescaut* but Richardson's own *Letters written to and for Particular Friends* (publ. 1741), a compilation made at the request of friends to enable "country readers" to "endite for themselves" and prescribing not only "the requisite style and forms" but also "how to think and act justly, and prudently in the common concerns of human life."

If Richardson's emergence as an imaginative writer is partly fortuitous, Henry Fielding's is no less so. Unlike Richardson, Fielding was a man of refinement and extensive social experience. Irritated by the success of *Pamela* with its narrow concentration on moral respectability, he set out to ridicule the book in *Joseph Andrews* (1742). The book remains an amusing skit on Richardsonian morality to the end; but as the work advanced Fielding obviously found more in the form, and especially in the character of Parson Adams, than he had expected. His description of the somewhat episodic *Joseph Andrews* as the "comic epic poem in prose" better fits *Tom Jones* (1749) which shows him more clearly as the conscious artist. The introductory chapters prefixed to each book not only endow the novel with a critical rationale but they constitute an important technical advance, for in them the author himself appears as the shaping spirit of the story. Hitherto authors (even Richardson) had commended their stories by claiming historical accuracy; Fielding asserts the author's right to manipulate his narrative in the interests of artistic truth. To suggest, however, that Fielding's achievement is chiefly technical would be to ignore the unity of his artistic and philosophic outlook. Unlike Richardson, Fielding is a literary novelist. Behind everything he writes one senses the pressure of his reading, but this has been so fully absorbed that it is never obtrusive or pedantic.

Yet another journalist who found fame as a novelist is Tobias Smollett, who, in his life as in his novels, somewhat resembles Defoe. Like Defoe he was multifarious—naval surgeon, unsuccessful medical practitioner, hack writer, translator and historian—and he too knew prison from the inside. Such a life, varied and arduous, is likely to produce a picaresque novel if it produces one at all, and though Smollett claims Le Sage's *Gil Blas* as the ancestor of *Roderick Random* (1748) the literary heritage is less important than the author's own experience. His good-hearted rogues move through a series of loosely connected episodes, much as Defoe's; but whereas Defoe's characters are colourless to the verge of anonymity, deriving their individuality from their surroundings, Smollett's Hawser Trunnion and Tom Pipes are creations in their own right, the products of an acute observation heightened to the point of caricature by the author's fancy. This is Smollett's contribution to the English novel. He lacks either Richardson's sense of tragic predicament or Fielding's delicate awareness of human comedy. His picture of life, especially life in the navy, is crude and insensitive, uncompensated by any consistent moral attitude; but his flair for the peculiarities of human behaviour, best seen in his last novel *Humphry Clinker* (1771), has been matched only by Dickens.

The development of the novel in the 18th century was rapid and diverse. In the 40 years after *Robinson Crusoe* it accommodated not only the *vraisemblance* of Defoe but the tragedy of Richardson, the comedy of Fielding and the humours of Smollett.

Throughout its development, however, it had retained basically the same framework of a more or less chronological sequence of events tending to an outcome foreseen by the writer. It remained for Laurence Sterne to complete the process by radically altering the framework in the creation of the first impressionistic novel. *Tristram Shandy* (1759–67) presents life not as a series of cause and effect but as a flux of irrelevances without relationship except in the consciousness of the person experiencing them. The coherence of events thus depends solely on the associations they set up in the minds of Sterne's characters; and since, according to Locke, the association of ideas is irrational, the pattern of events must be equally so. In his brilliant exploitation of absent-mindedness Sterne found the means to both pathos and comedy, but it is in comedy (as when the birth of the hero results from an association of ideas in his father's mind) that the technique is seen at its most effective. Sterne's "formlessness" is, of course, apparent, not actual: in fact no writer knew better what he was about. That he had no direct successors in unorthodoxy is perhaps fortunate, but the substitution of an emotional for a logical framework admitted overtones which persist in the work of the later novelists.

Of the multitude of lesser writers of fiction who people the latter half of the 18th century Horace Walpole is perhaps historically the most important. His *Castle of Otranto* (1765) stands first in the line of Gothic novels which culminates in the next century in the prose romances of Scott and the verse romances of Byron. Others in this tradition are Clara Reeve (*The Champion of Virtue*, 1777), Ann Radcliffe (*The Mysteries of Udolpho*, 1794), Matthew Lewis (*The Monk*, 1796) and, ironically, Jane Austen (*Northanger Abbey*, begun 1798, published 1818). The Gothic romance is another aspect of the cult of sensibility to be seen in Oliver Goldsmith's *Vicar of Wakefield* (1766), Henry Mackenzie's *The Man of Feeling* (1771) and Henry Brooke's *The Fool of Quality* (1764–70). Fanny Burney as a novelist is something of a triton among minnows. *Evelina* (1778), her most successful novel, manifests the qualities which made her a great diarist—a spontaneous interest in life and the cool appraisal of social convention. She learned something from Smollett's portrayal of eccentricity, but hers is a firmer sense of structure.

7. The Poetry of Pope.—The early poetry of Alexander Pope is mainly literary in conception. In the poems collected in his first volume (1717) he handles the genres of "local poetry" (*Windsor Forest*), the didactic poem (*An Essay on Criticism*), the heroic epistle (*Eloisa to Abelard*) and the mock-heroic (*The Rape of the Lock*) with remarkable sensitiveness and technical dexterity. The same qualities are evoked in his later poetry by themes which are primarily moral or social. The epithet intellectual which is too often applied to him obscures the true emotional source of his poetry. This lies in the diversity of his response to different kinds of theme. That the subjects handled in the *Moral Essays* (1731–35), the *Imitations of Horace* (1733–38) and *An Essay on Man* (1733–34) should be prosaic (or at any rate rational) is to be expected, since he is eminently a poet of his age. He is not, like Donne, a poet in search of a personally satisfying philosophy, nor, like Wordsworth, a poet who has discovered one; he accepts the intellectual attitudes of his time. His response to these, however, far from being intellectual, is vividly imaginative. By means of imagery and allusion he endows an abstract notion with both a universal significance and an immediate field of reference which affect the reader's emotions as well as his mind. Artistically he is perfectly at ease in the looser Horatian mode of the *Satires and Epistles* and the *Moral Essays*. These poems, now that their topical reference has lost its immediacy, brilliantly reflect both Pope's imaginative power and his genuine moral feeling. Much of Pope's contemporary reputation rested on his translation of Homer's *Iliad* (1715–20) and *Odyssey* (with W. Broome and E. Fenton, published 1725–26). By rendering Homer in the poetic idiom of his day he inevitably misses some of the intensity of the original. Nevertheless the version, with its copious notes and preface, displays not only his great powers of generalized expression but also a considerable insight into the nature of Homer's genius.

8. Other Satiric Poets.—Apart from its rational tone there is little resemblance between Pope's poetry and that of his friends, the other members of the Scriblerus club. This informal gathering, which included Swift, Arbuthnot, Gay and Parnell, was dedicated to writing the memoirs of the imaginary pedant Martinus Scriblerus. Some, though not all, of the satires in the club's three volumes of *Miscellanies* were the result of collaboration; but in their poetry these writers were remarkably diverse. Many of Swift's poems (which are seldom in heroic couplets) are distinguished by a playful fantasy and racy idiom which might be deduced from the *Journal to Stella*, if not from *A Modest Proposal*. The reputation of John Gay rests mainly on *The Beggar's Opera*, but the delicacy and charm of his *Trivia, or the Art of Walking the Streets of London* (1716), a mock-Georgic, and the wit of his *Fifty-one Fables in Verse* (1727) deserve wider recognition. Though his intent may be satirical, his genius lies in giving elegant and pointed expression to commonplace matters. Of the other Scriblerians John Arbuthnot, physician and man of letters, survives not as a prose satirist or the author of the philosophical *Know Thyself*, but as the recipient of Pope's *Epistle*, and as the friend whose wit helped to fertilize *Gulliver's Travels;* Thomas Parnell, a fine classical scholar who assisted Pope in his translation of Homer, belongs as a poet to the reflective rather than the satirical tradition of Augustan poetry. Older contemporaries of Pope were Samuel Garth, whose mock-heroic poem *The Dispensary* (1699) lent something to *The Rape of the Lock*, and Matthew Prior, diplomat and accomplished occasional poet, the grace of whose lyrics entraps the unwary into thinking them trivial. Pope had many imitators but few successors. Charles Churchill, author of *The Rosciad* (1761), possessed something of his animus but nothing of his sensitiveness. Oliver Goldsmith, though hardly a satirist, borrows Pope's pointed manner in *The Deserted Village* (1770) but lacks his intellectual vigour. After Pope the one great satirist is Johnson, whose *London* (1738) and *Vanity of Human Wishes* (1749) have all and more of Juvenal's melancholy force, and a finality of phrasing which renders their originals (Juvenal's third and tenth satires) almost trivial by comparison.

9. Reflective Poetry.—The 19th-century picture of Pope tyrannizing over Augustan verse while an ineffectual resistance movement strove to defy him is entirely false. The contemplative vein, which reaches back into the 17th century and runs through Lady Winchilsea, James Thomson, Gray and Collins to Cowper, is the obverse of the satirical tradition, and quite as copious if less spectacular. The cult of solitude and the contemplation of nature's solemnities is explicit in Shaftesbury's writings, and as early as 1726 James Thomson's *Winter* was winning general approval. *The Seasons* (1730), the longer poem which developed from *Winter*, is not only vivid in its detail but dramatic in its portrayal of natural phenomena and the landscape of Teviotdale. Thomson, a true Augustan, meditates on man's place in nature, but it is nature which steals the picture. Like *The Seasons*, both in theme and in the somewhat mannered use of Miltonic blank verse, are John Dyer's *Ruins of Rome* (1740) and *The Fleece* (1757), but Dyer had not Thomson's natural strength either of description or versification. He is happiest in the shorter octosyllabic "Grongar Hill" and "The Country Walk," poems which reveal clearly Dyer's early training as a painter. In the same tradition, though not in blank verse, is Goldsmith's *The Deserted Village* (1770) which finely vindicates the Augustan theory of generalized description in its depiction of both human and nonhuman nature. William Cowper's *The Task* (1785) is perhaps the last in this line of succession. But unlike Thomson, who sees in nature a demonstration of God's purpose, or Goldsmith, for whom the landscape of Auburn is primarily a sentimental symbol, Cowper turns to the hills and snowbound woods of Weston as a relief from religious terror. *The Task*, with its discursiveness, its placid tone and its masterly yet unostentatious handling of language, has the comforting tranquillity of poetry written for its own sake, seldom touched by the ratiocinative urge of Thomson or the sensuous brilliance of Christopher Smart, whose *Song to David* (printed 1763) and *Hymns* glorify in loving detail the particularities (rather than the universal plan) of nature

as symbolizing God's love for man. Cowper is, moreover, without the melancholy sublimity exploited *ad nauseam* in the fashionable and popular *Night Thoughts on Life, Death and Immortality* (1742) of Edward Young and rather less oppressively in Robert Blair's *The Grave* (1743). Blair, despite his subject, has a certain natural cheerfulness, and his Shakespearian rhythms disprove the notion that all Augustan blank verse was Miltonic. Melancholy sublimity also touches, though delicately, the *Nocturnal Reverie* of Lady Winchilsea, the *Night Piece* of Parnell, the odes of William Collins and, above all, Thomas Gray's *Elegy Written in a Country Churchyard* (published 1751). This poem, in the universality of its themes and its dignified yet intimate tone, epitomizes the Augustan ideal of reflective poetry.

10. Blake and Burns.—The two great lyric poets William Blake and Robert Burns lie, in some measure, though not entirely, outside the traditions of the century. Blake's *Songs of Innocence* (1789) and *Songs of Experience* (1794) are indeed the descendants of the fine hymns of Isaac Watts; but what the latter perceived by the light of moral doctrine Blake perceived intuitively. The success of Robert Burns was in part due to the increasing interest in primitive poetry to be seen in the Wartons, Bishop Percy and William Shenstone but this interest was often scholarly or dilettante, whereas Burns's rhythms and vocabulary sprang, not from any poetic theory, but from the tunes and vernacular speech of his native Ayrshire. Burns was no uneducated thresher-poet; like all great poets he derived inspiration from his predecessors. But because his background was unacademic his reading nourished his genius, without subduing it to the requirements of "learning" which hampered even such poets as Gray. Like Blake, Burns is in revolt against human limitations, and it is this which distinguishes these two poets from others of their age.

11. Belles-Lettres and Learned Prose.—The increase of travel and a new sense of leisure and connoisseurship among the wealthier families speedily produced in England a crop of those elaborate collections of letters and memoirs which had attained their apogee in France with Mme de Sévigné and the duc de Saint-Simon. In 1709 began the letters of Lady Mary Wortley Montagu; 17 years more saw the commencement of Lord Hervey's *Memoirs of the Reign of George II,* while Lord Chesterfield and Horace Walpole both began their series of *Letters* about 1740. Less modish but more pleasing, the letters of Gilbert White, published in 1789 as *The Natural History and Antiquities of Selborne,* show that the countryman's loving eye for detail was not confined to the romantic period; and the same quality appears in the letters of Thomas Gray. His delight in early flowerings, the odd corners of literature, or the landscape of the Alps and the English lakes he shared, in notable language, with a few intimate friends. Stylistically at the opposite extreme is Chesterfield, who in treating of manners and social amenity deliberately seeks a form of expression which is the perfection of tact, good order and *savoir faire.* The most frivolous and also the most pungent of letter writers is Horace Walpole, whose writings are an epitome of the history and biography of the Georgian era. He is in some ways a corrective to the self-complacency of his generation, a vast dilettante, a lover of Gothic, of curios, costly printing, old illuminations and stained glass.

The art of the memoir may be diversely savoured in the *Apology* of Colley Cibber, comedian, poet laureate and prince of Pope's dunces; the *Memoirs of my Life and Writings,* edited from the historian Edward Gibbon's six drafts and published by the earl of Sheffield; the *Diary* of Madame d'Arblay (Fanny Burney); and, above all, in the *Life of Samuel Johnson LL.D.* (1791) by James Boswell. Hitherto Johnson's conversation has monopolized an unfair share of the credit for the success of this great work: now, after the discovery of the Malahide papers and the publication of his journals, Boswell is established not merely as a paragon of industry and a mine of information, but as a master of English prose.

Many types of literary production, now familiar, first achieved classical form during the Johnsonian period. Among them may be mentioned the economic treatise, exemplified in Adam Smith's admirably symmetrical *Wealth of Nations* (1776), the historico-philosophical tableaux of William Robertson and Edward Gibbon, the light political parody, of which the poetry of *The Rolliad* series (poems published in *The Morning Herald* and *The Daily Advertiser*) and the periodical *The Anti-Jacobin* afford excellent instances; and, at the other extreme, the ponderous archaeological or topographical monograph, as exemplified in James Stuart and Nicholas Revett's *The Antiquities of Athens,* Robert Wood's colossal *Ruins of Palmyra,* or the monumental *History and Antiquities of the County of Leicester* by John Nichols.

It was not until the third quarter of the century that English literature freed itself from the imputation of lagging hopelessly behind France, Italy and Germany in the serious work of historical reconstruction. David Hume published the first volume of his *History of England* in 1754. William Robertson's *History of Scotland* appeared in 1759 and his *History of the Reign of the Emperor Charles the Fifth* in 1769. Edward Gibbon's *The Decline and Fall of the Roman Empire* came in 1776–88. Hume was perhaps the first English historian to attempt to give his work a modern interest and a modern style. Robertson brought to bear on his studies the gifts of compromise which he had learned as moderator of the general assembly, and an academic English which he wrote with all the greater care because it was unfamiliar to him. Even more portentous in its superhuman dignity was the style of Edward Gibbon, who combined with the skepticism of Hume and Robertson an ordonnance and a pleasing malice all his own.

The increasing transparency in working prose of this period is shown in the writings of theologians such as Joseph Butler and William Paley, and of thinkers such as George Berkeley and David Hume. Butler and Paley are merely two of the ablest and most characteristic apologists of that day, both great stylists and both successful in repelling the enemy in controversy. Free thought, however, was beginning to sap their defenses in various directions, and in Thomas Paine, Joseph Priestley, Richard Price, William Godwin and Sir James Mackintosh they found more formidable adversaries than in the early deists. But the greatest champion of continuity and conservation, the great eulogist of the unwritten constitution and the perfect master of emotional prose was Edmund Burke, one of the most commanding intellects in the field of political letters—and in this respect a striking contrast to "Junius" (*q.v.*), the pseudonym of an unidentified author, whose mechanical and journalistic talent for invective had a quite ephemeral value.

12. Drama.—An age of great actors, the 18th century was poor in drama. With the exception of Fielding, and perhaps Goldsmith, the men of literary genius, though many wrote plays, were not men of the theatre; and competent dramatists such as Nicholas Rowe, Richard Cumberland and John Home, whose *Douglas* (acted 1756) retained its popularity for a century, were not men of literary genius. The work of Richard Brinsley Sheridan alone possessed both the literary and theatrical qualities necessary for great drama. In *The Rivals* (acted 1775) and more especially in *The School for Scandal* (acted 1777) he commands Restoration wit without its lubricity, and the fashionable poseurs of his age (notably the "man of sentiment") are brisker and more honest comic game than rakes and cuckolds.

In any age these two plays would have been distinguished: their complete pre-eminence in the 18th century may be seen by comparing them with the one play that even approaches them—Goldsmith's *She Stoops to Conquer* (acted 1773). (N. C.)

VI. THE ROMANTIC PERIOD

1. The Nature of Romanticism.—The term "the romantic movement," embracing some but by no means all of the writers who flourished in the last years of the 18th century and the first decades of the 19th, is somewhat misleading: it was hardly a "movement," and the word "romantic" is so vague and general as to be almost meaningless. The great writers of the period differed widely, not only in the nature of their work but also in their critical approach. They did not apply the word "romantic" to themselves, and the recognition they met with, both from the general public and from the critics, was slow; as J. R. Sutherland has

put it, "the notion of a parched and thirsty nation waiting eagerly for the prophet Wordsworth to smite the rock and let forth the living water is at variance with the facts. Wordsworth had first to persuade the public that it was thirsty before he could get it to drink his pure mountain streams at all." (*A Preface to 18th-Century Poetry*, 1948.)

With these reservations in mind, it may be possible to isolate some of the distinguishing features of English romanticism. These can best be seen in the poetic theory and practice, and have obvious correspondences with the political and social upheavals of the time: such movements as the American and the French Revolutions, with their emphasis on natural rights and the importance of the individual, are reflected in the attitudes of writers, and are implicit in many of their statements about poetry. Where the main trend of 18th-century poetics had been to praise the general, to see the poet as a spokesman of society, addressing a cultivated and homogeneous audience, and having as his end the conveyance of "Truth," the romantics found the source of poetry in the particular, unique experience, and saw the poet as an individual distinguished from his fellows by the intensity of his perceptions, taking as his basic subject-matter the workings of his own mind. The implied attitude to an audience varied accordingly: although Wordsworth maintained that a poet did not write "for Poets alone, but for Men," for Shelley the poet was "a nightingale who sits in darkness and sings to cheer its own solitude with sweet sounds" and Keats declared "I never wrote one single line of Poetry with the least Shadow of public thought." As for the purpose of what was written: poetry was regarded as conveying its own truth; sincerity was the criterion by which it was to be judged. Provided the feeling behind it was genuine, the resultant creation must be valuable.

The emphasis on feeling is found in most of the romantic definitions of poetry. Wordsworth called it "the spontaneous overflow of powerful feelings," and in 1833 John Stuart Mill defined "natural poetry" as "Feeling itself, employing Thought only as the medium of its utterance." It followed that the best poetry was that in which the greatest intensity of feeling was expressed, and hence a new importance was attached to the lyric. The degree of intensity was affected by the extent to which the poet's imagination had been at work; as Coleridge saw it, the imagination was the supreme poetic quality, a quasi-divine creative force which made the poet a godlike being. Romantic theory thus differs from the neoclassic in the relative importance it allotted to the imagination: Samuel Johnson had seen the components of poetry as "invention, imagination and judgement" but William Blake wrote "One Power alone makes a Poet: Imagination, the Divine Vision." The judgment, or conscious control, was felt to be secondary; the poets of this period accordingly lay great emphasis on the workings of the unconscious mind, on dreams and reveries, on the supernatural and on the childlike or primitive view of the world, this last being regarded as valuable because its clarity and intensity had not been overlaid by the restrictions of civilized "reason." A further sign of the diminished stress placed on judgment is the romantic attitude to form: if poetry must be spontaneous, sincere, intense, it should be fashioned primarily according to the dictates of the creative imagination. Wordsworth advised a young poet, "You feel strongly; trust to those feelings, and your poem will take its shape and proportions as a tree does from the vital principle that actuates it." This organic view of poetry is opposed to the classical doctrine of "genres," each with its own linguistic decorum; and it led to the feeling that poetic sublimity was unattainable except in short passages—a view also expressed by Edgar Allan Poe in his statement that "The absolute effect of even the best epic under the sun is a nullity."

Hand-in-hand with the new conception of poetry and the insistence on a new subject-matter went a demand for a new diction. Wordsworth and his followers, particularly Keats, found the prevailing poetic diction of the later 18th century stale and stilted, or "gaudy and inane," and totally unsuited to the expression of their perceptions. It could not be, for them, the language of feeling, and Wordsworth accordingly sought to bring the language of poetry back to that of common speech. His theories of diction have been allowed to loom too large in critical discussion: his own best practice very often differs from his theory, and he should have realized that, as Coleridge was to point out, it applies only to certain kinds of poem, and that the great poets have always departed from "the real language of men" when they chose. Nevertheless, when Wordsworth published his preface to *Lyrical Ballads* in 1800, the time was ripe for a change: The flexible diction of earlier 18th-century poetry had hardened into a merely conventional language and, with the notable exception of Blake and Burns, little first-rate poetry had been produced in Britain since the 1740s.

The insistence on the importance of the individual led to an extension, in some directions, of literary sympathies. Most prominent and characteristic was the concern with external nature, which may be connected with the social changes being brought about by the Industrial Revolution—especially with the movement of the population from rural areas into rapidly expanding, ugly cities. In the presence of Nature, man considered himself closer to God; in a simple environment, spontaneous feeling was still possible. In that condition, said Wordsworth, "our elementary feelings co-exist in a state of greater simplicity . . . the passions of men are incorporated with the beautiful and permanent forms of nature." Furthermore, Nature was felt to possess a mystical significance, to be a kind of divine hieroglyph. Communion with, and understanding of, Nature led the poet toward an understanding of his own soul. A further area for the exploration of the workings of the human mind was the supernatural and marvelous. The strange and mysterious was a prime source of sublimity, and the remote, whether in time or space, formed the subject-matter or the setting of much romantic writing.

Neither the suddenness nor the novelty of romanticism should be exaggerated. All the critical theories put forward are to be found in the 18th century and earlier—most important, perhaps, being the 18th-century interest in Longinus and the concept of sublimity—but the main difference lies in the shift of emphasis: what had been marginal in earlier theory became central, and vice versa. It must be emphasized, moreover, that many of the writers of this period—Jane Austen, for example, or George Crabbe or Thomas Love Peacock—can hardly, by any stretch of the word, be called romantic, while much of what seems to be most enduring in the novels of Scott is nonromantic. Again, the great romantics differed among themselves: Byron was eloquent in his praise of 18th-century poetry, and at times vehement in his attacks on Wordsworth and Keats; Shelley found Wordsworth lacking in imagination; and Blake disagreed sharply with many of Wordsworth's critical theories.

2. Poetry.—*Wordsworth and Coleridge.*—The meeting in 1795 of William Wordsworth and Samuel Taylor Coleridge was the beginning of one of the most fruitful of all literary friendships; during their early and most productive years they acted as a constant stimulus to each other. The anonymous appearance, in 1798, of their *Lyrical Ballads* marked a new departure in English poetry, and Wordsworth's preface to the enlarged edition of 1800, defending his "experiments," was a manifesto of the new outlook. Wordsworth's contribution to the volume was, they had agreed, to be mainly poems "giving the charm of novelty to things of every day," and it is to these that his remarks on diction and subject-matter mainly apply, and these which aroused most hostility. Certainly, in the earliest versions of these poems, Wordsworth was not always sufficiently sure of himself: his extreme simplicity and matter-of-factness sometimes lead him into bathos, and he does not always succeed in raising them above the level of anecdote; but at their best, as in "The Idiot Boy" or some of the lyrics added in 1800, they have an extraordinary immediacy and power.

In his account of the work Coleridge says "Wordsworth added two or three poems written in his own character, in the impassioned, lofty, and sustained diction which is characteristic of his genius"; foremost among these (and, with Coleridge's "Ancient Mariner," the finest of the poems in the volume) is "Tintern Abbey," a poem in blank verse in which Wordsworth gives superb expression to his feeling for nature, and the mystical "presence," the "sense sublime / Of something far more deeply interfused"

that guides, sustains and consoles him. All Wordsworth's best work is suffused with this attitude to nature. He has a remarkable gift for conveying the inner significance of a scene or an apparently trivial event. As he explained in his long autobiographical poem *The Prelude* (1805; later revised, and published 1850), there are in our existence "spots of time"—moments when, with sudden clarity, we "see into the heart of things," and by which "our minds are nourished and invisibly repaired." His ability to present these moments is his greatest strength. Poetry, he wrote, "takes its origin in emotion recollected in tranquillity"; the quiet meditative gravity with which he is able to recall and recreate his moments of luminous perception forms a perfect illustration of this statement. Equally notable are the solemn dignified pathos of "Michael" (1800), the moving simplicity of the "Lucy" poems, the excitement of many passages of *The Prelude*, and the eloquence of the ode "Intimations of Immortality" (1807). Wordsworth is an uneven poet: when his imagination fails he can be banal or silly, or indulge in empty moral rhetoric and inflated diction; but of the greatness of his achievement there can be no doubt.

A meditative quality also informs some of Coleridge's best work: in "Frost at Midnight" (1798), for example, he relates his state of mind to a particular time and place, moving easily and conversationally through an association of ideas. The power of a less consciously controlled association is seen in his great "magical" poems—"The Ancient Mariner," "Kubla Khan" (1797; published 1816) and "Christabel" (1797–1800; published 1816); 20th-century studies have demonstrated the extent to which Coleridge's amazingly wide reading is incorporated in these poems. "The Ancient Mariner" is his finest achievement, with its narrative skill, its evocation of a supernatural, nightmarish atmosphere, its richness of image and symbol, and its complex and all-pervading moral tone. A fine poem of a very different kind is "Dejection: an Ode" (1802), in which Coleridge returns to a more meditative manner to lament his spiritual apathy, the loss of his poetical faculty, the "shaping spirit of Imagination." This poem is, of course, paradoxical, in that it is a great poem about the inability to write poetry; but it is true to say that, for whatever reason, Coleridge's poetic gift began to fade about this time. He was a man of immense potentialities, few of which were realized; yet his achievement remains a great and influential one. Hazlitt said of him "If Mr. Coleridge had not been the most impressive talker of his age, he would probably have been the finest writer." His *Table Talk* (published 1835) gives some idea of his gifts and the range of his interests, as do his lectures, essays, and collections of notes such as *Anima Poetae* (published 1895)—theology, politics, philosophy, literature are all discussed. Most important are his *Biographia Literaria* (1817), a rambling but extremely stimulating discussion of poetry and criticism, and his essays and lectures on Shakespeare. He was one of the most intellectually formative writers of his day, and has had a profound influence on later criticism.

Other Poets of the Early Romantic Period.—Popular though they once were, the lesser poets of this generation merit little discussion now. The sonnets (1789) of William Lisle Bowles were received with great enthusiasm by Coleridge and Wordsworth, and may have acted as an inspiration to them, but they are flat and insipid, though pleasant. Thomas Campbell is chiefly remembered for his patriotic lyrics such as "Ye Mariners of England" and "Hohenlinden" (1807), and for the critical preface to his *Specimens of the British Poets* (1819). Samuel Rogers has survived for his brilliant table talk (published 1856, after his death, as *Recollections of the Table Talk of Samuel Rogers*), rather than for his poetry. One of the most popular poets of the day was Thomas Moore, whose *Irish Melodies* began to appear in 1807; the income that a successful author could command in the early 19th century is indicated by the fact that he got 100 guineas for each of his 130 Irish songs. His highly coloured, oriental fantasy *Lalla Rookh* (1817) was also immensely popular. Robert Southey was closely associated with Wordsworth and Coleridge, and was looked upon as a prominent member, with them, of the "Lake School" of poetry; but his grandiose epic poems, such as *Thalaba the Destroyer* (1801) and *The Curse of Kehama* (1810), scarcely repay the reading. Some of his shorter poems, such as "The Battle of Blenheim," still live; but his fame is based on his prose works—the vigorous *Life of Nelson* (1813), the *History of the Peninsular War* (1822–32) and his children's tale, "The Three Bears."

A reminder of how various was the literary achievement of this period is provided by George Crabbe. He has been called the last of the Augustans, and his roots—his sensibility, his values, much of his diction, his heroic couplet verse-form—are very firmly in the 18th century. He differs from the earlier Augustans, however, in his subject-matter, concentrating on realistic unsentimental accounts of the life of the poor and the middle-class. He shows considerable narrative gifts in his collections of verse tales (in which he anticipates many short-story techniques) and a great power of description, which is put to atmospheric and psychological use. His main works, *The Village* (1783), *The Borough* (1810), *Tales in Verse* (1812) and *Tales of the Hall* (1819), gained him great popularity in the earlier 19th century (the publisher John Murray gave him £3,000 for the last-named work); and after a long period of neglect he is widely recognized once more as a major poet.

The Later Romantics: Shelley, Keats and Byron.—When *Lyrical Ballads* appeared, Shelley was six years old, Keats three and Byron ten; they grew up in the shadow of the Napoleonic Wars and, for them, the overthrow of Napoleon in 1815 was an event comparable to the storming of the Bastille, in 1789, for the older poets. Shelley in particular was passionately interested in politics, coming early under the spell of the anarchistic views of William Godwin, whose *Political Justice* had appeared in 1793. This revolutionary ardour, a somewhat vague zeal for the liberation of mankind, is the basis of Shelley's critical essay *The Defence of Poetry* (1821) in which he eloquently claims that "the most unfailing herald, companion, and follower of the awakening of a great people to work a beneficial change in opinion or institution, is poetry," and in which poets are hailed as "the unacknowledged legislators of the world." This fervour burns throughout the early *Queen Mab* (1812), the long *Revolt of Islam* (1818) and the noble lyrical drama *Prometheus Unbound* (1820). Shelley is the poet as prophet, as the fine *Ode to the West Wind* (1819) makes clear; but with his prophetic spirit goes a vagueness which too often results in a blurring of the effect. His favourite words and images are "spirit," "breath," "being," "cloud," "faint," "sweet," "bright," "domes," "stars"; the celestial machinery of rainbows, visions and atmospherics is perhaps employed to excess; his epithets tend to be nebulous and imprecise. He is, in fact, an emotional poet, and critics have complained of his "weak grasp upon the actual," of the fact that he frequently seems simply to strike an attitude and let poetic clichés come pouring out. This is true of his weaker work: but Shelley should not be judged as if he were trying to write like Donne or Keats. His material is transcendental and cosmic, and his expression is appropriate to it. He had great technical brilliance ("Shelley is one of the best *artists* of us all," said Wordsworth in 1827), and, though the elements of self-pity and immature posturing in him have caused a decline in his reputation, he is, at his best (as in many short lyrics, the "Letter to Maria Gisborne," the "Lines Written Among the Euganean Hills," much of *Adonais*, 1821, "Ozymandias," and "The Triumph of Life"), a poet of compelling excitement and power.

The reputation of John Keats is less liable to fluctuation, supported as it is by his superb letters. He is a richly sensuous poet, whose early work, such as *Endymion* (1818) was overloaded with imagery—"a trial of my Powers of Imagination" he called it—and produced an over-luxuriant, cloying effect. But Keats was able to discipline himself: even before Feb. 1820, when he first began to cough blood, he must have known that he had not long to live, and he devoted himself to the perfection of his art with feverish intensity. He experimented with many kinds of poem: *Isabella* (published 1820), an adaptation of a tale by Boccaccio, is a tour de force of craftsmanship, a deliberate attempt to reproduce the medieval atmosphere; his epic fragment, *Hyperion* (begun in 1818 and abandoned, published 1820; later begun again

as *The Fall of Hyperion*, 1819) has graveness and sonority, but Keats found the style too Miltonic—"I wish to give myself up to other sensations" he wrote. Some of these "other sensations" are found in the poems of Keats's *annus mirabilis*—1819—"The Eve of St. Agnes" and the great odes, "To a Nightingale," "On a Grecian Urn" and "To Autumn." These, with the revised *Hyperion*, represent the summit of Keats's achievement, showing what has been called "the disciplining of sensation into symbolic meaning," the complex themes being handled with a concrete richness of detail.

The last of the "great five" romantic poets, Byron, presents important differences from the others. In *Don Juan* he wrote, without irony:

> Thou shalt believe in Milton, Dryden, Pope;
> Thou shalt not set up Wordsworth, Coleridge, Southey,

and his antipathy to Keats has already been mentioned. Moreover, there are striking contrasts within his own work: in *Childe Harold* (1812–18) he appears as the archetypal romantic, bearing "the pageant of his bleeding heart" through Europe; and the element of melodrama, of histrionic misanthropy, appears again in such poems as *Manfred* (1817) and *Cain* (1821). Very different is the detached, ironic Byron, a master of witty, almost impudent, satire. His poems in *ottava rima*, notably *Beppo* (1818) with its colloquial ease and gaiety, *Don Juan* (1819–24), a kind of gay and comic picaresque novel in verse, and *The Vision of Judgement* (1822) represent, with a handful of lyrics, his lasting achievement.

Minor Poets of the Later Period.—Of the lesser poets of this generation the best is undoubtedly the Northamptonshire peasant John Clare. He led a wretched life of poverty and eventual madness, marked, however, by a noble dedication to his poetic gifts. He has a natural simplicity and lucidity of diction, an intent observation and an almost classical poise; and the unassuming dignity of his attitude to life makes him one of the most quietly moving of English poets. Thomas Lovell Beddoes, whose violent imagery and obsession with death and the macabre recalls the Jacobean dramatists, represents an imagination at the opposite pole from that of Clare. He had considerable metrical virtuosity, seen at its best in the songs and lyrical passages from his shapeless tragedy *Death's Jest-Book* (begun 1825, published 1850), but his sensationalism becomes tiring. Another minor writer who found his inspiration in the 17th century was George Darley, some of whose songs from *Nepenthe* (1835) keep their place in anthologies. The comic writer Thomas Hood once enjoyed a great vogue, but is now little read; but such poems of social protest as *The Song of the Shirt* (1843) and "The Bridge of Sighs," and the graceful *The Plea of the Midsummer Fairies* (1821), are by no means negligible.

3. The Novel.—At the turn of the century the cult of sensibility, manifested in the novel of feeling and of terror, dominated prose fiction. The Gothic novel, with its elements of German romanticism, lingered on in such writers as Mary Shelley (the daughter of William Godwin), whose *Frankenstein* (1818) deals, among its scientific horrors, with themes of social injustice and the "noble savage"; and Charles Maturin, whose *Melmoth* (1820) has, with all its absurdity, a striking intensity. But the novelists dealt with the supernatural far less adequately than did the poets, and their works were ripe for deflation. The spirit of parody, indeed, formed a starting point for the writing of Jane Austen, one of the greatest of English novelists. She early perceived the falsity of the novel based on a carefully cultivated emotional response, and set out to demolish its pretensions by burlesque and irony. *Northanger Abbey* (begun 1797, published 1818) satirizes the Gothic novel, among other things, with complex irony, and *Sense and Sensibility* (also begun 1797, published 1811) examines the qualities indicated by its title with sympathy and understanding. As her novels grow in depth, so do her moral perceptions; her art is perhaps seen at its finest in *Emma* (published 1816). She is the spokesman for sanity and intelligence, and, like every good satirist, she builds up positive values against which the objects of her attack can be judged. She is by no means a simple novelist: the limitation suggested by her narrow range of setting and characters is illusory; working within these chosen limits, she observes and describes very closely all the subtleties of the relationships between people. She continually analyzes, seeking to discover the principles of conduct that animate us; but her analysis is predominantly sympathetic, her vision basically comic. She is a master of dialogue, and writes with great economy, hardly ever wasting a word.

The same cannot be said of the sprawling talent of Sir Walter Scott. His earlier literary career is as a successful poet, with such narrative historical romances as *The Lay of the Last Minstrel* (1805), *Marmion* (1808) and *The Lady of the Lake* (1810); then he turned to prose with *Waverley* (1814), a tale of the Jacobites. In the next 15 years he wrote more than 20 novels; taken all together, they are a great achievement, and perhaps are best judged by their cumulative effect. Scott is at his best when dealing with recent Scottish history, as in *Old Mortality* (1816), *The Heart of Midlothian* (1818) and *Redgauntlet* (1824), but there are good things in all the novels, though too often the reader has to work his way through a pseudo-antique diction. Scott had great dramatic gifts, but little sense of form: he admitted in his journal that he could never manage a well-contrived story because "Woe's me, the ideas rise up as I write." The ideas vary in quality and maturity, and the reader is continually conscious of the lack of discipline; but Scott has many touches of greatness.

An entirely different approach to the novel was that of Thomas Love Peacock, a satirist for whom plot meant practically nothing; all his novels are conversation pieces, in which many of the pretensions of the day are laid bare in the course of witty, animated, and genial talk. *Nightmare Abbey* (1818) exposes the tawdriness of many of the trappings of romanticism with great good humour and, by including as characters, under thin disguises, such writers as Coleridge, Byron, and his intimate friend Shelley, Peacock is able to poke fun at some elements of romantic poetry. The more serious side of his satire is shown in such passages as Mr. Cranium's lecture on phrenology in *Headlong Hall* (1816). His seven novels have a lasting freshness and vitality. Among the lesser novelists may be mentioned Maria Edgeworth, whose realistic didactic novels of the Irish scene inspired Scott; Susan Ferrier, a Scot with her own vein of racy humour; and another Scot, John Galt, whose *Annals of the Parish* (1821) is one of the minor works that deserve to survive. This judgment applies also to the remarkable *Confessions of a Justified Sinner* (1824) by James Hogg, a powerful story of Calvinism and the supernatural.

4. Miscellaneous Prose.—The romantic emphasis on individualism is reflected in much of the prose of the period, particularly in criticism and the familiar essay. Among the most vigorous and forthright, and least mannered, writing is that of William Hazlitt. He is an energetic, enthusiastic and subjective critic, wide in his range, though lacking in sympathy when writing of the 18th-century poets. His most characteristic work is seen in his collections of lectures *On the English Poets* (1818) and *On the English Comic Writers* (1819), and in *The Spirit of the Age* (1825), which is especially valuable for its portraits of Hazlitt's contemporaries. Charles Lamb is an even more personal essayist, a zealous but undiscriminating critic. In the *Essays of Elia* (1823) and *Last Essays of Elia* (1833) he projects, with apparent artlessness, a carefully managed portrait of himself—charming, whimsical, witty, sentimental, warm-hearted, nostalgic, sociable. Confession of a different kind is provided by Thomas De Quincey, a self-conscious stylist whose addiction to opium resulted in the sensational and eloquent *Confessions of an English Opium Eater* (1821), a morbid, nightmarish, but often magnificent work which has kept its popularity. His shrewd *Reminiscences of the English Lake Poets* (1834–39) is of great interest to the student of the period. A critic and essayist of lighter weight is Leigh Hunt, the intimate and loyal friend of Keats. His outspoken journalism, particularly in his *Examiner* (1808–21), was of considerable influence.

Of the writers who might be called surviving classicists, the best is Walter Savage Landor, whose detached, lapidary style is seen at its best in some brief lyrics, and in the series of erudite *Imaginary Conversations* which began to appear in 1824. The

classic, or rather the anti-romantic, point of view received its most pungent expression in the pages of the quarterlies: the Whig *Edinburgh Review* (begun 1802), edited by Francis Jeffrey, was followed by its Tory rivals the *Quarterly Review* (1809) and the monthly *Blackwood's Magazine* (1817). All three expressed clearly, trenchantly, and often savagely, their adverse views of the romantic poets. Their criticism was by no means always unjust, and summed up much contemporary opinion; but the reviewers were too willing to judge the new poetry by their own rigid and settled standards, and had insufficient ability or desire to separate the good from the bad. These journals were not, however, confined to literature, and in their attacks on many kinds of prejudice and abuse they set a notable standard of fearless and independent journalism. A similar independence was shown by William Cobbett, whose *Rural Rides* (collected in 1830 from his *Political Register*) gives a picture, in forceful and clear prose, of the English countryside of his day.

5. Drama.—The drama of the period is hardly worth serious consideration. All the great romantic poets essayed tragedies or "lyrical dramas," and one or two of these, such as Coleridge's *Remorse* and Byron's *Marino Faliero*, made brief appearances on the stage; but on the whole these plays were unacted and often virtually unactable. The poets, and such writers as Lamb and Leigh Hunt, had little dramatic gift or knowledge of stagecraft, they tended to look down on the contemporary theatre, and their preoccupations were with themes of the past rather than the present, so that much of their work seems like a pastiche of Jacobean drama. A partial exception is Byron, who, as one of the management committee of Drury Lane, was interested in the theatre and had some talent for it, but even his dramas are lacking in any vitality. Shelley's *The Cenci* (1819; first acted 1886) has some power and passion, but it is crudely melodramatic and has so many echoes of Shakespeare as to read almost like a burlesque.

As for the "acting theatre," the great bulk of performances, both at the two patent theatres of Covent Garden and Drury Lane, and at the many minor theatres, were of melodramas, farces, pantomimes and extravaganzas, spectacularly staged before rowdy and sensation-hungry audiences. Writers of more serious plays, such as the dully conscientious Joanna Baillie with her *Plays of the Passions* (1798–1812), showed little talent; and although some of the farces of the day are lively enough, the "regular" comedies are solemn and dreary. The early 19th century, in fact, failed to produce either tragedy or comedy of any lasting worth; against this lack, however, might be set the introspective and often illuminating dramatic criticism of such writers as Coleridge, Hazlitt and Lamb. (R. P. C. M.)

VII. THE 19TH CENTURY

1. The Aftermath of Romanticism: Tennyson.—It is common practice in literary histories and in university syllabuses to divide the 19th century into two periods called respectively "romantic" and "Victorian," of which the first, beginning in the 18th century, is taken to have ended about 1830–32. Keats had died in 1821, Shelley in 1822, Byron in 1824 and Scott in 1832; and in 1830–32 Tennyson, the representative Victorian poet, published his first volumes of verses. Coleridge, though he lived on another couple of years, had done all his best work long before; and the same is true of Wordsworth, although he survived into the heart of the century (1850). This division is convenient and has been adopted in the present article. Yet there was no real break; 1830–32 is not a turning-point in cultural history like 1660 or 1798, or like 1914 (or 1920) when the 19th century came to its true end. The story of the century is a story of continuous development—or let us say "evolution," since that is the key-word of 19th-century thought. The impress of the romantics and particularly of the first-generation romantics—Wordsworth, Coleridge and Scott—is upon all the work of their Victorian successors. It was, above all, Wordsworth and Coleridge who had fixed the notions about the nature and functions of the poetic imagination, about poetic subject-matter and diction and about the dignity of the poet's office, which dominated the rest of the century. Wordsworth's description of poetry as "the spontaneous overflow of powerful feelings"; his account of the poet as "a man speaking to men: a man, it is true, endowed with more lively sensibility, more enthusiasm and tenderness . . . than are supposed to be common among mankind; a man pleased with his own passions and volitions . . . , delighting to contemplate similar volitions and passions as manifested in the goings-on of the Universe"; his avowal of the purpose of his poetry as being "to teach the young and the gracious of every age to see, to think, and feel, and, therefore, to become more actively and securely virtuous"; Coleridge's statement of "the two cardinal points of poetry, the power of exciting the sympathy of the reader by a faithful adherence to the truth of nature, and the power of giving the interest of novelty by the modifying colours of the imagination"—these were central articles of the poetic creed to which nearly all the later poets of the century, and virtually all their readers, subscribed. Poetry was to reveal truth, and the poet was revered as prophet and priest—typically (to use Matthew Arnold's phrase about Wordsworth)

> a priest to us all
> Of the wonder and bloom of the world.

Never, before the 19th century, was it so generally taken for granted that nature—in the sense of the outward world—should be a poet's main source of inspiration and the main source of edification to the reader. Hence the extraordinary success attained by many writers of the century in rendering nature with fidelity and insight—a success not confined to the poets, though seen to great advantage in Tennyson, Matthew Arnold, Gerard Manley Hopkins, George Meredith and others, but found also in prose-writers such as Emily Brontë, John Ruskin, George Eliot, Charles Kingsley, Thomas Hardy, R. L. Stevenson and others. From Tennyson's "jewels five-words long"—"Laburnums, dropping-wells of fire"—to Hopkins' "mealed-with-yellow sallows" or "Fresh-firecoal chestnut falls"; or from Ruskin's description of his first sight of the Alps to Hardy's Egdon Heath—wherever we turn, 19th-century literature surpasses in command of "natural magic" that of any preceding century.

Behind all this lay the nature religion bequeathed to the century by Wordsworth; nature was holy and morally uplifting, and the more successful a writer was in penetrating its inmost life, or "disimprisoning the soul of fact" (as Carlyle put it), the more good he was doing to mankind. Moreover in spite of the Oxford movement and other manifestations of religious vitality the century did witness an immense falling-away from the traditional religious certainties and Wordsworthianism furnished many doubters with an interim religion. Meanwhile the increasing urbanization of English life and the defilement of the countryside by industry created a longing for escape and compensation, and readers expected their poets and novelists to restore to them "the freshness of the early world" and to lay them "where they lay at birth./On the cool flowery lap of earth." Thus much of Victorian literature is agnostic at heart, yet nature-loving, highminded and morally serious. Tennyson is the representative poet of the age because he expressed most completely its misgivings as well as its underlying assumptions. *In Memoriam*, it has been truly said, represents the 19th-century mind just as Milton's *Paradise Lost*, Pope's *Essay on Man* and Wordsworth's *The Prelude* represent their respective periods. Immortality, progress, the providential scheme and the beneficence of nature are all questioned, yet Tennyson preserves his faith in human dignity and his passionate joy in landscape and the seasons; he doubts not that "through the ages one increasing purpose runs" and he looks forward to "one far-off divine event." Tennyson, although he had begun (according to an early critic) by "out-babying Wordsworth and out-glittering Keats," was no namby-pamby aesthete or escapist; as A. C. Bradley rightly said, "with the partial exception of Shelley, Tennyson is the only one of our great poets whose attitude toward the sciences of Nature was what a modern poet's ought to be." He kept abreast of scientific thought and flinched from none of the implications of the nebular theory, of geology, of embryology, and of the evolutionary theory (which he versified several years before the *Origin of Species* came out in 1859). And though he was aghast at some of these implications,

A warmth within the breast would melt
The freezing reason's colder part,
And like a man in wrath the heart
Stood up and answer'd 'I have felt.'

It was characteristic of Tennyson that he immediately added to this the words "No, like a child in doubt and fear," for which he was highly praised by the devout agnostic Henry Sidgwick. The heart, if it is to triumph over the head, must at least be humble, and Sidgwick considered that Tennyson had given beautiful and moving expression to "the indestructible and inalienable minimum of faith which humanity cannot give up because it is necessary to life." Tennyson's reputation has long emerged from the temporary eclipse it suffered and even those who may not be interested in him as a representative Victorian must admire the *curiosa felicitas* of his renderings of nature, and the skill with which his descriptions are made to symbolize states of feeling.

2. The Aftermath of the French and Industrial Revolutions.—Tennyson has been discussed at some length because he is the most obvious link between the romantic and Victorian ages and because of his representative position. In the following section it is proposed to discuss, against the background of the period, only those authors who seem, a century or so later, to have best stood the test of time.

It has been said that the 19th century, in one respect, was merely the aftermath of the French Revolution. As we have seen, the English poetry of that century took its character mainly from the Romantic age—that is to say, from a period of upheaval and excitement which, originally sympathetic with French aspirations, later on swung away into conservative reaction. In England the Radicals and Liberals made it their concern to salvage from the tidal wave of reaction whatever of permanent value the French Revolution had bequeathed. Although the working-out of this process belongs to political and social rather than literary history, it cannot be ignored by anyone who wishes to discover the mainsprings of the literature. The enduring social injustice and inequality, the extremes of wealth and poverty, the persistence of class privilege, the illiteracy of the masses—the sense of all these and many other evils lay heavily on the Victorian conscience. It was against all this that the Revolution had originally fought, and, though in its 20-year struggle with Napoleon England had been forced, under Burke's guidance, into a reaffirmation of its immemorial way of life, class structure and all, yet "the condition-of-England-question" was bound to assert itself with ever-increasing urgency. For the 19th century witnessed not only the aftermaths of the French and the romantic revolutions but also of the Industrial Revolution. Thus one major theme and preoccupation of Victorian writers, especially the "prophets" but of some of the novelists too, is indignation at the state of the underprivileged and the sense that something must be done before it is too late. This appears in writers as various as Bentham, Carlyle, Ruskin, Dickens, Disraeli, Mrs. Gaskell, Charles Kingsley, Matthew Arnold, William Morris and many others. Naturally these spiritual physicians do not all diagnose the same complaint or prescribe the same cure. Bentham urges reform of the legal code to ensure the triumph of utilitarianism—the "greatest-happiness principle"; Carlyle urges more faith, more sincerity and more leadership; Ruskin and Morris, more love of nature and of art and more pride in craftsmanship; Kingsley, Christian Socialism; Arnold, the spread of culture among the barbarians and philistines and the education of the populace; while the novelists proper are content on the whole to satirize abuses (work-houses and prisons, legal corruption, the squalor of the industrial slums, the inhumanity of *laissez-faire*, etc.) or to bring them home unforgettably to the readers' imagination.

Considering what great changes were going on throughout the 19th century—changes in the pattern of living, working and thinking—and what even greater changes were being prepared far below the surface, it is extraordinary with what tenacity the old social hierarchy persisted in England and how fully its underlying assumptions and valuations were taken for granted by the imaginative writers. Novelists, poets and critics all presuppose a world resting on the foundation of a working class and a "servant" class, and it goes without question that they themselves are above that level and enjoy privileges and liberties not shared by the masses. Even in the wild and visionary world of Emily Brontë there is a "Thrushcross Grange" as well as a "Wuthering Heights" and the story moves between those poles. In Dickens the Dedlocks may be on the way out but the Rouncewells, the Merdles and the Chereybles have replaced them and Society—however little Dickens may like it—is still in the ascendant. The world of Victorian literature is still largely feudal, though it reflects the uneasiness which in the real world was blighting the old order. The Victorian age was uncomfortably aware that England, in the process of becoming the workshop of the world, of acquiring steamships, railways and spinning jennies and of accumulating capital and doubling its population, had lost much that had given an ideal character to past times and particularly to the middle ages. For the very reason that the age of faith and chivalry had gone for ever, people read Scott's novels with avidity and built themselves (if and when they could afford it) palaces of art and lordly mansions in bogus Gothic or Renaissance styles.

3. The Cult of the Medieval.—Looking round for dominant threads in the carpet-pattern, one notices the cult of the medieval as significant. It appears in the poets—in Tennyson, in the Rossettis and other Pre-Raphaelites, in Carlyle, Ruskin and Morris; in the architecture of Pugin, Barry, Waterhouse and Gilbert Scott. Carlyle's *Past and Present* (1843) contrasts (unfavourably) the present—disfigured by materialism, irreligion, utilitarianism, dilettantism and the "Cash nexus"—with the past, typified by the life of the monks of St. Edmundsbury, ordered in faith and reverence under the wise despotism of their abbot. Ruskin proclaimed the gospel of Gothic to an age of machinery, declaring it to be superior because it expressed not only the faith of the builders, but also the craftsmen's joy in nature and in the exercise of their own skills. Morris dreamed of a London small and white and clean, and of a Nowhere where men might perhaps recapture the lost glamour of living.

4. Carlyle.—The predominance of Thomas Carlyle on the first half of the century, however, has an even wider and deeper meaning. His prophetic mission was to assert eternal Providence, and to recall a doubting and godless generation to faith in God: not the God of church or creed, but the God of the Old Testament and of Calvin. Speaking as one having authority, he bade men not discard the living truth along with the outworn "Hebrew old-clothes." In his *French Revolution* (finally published in 1837, but written three or four years previously) he dramatized history as the working-out of judgments upon sin and corruption; and in *Sartor Resartus* (1833–34), *Past and Present* and *On Heroes, Hero-Worship and the Heroic in History* (1841) he proclaimed nature to be the garment of the living God and summoned his readers to hear "the voice of old Eternities, far-sounding through thy heart of hearts" and to reverence those great men who had heard that voice more clearly than others. Carlyle's influence dwindled in his later years as people grew tired of his hysterical style and his excessive admiration for strong men. But he had inspired and invigorated a whole generation, and his prose at its best had been moving and expressive—the incarnation of a highly individual vision.

5. Bentham and Coleridge.—Though Carlyle subscribed to no particular religious creed he made continual use of religious language, and the whole trend of his message associated him with the believing side in the 19th-century struggle. J. S. Mill once called Bentham and Coleridge the two seminal minds of the century and said that most men of the time had learned how to think from one or the other. Bentham is the father of the philosophical radicals, the Utilitarians and the freethinkers—men like James and John Mill, Herbert Spencer, G. H. Lewes, Leslie Stephen, John Morley, Frederic Harrison—of all, that is, who believed that the religion of the past had had its day and that the self-interest of man could be schooled in the direction of orderly progress by education, free discussion and the enlightened application of utilitarian standards. Bentham asked of any belief or institution "Is it true? Is it useful?" Coleridge, on the other hand, asked "What does it mean? How did it arise?" In his later

prose works, *Aids to Reflection* (1825) and *The Constitution of Church and State* (1830) he showed that old meanings in religion and politics might still be valid though outward forms had changed, and claimed that the spiritual realities which must be spiritually discerned are invulnerable to all attacks by the mere "understanding," the mind of the flesh. The influence of Coleridge's philosophical rehabilitation of Christianity can be seen all down the century in writers like John Sterling, Julius Hare, Thomas and Matthew Arnold, and F. D. Maurice; Carlyle conspicuously carried on his work, notably in exalting German idealism over the native empirical tradition; and even J. S. Mill, who belonged to the opposite camp, freely acknowledged that Coleridge had deepened and enriched his understanding of the forces that hold men together in society.

6. Carlyle and Newman.—J. A. Froude, the disciple and biographer of Carlyle, placed Thomas Carlyle and John Henry Newman where Mill had placed Bentham and Coleridge, at the head of the two currents of thought that interested him most. These two prophets may now both seem men of God, though Newman looks to Rome and Carlyle to Goethe and the starry heavens. To Froude, naturally, they seemed to be facing in opposite directions. Yet they both belong to the Coleridgean stream and Newman himself counted Coleridge, together with Wordsworth and Scott, among the influences making for spiritual reawakening. This is no place for an account of the Tractarian movement or of Newman's secession from the Church of England, central though these were in the spiritual life of the age. It is surprising how little they are reflected in the literature of the time; one can point only to Froude's *Nemesis of Faith* (1849) and *The Oxford Counter-Reformation* (1881), Charlotte M. Yonge's *The Heir of Redclyffe* (1853), J. H. Shorthouse's *John Inglesant* (1880), a few passages in Mark Pattison's *Memoirs* (1885) and the opening of Matthew Arnold's *American Discourse* on Emerson (1885). But in the *Apologia* (1864) Cardinal Newman produced what is not merely a masterpiece of prose style (prose "of the centre," not eccentric like Carlyle's) but one of the two or three most significant books of the century. (*See* OXFORD MOVEMENT.)

7. Darwinism.—In any short list of such "significant" books one would have to include Charles Darwin's *Origin of Species* (1859). It may be said that this is not "literature" but science and indeed that is true if the meaning of literature is confined to that which proposes for its immediate object pleasure rather than truth. Works of science, however—and of philosophy and history—did still, in the 19th century, often possess literary value and the *Origin* was perhaps the last major scientific work written to be understood and enjoyed by the "general reader." Darwin did not invent the theory of evolution; his importance is that he demonstrated how it worked. He familiarized his age with the concepts of natural selection in the struggle for existence and of continuous development by the operation of constant natural laws. Powerfully aided by his disciple and champion T. H. Huxley he changed the traditional world-picture and compelled men to readjust all their cherished assumptions about God, nature and man. After Darwin there comes a gradual ebbing of the old religious certainties—a God who creates, designs and sustains and loves; man made in His image; a hereafter of rewards or punishments—and a growing suspicion that we are caught up in the workings of impersonal forces moving inscrutably to an unknown end, without regard for human feelings and hopes. All this, of course, did not come about immediately; at first, indeed, it seemed that Darwinism fitted in well with Victorian optimism about the inevitability of "progress." But, both in life and in literature, the old unthinking exuberance and confidence give place in the latter half of the century, to "doubts, disputes, distractions, fears." If Dickens' *The Pickwick Papers* (1836–37) and Thackeray's *Vanity Fair* (1848) may be taken to represent the first half of the century, George Eliot's *Middlemarch* (1871–72) and Thomas Hardy's *Tess of the D'Urbervilles* (1891) or *Jude the Obscure* (1895) may stand for the second half.

8. The Novel.—The Victorian age produced hordes of eminent persons and many truly great ones, but in pure literature its achievements reached the highest level mainly in the novel. A good deal of its poetry has worn rather thin, and some reputations —notably that of Robert Browning, so highly regarded up to the turn of the century as a subtle psychologist, deep thinker and invigorating lay-preacher—have been somewhat eclipsed. In drama its contribution was inferior; in criticism, with the conspicuous exception of Matthew Arnold, it was uncertain of its direction. But in Charles Dickens, Emily Brontë, George Eliot, Thomas Hardy and Henry James it produced novelists of the first order: while it produced also a large company of others who, though sometimes in their own way unsurpassed, belong mainly to a rank below the very highest, *e.g.*, Anthony Trollope, Charlotte Brontë, Wilkie Collins, Lord Lytton, William Hale White ("Mark Rutherford"), R. L. Stevenson, Mrs. Humphry Ward, George Moore and George Gissing, besides others already mentioned. It used to be customary to class Thackeray with Dickens as the two best Victorian novelists, just as Tennyson and Browning were regarded as the two best poets. Later, Dickens was decried as melodramatic, sentimental and chaotic, and Thackeray was preferred as the more civilized and self-conscious artist. Similarly, Browning for a while gained ground for his intellectual toughness, his rugged speech-rhythms and his abandonment of the garland and singing-robes. But, seen from a modern vantage-point, Tennyson overtops Browning, and Dickens begins once again—though perhaps for new reasons (because his popularity in his lifetime was a thing apart, never reached before and never to be repeated)—to appear the supreme Victorian novelist. He began in *Pickwick* as the arch-entertainer in the comic-picaresque style of the 18th century; he ended with the group of far more subtle, intricate and sombre novels on which his reputation now mainly rests: *Bleak House* (1852–53), *Little Dorrit* (1855–57), *Great Expectations* (1860–61) and *Our Mutual Friend* (1864–65). All the while he was *en rapport* in a unique way with his public, sometimes altering his plots to suit what he knew they wanted or to avoid breaking their hearts—for he knew that his characters were real people to his readers, as they were also to himself. Every sophisticated reader now knows all about Dickens' defects; he draws caricatures rather than portraits; his characters are stage types who repeat catch phrases *ad nauseam;* his plots are fantastic and contrived, not growing organically out of character and situation; his sense of artistic form is crude; he knows only his own social milieu—the urban lower-middle class; his scale of moral values is unenlightened; he takes sides for and against his own creations —idolizing them, weeping over them or lecturing them, to the reader's embarrassment. And this is true; yet he is big enough to triumph over it all. For sheer vitality, inventiveness and imaginative force combined with great powers of observation and a flair for significant detail, he is without an equal among English novelists. Already he is being praised as a symbolist, though an intuitive rather than a conscious one. He felt his way imaginatively into scenes and situations, and by charging them with feeling made them expressive emanations of the Dickens vision. Moreover his linguistic virtuosity is astonishing; no man has ever invented so many quite distinct dialects of the English language.

It has been said that with "George Eliot" the English novel came of age. And certainly there is in the work of this distinguished woman, Marian Evans, an intellectually adult quality which distinguishes her from all her predecessors and contemporaries. She was in every sense an "intellectual," familiar with all the advanced thought of her time—scientific, philosophical and theological; a reader of Comte and Darwin, and the first translator of D. F. Strauss's *Leben Jesu* (1835), a work which pained the orthodox by its relentless "demythologizing" of the Gospels. If the *Westminster* wanted a review of the latest work on sociology or the higher biblical criticism, it was to Miss Evans that they turned. Brought up in the evangelical tradition, and experiencing in her youth a strong urge toward saintliness and the service of mankind, she had been shocked out of orthodoxy by an early encounter with "emancipated" thinking and became an agnostic. In this she was typical of her age, in which so many men repeated a similar pattern of development; and she was no less typical in retaining, in spite of her loss of simple faith, a firm hold upon

the ethical core of religion and a lifelong sense of mission to mankind. She also retained—and this was what made her a novelist—an intense affection for the well-remembered scenes and people of her Warwickshire childhood and an imaginative sympathy with their beliefs. All these qualities, combined with a powerful grasp of psychological and sociological realities, enabled her to produce a series of admirable novels, including *Adam Bede* (1859), *The Mill on the Floss* (1860) and, most notably, *Middlemarch* (1871–72), now generally felt to be the greatest single English novel of the period.

It is natural to think next of Thomas Hardy, since both he and George Eliot were at their best in dealing with rustic life, scenes and character. But Hardy's world is both narrower and wider than hers; narrower in that he is at home in a less complex rural society, but wider in that it includes an imaginative dimension lacking in hers. Hardy, unlike George Eliot, was essentially a poet; it is from the heart rather than the head that his best work comes. Hardy's real subject-matter is states of feeling, rendered against a landscape-background coloured by these feelings. Since Hardy was of a desponding temper the colouring tends to be sombre, and his themes tragic. Hardy's sense of the strange disease of modern life, of the withdrawal of God and the indifference of nature, expressed itself in the vision of man caught up in the web woven by an inscrutable destiny—a vision implicit in all his work but expressly stated in his remarkable epic drama *The Dynasts* (1910). Yet he was as keenly alert to every aspect of nature's goings-on and depicted them with as much imaginative faithfulness and intensity as if he had retained Wordsworth's trust in nature's beneficence. There is nothing in English prose to equal his rendering of the night-sky or the thunderstorm in *Far From the Madding Crowd* (1874), Egdon Heath in *The Return of the Native* (1878) and winter at Flintcomb Ash or early mornings at Talbothays in *Tess of the D'Urbervilles* (1891)—to mention only a few of his triumphs of this kind. Faults of style he may have: his prose, like his verse, is sometimes cumbrous and gnarled. But the brooding intensity of his mind, which stamps his individuality on everything he touched, raises Hardy to a very high place among the novelists of the later 19th century, a position hardly challenged—as we see it now—by his far more brilliant, witty and sophisticated rival George Meredith; and challenged only by Henry James, who carried the art of the novel to the utmost limits of technical, psychological and ethical subtlety, but whose *oeuvre* falls mostly outside this century and is discussed in *The 20th Century*, below. The same qualities of intensity and authenticity give Hardy as poet (and verse was always his own preferred mode of expression) a rank among the later poets of the century shared perhaps by Gerard Manley Hopkins and not attained even by Algernon Charles Swinburne who was once considered the leading poet of the time, but who, despite all his sound and fury, is now generally felt to signify very little.

9. Critics.—It would be generally agreed that of all the eminent Victorians Matthew Arnold has worn best, and in saying this one is thinking of his prose rather than his poetry, and in particular (in the present context) of his literary criticism. Only Leslie Stephen and Walter Pater can be thought of as approaching a similar order of importance, with critics like Walter Bagehot, R. H. Hutton, W. E. Henley, R. L. Stevenson, J. Addington Symonds, Edmund Gosse and others well in the rear.

Leslie Stephen is, in virtue of his whole output, a very considerable figure. In his great *History of English Thought in the XVIIIth Century* (1876) he set a high standard in what was then a new venture but is now an accepted type of study: the "history of ideas"; and throughout the numerous essays collected in *Studies of a Biographer* (1899–1902) and *Hours in a Library* (1874, 1876, 1879), in his many biographies in the *Dictionary of National Biography* (which he edited), and in the group of lives he contributed to the "English Men of Letters" series, which he inaugurated with the biography of Samuel Johnson in 1878, he showed unusual knowledge and skill in relating writers to their historical, social and intellectual backgrounds and a sure instinct in discerning what one may call their moral core. Stephen is also interesting as one of that band of honest doubters—including T. H. Huxley, W. K. Clifford, Henry Sidgwick and others—whose highmindedness and good lives did so much to make agnosticism respectable. He published *An Agnostic's Apology* in 1893.

Walter Pater, the high-priest of aestheticism, was of a quite opposite cast of mind. In him evaluation is subordinate to "appreciation." When he writes, say, of Wordsworth, Coleridge, Charles Lamb or Sir Thomas Browne, his aim is not to indicate their moral stature or assign them their place in any scheme of values, but to divine their uniqueness, disengage and distil it and then present it "flask'd and fine" as the quintessence of each writer: "the property each has of affecting one with a special, a unique, impression of pleasure." Often Pater does this supremely well, yet one is quickly sated with the criticism that treats all subjects alike as if they were the "strange dyes, strange colours, and curious odours" to which, as Pater teaches, we ought to be so alert during every counted minute of our lives. It was Matthew Arnold who gave criticism a firm basis of principle and a sense of direction; and a great deal of the best subsequent criticism has taken its cue from him.

Arnold's special distinction was that he combined a poet's sensibility with a civilized intelligence and a fine flair for diagnosing the cultural and spiritual condition of the modern world. He lifted criticism out of its rut of provinciality and brought it into line with contemporary European thought. He was a true son of Thomas Arnold of Rugby and had both the scholar's keenness to learn, and the educator's zeal to propagate, the best that had been said and thought in the world. As a sage and humanist he perceived the malaise of his times; as a poet and man of feeling he was well-acquainted with the *Zeitgeist* and had felt its spiritual east winds. Thus, for example, in an age when romanticism was in its heyday and nobody asked whither it was tending, Arnold drew attention to the defects of much of the admired romantic poetry—its lack of sanity, substance and form—and pointed instead to the classic virtues of sobriety, weightiness and unity. Moreover, ever alert to the signs of the times, he perceived that with the weakening of old religious beliefs and the advent of a new, undisciplined democracy literature might have to play a far more important part than ever before in maintaining civilized standards. He therefore tried to furnish his readers with criteria or "touchstones" whereby to distinguish the good from the bad, the very best from the inferior or counterfeit. The marks of supreme excellence, he taught, were high seriousness, truth of substance and superiority of accent and movement (the grand manner), the power of moral interpretation (concern with "the question how to live") and the power of "natural magic" (the gift of awakening in us a sense of the beauty and mystery of things).

Arnold interpreted "criticism" very widely, extending it (as in the essay "The Function of Criticism at the Present Time"; in *Essays in Criticism*, first series, 1865) to include the whole work of an informed intelligence playing freely upon all stock notions, habits and institutions—in fact, equating it with his other concept of "Culture," which was to perform the like functions. In *Culture and Anarchy* (1869) Arnold turned his attention in this way to society and proposed Culture, in particular liberal education for the middle-classes, as the best remedy for the rawness, smugness and narrowness of the "Philistines." In place of "the dissidence of Dissent," he proposed "Sweetness and Light"; and he wanted to balance "Hebraism," a way of life limited by rigid codes and dogmas, with "Hellenism," or the kind of intelligence that tries to see things as they really are.

It is also characteristic of Arnold that in spite of his intellectual emancipation, he should recognize in religion the highest form of culture. He felt, however, that although men could not do without Christianity, they could not do with it as it was; and so he set himself—in a series of remarkable works (*St. Paul and Protestantism*, 1870, *Literature and Dogma*, 1873, *God and the Bible*, 1875 and *Last Essays on Church and Religion*, 1877)—to find for religion a sure foothold in spiritual experience and need, at the same time separating and winnowing away all that he felt to be inessential accretion or *aberglaube*. Opinions differ about the permanent value of these books, but not about the sincerity of Arnold's effort to restore life and conviction to religion at a time when it was

threatened on the one side by a too easy infidelity and on the other by a too easy acceptance of sacred clichés.

10. Conclusion.—Space forbids discussion or even mention of many other writers of importance and interest. It was not the intention here to touch upon every name, and readers are also recommended to consult the separate biographies. It may be said in conclusion that in the last two decades of the 19th century there were unmistakable signs that the original impulses which launched the century on its course—romantic, evangelical, humanitarian, melioristic—were losing momentum; poets were becoming aesthetes instead of legislators of mankind, and novelists were turning more and more inward for their subject-matter. In Oscar Wilde there is the iridescence of decay; in Samuel Butler the overthrow of many Victorian sanctities: the home, parental and filial affection, the church; in Rudyard Kipling the bugle notes of the new imperialism; in George Bernard Shaw the conversion of drama into an instrument of propaganda and discussion. The 19th century, however, lingered precariously on till World War I swept away its Edwardian aftermath. (B. Wy.)

VIII. THE 20TH CENTURY

The turn of the century was a period of complex change, and combinations of social forces and tendencies which were not at first apparent have left their mark on the literature of the time. Rudyard Kipling's verse, for example, expresses pride in the expansion of Britain overseas and an attitude to subject races which was at once generous and condescending; but it also, frequently, looks back in admiration to the mid-Victorian virtues of industriousness, determination and technical skill, upon which that expansion had been based, and registers with recurrent disquiet a sense that these sterling qualities were being replaced by a plutocratic age in which the financier stood for more than the industrialist or pioneer. Kipling was no vulgar imperialist, as is sometimes thought, and his writing reflected the anxieties as well as the buoyancy of his period.

1. Early 20th-Century Novelists.—Other writers, especially novelists, also showed a sense of the less attractive side of overseas expansion. Joseph Conrad's most outstanding novel, *Nostromo* (1904), resumes Kipling's theme, in that it records how finance imperialism (American in this case) operates in a Central American country to replace the colourful barbarism of the past with nothing better than a soulless and sterile materialism. Conrad is vastly superior to Kipling in this respect on account of his comprehensiveness and detail and more still because of his large and considered sense of human integrity and standards of conduct: qualities which he brought from his years as a sea captain and writer of sea-stories to the broader field of his later novels of contemporary society. Yet the deepest note which runs through his fiction is of a society which, for all its range and power, is losing its values and often prefers the elaborate sham to genuine human worth.

If this note of frustration, almost of despair over human ideals, is Conrad's response to the plutocratic society of his time, it is significant that Henry James, in a series of novels which are the outstanding literary product of the turn of the century, shows much the same sense. The world he depicts is the world of *fin-de-siècle* and Edwardian plutocracy itself (the great country manor, the great town house); and repeatedly, the progress of his novel reveals, beneath the elegant surface of contemporary life, something of its sordid reality: greedy connoisseurship in *The Spoils of Poynton* (1897), calculating promiscuity in *What Maisie Knew* (1897) and *The Awkward Age* (1899), materialism and selfishness in *The Wings of the Dove* (1902) and *The Golden Bowl* (1904).

Like Conrad, James brought to the world in which he lived and which he took for his scene standards which were both stricter and richer than what he found in it. In his case, these largely came from what was best in the culture of New England, and the impact of Europe on the American visitor was a recurring theme in his work. One value, however, which does not show explicitly in Conrad's work (though it does implicitly) is conspicuous in James's: it is a conviction of the importance in life of art itself taken in its highest sense. This does not merely reflect James's affiliation with Walter Pater, Oscar Wilde and the poets of the 1890s and the Yellow Book in which they published (an affiliation in itself real enough). It is a reminder of something that lies behind his whole effort as a writer. In his deep and discriminating study of the great French novelists of the 19th century, in his attention to a range of other continental writers (particularly Turgenev), and also in his work as a critic and his habit of intent and sustained reflection on the whole task and art of the novelist, James marks a departure from the general character of English literature in the mid-19th century.

At that time, English literature proceeded largely along its own lines; but by 1900 a steady tide of influence was beginning to flow from the mainland of Europe. The desire of James and Conrad to excel specifically as artists (a desire which makes the work of both, at times, leave a sense of over-calculation, of material which has been much elaborated but in essence is thin) is a sign of the influence of a self-conscious and sophisticated writer like Flaubert. All in all, indeed, the influence of the continent was a pervasive quality of development at the time and may be clearly traced in the novels of George Moore and Ford Madox Ford (notably his *The Good Soldier*, 1915) and in the early writings of Somerset Maugham and John Galsworthy. In particular, Arnold Bennett's novels plainly reveal the impact on English writing of such French "realist" authors as Émile Zola.

Interest in continental writing did not lead to mere slavish imitation. James studied Flaubert and learned from him, but condemned him on balance. Arnold Bennett studied Zola while resolving to avoid what he saw as Zola's faults. Nor, of course, was there a break with tradition and continuity in English itself. James salutes George Eliot as the great master of English fiction and Bennett acknowledges how much he owes to Herbert Spencer; their deference to the broad and humane novelist and the comparatively shallow sociologist respectively, is in fact a measure of both the nature and the merit of their own work. But Bennett's achievement was a genuine one. His fiction usually sacrifices a large and humane sense of life for realism of scenic detail and the study of impersonal social forces; but to depict these things is not without value; and it was Bennett, more than anyone else writing fiction in the early 1900s, who revealed the other side of Edwardian opulence, the drab and deprived life of the lower classes whose emergence over the next 50 years was to have such an effect on writing itself.

Other developments in English writing in the early years of the 20th century also show a complex interaction of indigenous forces and influence from the mainland of Europe. The vogue for Ibsen, which came to dominate the English theatre in the 1890s, led to the early plays of George Bernard Shaw; but although these works, like a number of Ibsen's, took contemporary social abuses as their subjects, they abandoned Ibsen's own concise and realistic speech for witty and flamboyant diction such as may be traced back in the English (or more particularly Irish) theatre as far as Congreve. The early writings of James Joyce show the influence of his years in Paris, and link him, somewhat unexpectedly, with admirers of contemporary French writing like the poets of the 1890s. Like that of Joyce, D. H. Lawrence's main work belongs to a somewhat later period, and his intense individuality is always conspicuous in it; but admiration for recent Russian fiction is not only overtly expressed in his letters; it is also visible in the humane sensitiveness and fluid organization of his own work. Finally, it must be remembered that this is the period when, through the translations of Constance Garnett and others, a large body of Russian fiction was becoming available in English for the first time.

2. The Early Development of "Modern" Poetry.—How "modern" poetry emerged in the early years of the 20th century is not wholly clear. Certainly it will not do to think simply of a swift and drastic substitution, by Ezra Pound and T. S. Eliot during the 1910s, of realism and intelligence for escapist romanticism. The "Decadent" poets of the '90s were both more isolated in England, and more cosmopolitan in their interests, than is usually realized; and Pound's earliest affiliations were in part to them. Throughout these years there was also a clear tradition of verse

(going back at least as far as W. E. Henley's *Hospital Verses* of the early 1870s), that dealt with everyday life and that used everyday language. Thomas Hardy, Edward Thomas and the early Lawrence were more perceptive, thought more deeply and had greater power than some of their contemporaries represented in the "Georgian" anthologies (1912–22) which, started by Rupert Brooke, John Drinkwater, Harold Monro, Wilfred Wilson Gibson and Edward (later Sir Edward) Marsh, contained poems by Brooke, W. H. Davies, Drinkwater, John Masefield, Walter de la Mare (whose poetic gifts were perhaps as great as any, but who stood outside the main stream of "modern" poetic development) and Lawrence himself, as well as by many others; but they did not stand in contrast to them.

Indeed, their work itself hints at weaknesses decisive elsewhere: both Hardy and Thomas were inspired by nature and the countryside, but it is a nostalgic inspiration of something receding into the past. In this they were responding to the facts of English social change; and these facts meant that nature and the countryside could no longer be a source, to young poets, of deep and vital inspiration. The writer's contact tended more and more to be with nature seen superficially and with the townsman's or the weekender's eyes (the same trend is clear, during these years, in the fiction of H. G. Wells). As such, the work of these poets easily became, in the hands of others, diluted and middle-brow; against just this the avant-garde authors of the period reacted.

W. B. Yeats, before 1900, saw something of this "middle-brow" quality in much 19th-century poetry, and declared that true poetry could be written for scholars or for peasants, but for none in between. In so saying, however, he was responding to the indigenous peasant life of his own country on the one hand, and (once again) to France as a source of literary ideals on the other. He turned from English Victorian poetry to the difficult, uncompromising and esoteric work of the French Symbolists: though their work was probably best known to him through Arthur Symons' *Symbolist Movement in Literature* (1899). During the 1900s Yeats himself, both in his poetry and in plays written then for the Abbey theatre in Dublin, was breaking with literary mannerisms such as he owed to the Pre-Raphaelites. He was not, however, turning to the "ordinary life" or "ordinary speech" of the middle class, but to a diction he found among the peasantry or the landed ascendancy of Ireland and to an austere ideal of poetic drama which stood in distinguished remoteness from the everyday.

Pound and Eliot, in their early years in England from about 1910 onward, constituted a sophisticated minority movement which looked to the Symbolists (particularly Jules Laforgue) as models; and their poetry must be seen as a deliberate reaction against conventional work accepted by a large "middle-brow" audience. Moreover (again taking its direction from French poets like Laforgue), it turned emphatically away from nature and the countryside, to represent the attitude and interests of a metropolitan intellectual élite.

3. The Impact of World War I and Its Aftermath.—Even as early as Eliot's *Waste Land* (1922), however, something of a fundamentally different kind was to be observed. What this poem attempted lay quite outside the interests of the Symbolists. It sought to depict the plight or indeed breakdown of a whole civilization. The fundamental change of direction must be seen in the context of the war and of the sharp break between early war poets like Julian Grenfell and Rupert Brooke, and later ones like Siegfried Sassoon, Wilfred Owen and Isaac Rosenberg. The earlier war poets inherited their attitude from literature about wars in the past. What dominates the work of the later poets is a deeply tragic sense of the fearful sufferings which the war brought, as in Owen; or (as in Sassoon) angry satire at those who endorsed it and prolonged it.

To notice that Owen was a much more insular poet than Pound or Eliot (who were, after all, expatriates of the United States) is not to deny his great talent, less as it undoubtedly was than theirs. But for a comprehension of the general development of literature at the time, what is of importance is that in their poems after 1920 (Eliot's *The Waste Land* and Pound's first *Cantos*, which were written then), these poets show much the same sense as his that mankind had come to a time of crisis and that the urgent call on the poet was to confront it. The poet's utterance rises from social satire almost to prophecy. Yeats also moved in the same direction at the same time: as is plain in poems such as "The Second Coming" or "Meditations in Time of Civil War." This last title, however, is a reminder that in his case the change was prompted less by the war itself than by the so-called "Troubles" in Ireland of which the war was in fact the occasion.

The war had a decisive effect on fiction as well as on verse, though its impact shows itself in rather different ways. The early years of the 20th century saw a resurgence of Liberalism, which enjoyed a great victory in the election of 1906, and which in the following years took decisive steps toward the Britain of social welfare and constant material advance which, to a large extent, became the pattern for the future. This movement was not political in any narrow sense: it was reflected in the whole intellectual life of the great houses of leading aristocratic families (James's novels reflect something of this), and in intellectual groups like the "Souls" of Oxford; and it has a certain affinity even with more radical groups like those of the Fabian society (*q.v.*), with which Beatrice and Sidney Webb and George Bernard Shaw were associated. In the sphere of culture, the word "liberal" during these years justly points to a whole movement of emancipation, humanity and concern for the higher and finer values of living.

This intellectual liberalism is clearly reflected in the novels of E. M. Forster and Virginia Woolf (though "Bloomsbury," with which their names are so often associated, was no isolated phenomenon, but part of a much larger whole; *see* BLOOMSBURY GROUP). The values of these writers were the values of the liberal intellectual in the best sense, yet their work also leaves an impression—clear even in a work from the years of mounting tension before World War I, such as Forster's *Where Angels Fear to Tread* (1905)—that there was a weakness in their position which might make it untenable in the postwar world. In Forster, as in Virginia Woolf's first postwar novel, *Jacob's Room* (1922), a concern for freedom, individuality and personal development goes with a clear and deep fear that these things have not the strength to survive the violent modern world which now surrounds them. Even in the best work of these authors some may discern a lack of vitality, of conviction, which it is fair in the last analysis to see as a 19th-century idealism (Virginia Woolf was the daughter of Leslie Stephen) conscious of how it may go down to the new century in which it finds itself.

The novels of Aldous Huxley showed a more advanced condition of the same anxiety. Their portrait of liberal intellectual culture is in a sense more brilliant than that of Forster or Virginia Woolf; but it reads like the work of one who has lost faith in what he depicts, and is preoccupied with its destruction everywhere at the hands of the violent and irrational. "Since the war we wonder at nothing" (*Crome Yellow*, 1921). The sometimes almost hysterical satire of Huxley's later novels (*e.g., Eyeless in Gaza*, 1936) led naturally to the cynical attitude to liberal culture of the novels of Evelyn Waugh, in the context of which the feeling conveyed by the novels of Graham Greene (*q.v.*), that (except for the Roman Catholic Church) life is drab, futile and horrifying, is also rightly seen. The highly stylized novels of Ivy Compton-Burnett, with their picture of upper-class life surviving as a witty mask covering only the seedy and needy, reflect the underlying social change in another way; and the same may be said of the earlier novels of Wyndham Lewis, with their disgusted sense of the triviality underlying much cultivated or "intellectual" life. Seen as a whole, however, Lewis' work registered an awareness of the coming of a machine age and a megalopolitan society such as has no parallel in modern fiction, and is a notable positive achievement.

James Joyce.—The two most remarkable novelists of the period were Joyce and D. H. Lawrence. The work of Joyce can be understood only as a complex interaction between changing literary influences, and a lifelong preoccupation with everyday life in his native city of Dublin. Joyce's first affiliations, as his early and somewhat emotional prose style makes plain, were with the '90s; and something of the conviction that art is and should be a self-contained world of its own remains through all his work. Some of

his later and in a sense drabber realism he may have owed to Ibsen, or to French realist fiction such as that of Zola. But his own unique style seems to have emerged from his contact with Pound and Eliot, or rather perhaps (during his years in Paris) with the French writers who influenced them. In his later prose (*Ulysses*, 1922; and *Finnegan's Wake*, 1939) Joyce shows the reaction from easy-to-follow middle-brow writing to a style in which art makes no compromise with the general reader. Ideas now come one after another in unadorned juxtaposition, a word which is a key to the kind of sequence aimed at in Pound's and Eliot's verse. The effect is to present the reader with what seems like the illogical thinking of a character who is bringing his thought only half into conscious focus. Developing from this (which has affinities with the "stream of consciousness" technique of Dorothy Richardson and Virginia Woolf) Joyce creates, through an extraordinary tissue of puns, words fused into one another, and verbal coinages (often made possible through amalgamating standard and Irish pronunciations) a dense texture of associations and cross-reference in thought which purports to represent the full richness and subtlety of the mind at work. Yet *Ulysses* is no empty piece of virtuosity. Joyce's constant experimentation with language was for a purpose. It created a picture of everyday life in Dublin which for its detail, comprehensiveness and realism has few equals in literature anywhere, and which is also outstanding for the humour and humanity which plays throughout it.

D. H. Lawrence.—If Lawrence's work as novelist, poet and critic is considered as a whole, it is plain that no writer except perhaps Yeats can challenge him as the outstanding literary genius of the earlier 20th century. Coming from an environment in the north midlands where rural life was fast giving place to modern industrialism, it is not surprising that he was influenced by other writers, like Thomas Hardy and Richard Jefferies, with somewhat similar backgrounds. But from the start, Lawrence was a sophisticated and avant-garde writer. He was acquainted with the modern realist trend (though reacting vigorously against it); he admired the Russian novelists; and he was in touch with a great deal of contemporary development in continental fiction. His earlier novels frequently rehearse just his own development in this direction: the central character, each time, outgrows his or her provincial origins and acquires a modern, urbanized, even cosmopolitan, consciousness (this is clearest in *The Rainbow*, 1915).

As his own consciousness of modernity extended, however, Lawrence seems to have become more and more aware of its dangers, and to have set himself more and more urgently to find a solution to them through his fiction. The central reality of his novels became his conviction that men were being deprived of traditional sources of strength, balance and genuineness, and could replace these only by finding the vital source of life-giving spontaneity which they had in their own innermost being. Lawrence's concern with sexual life is subordinate to this: for him, men engaged in sexuality were (or could be) most genuinely themselves, and least puppets of a machine age. It was natural enough that, with these convictions, Lawrence should be led toward new technical interests in writing: and the fluid, episodic constructions of his later novels (notably *Women in Love*, 1920), with their looser, even seemingly disjointed characters which yet have an intense reality and individuality, reflect contemporary trends in literature (largely foreign) put to substantial use in his work.

But, as he recognized himself, Lawrence was a devoted seeker after the deepest values of life, rather than an assured master of them. His successes came out of the intensity of his effort, and his work often displays the stridency, even bullying insistence of the neo-moralist, which when he was truest to himself he rejected. He is not to be admired without reserve. Yet it is easy, and venial, to forget this, in the face of his astonishing vitality, his often sharply perceptive intelligence, his unrivaled sense of life's potentialities for good, and his unremitting stand for what was real in it against what was sham.

W. B. Yeats.—It is illuminating to see Yeats beside Lawrence, for he too was driven from adherence to a local and traditional culture by a sense of the threat of modernity. His earliest work was the product of Celtic legend, Irish peasant life and the aesthetic movement. But more and more, as the early years of the new century went by, he was drawn into the modern world: first, by the part he played in the Irish Dramatic movement, then by the Easter rebellion of 1916 and "Troubles" before and after 1920, more generally by the wider consciousness to which these things led him, the awareness that in the modern world, the values which he most cared for both as artist and as at heart an aristocrat were being threatened by (in his own phrase) "new commonness" on every hand.

Troubled more and more deeply after 1920 by his sense of living in a hostile world, Yeats turned again, though in new ways, to the occultism and esoteric philosophies of history which had interested him (as occultism had interested several of the Symbolists in France) when he was a young man; and it was out of such material as this that he developed his later views, expressed in *A Vision* (1925). But Yeats's elaborate cyclic philosophy of history served an end beyond itself. It was the rationale that Yeats erected for his intuitive sense of spontaneity and life-affirming energy arising from bare existence. His later poetry may be seen as an assertion that human vitality was in itself an answer to what was de-humanizing in modern life. It was an assertion which looked back to the Irish ascendancy of the 18th century, and it was different from Lawrence's in being more hieratic, stylized and rhetorical; but it had almost equal force and conviction.

Other Writers.—Many other authors of the period between 1920 and 1950 also responded in somewhat similar ways to what they sensed as a period of crisis in human affairs. L. H. Myers showed, especially in *The Root and the Flower* (1935), much of Lawrence's concern for what is real and what is sham in human life (though he lacked Lawrence's gift for fiction). Joyce Cary deals recurrently with the disintegrating impact of war and modernity; but re-asserts the power of human emotion, and especially of social and family affection, to overcome it. George Orwell (*q.v.*) was notable as a satirical and political writer (*The Road to Wigan Pier*, 1937; *Animal Farm*, 1945; *1984*, 1949). The left wing poets of the 1930s—W. H. Auden, Stephen Spender and Louis Macneice—represent another though shallower current in the same movement.

4. The Changing Scene After 1930.—To turn to these authors, however, is to discover something fundamentally new in the literary scene; for their work shows that, broadly speaking, the avant-garde movement represented by Joyce, Pound, Eliot, Virginia Woolf and Lawrence was dominant for a short time only. Cary was a predominantly traditional novelist, writing in the chronicle form and owing much to that most English of novelists, Dickens. The leftwing poets of the '30s may have retained something of the terse colloquialism and satire of Pound or Eliot, but they more or less abandoned the whole Symbolist side of their work. The main influences on new verse seemed now to be Hardy, and Gerard Manley Hopkins, whose poems were written between 1860 and the late 1880s, but were not published until Robert Bridges edited them in 1918. Hopkins' obscurity and compression, however, and his bold and varied experiments with language, had less influence than his rhythmic fluency and rich descriptiveness. The general effect was that, after 1930, both fiction and verse began to make a significant break with the immediate past, and to regain much more continuity with trends in English writing throughout the 19th century. The period of strongest influence from the continent was over. Almost imperceptibly, new writing came to be dominated by something much more popular in quality, and read as if written for an audience both larger and less cosmopolitan than that of the avant-garde.

Ultimately, what lay behind these literary changes may well have been certain basic developments in the whole social structure of Britain. It must be remembered that Pound and Eliot (like, indeed, Yeats and the poets of the '90s before them) were a small élite group writing against what they saw as an effete literary "Establishment." The same ends were being served by the most significant critical organs of the time, such as A. R. Orage's *The New Age*, which ran from 1908 and published much of Pound's early criticism, T. S. Eliot's *Criterion* (from 1921), and Edgell Rickword's *Calendar of Modern Letters*, which ran in the 1920s and published some of Lawrence's major critical pieces. The

whole literary movement of the first 30 years of the century must be seen as comprising not only a body of original work, but also a comprehensive set of principles about writing expressed by such critics as T. E. Hulme, and a new evaluation of English literature based upon these principles. All in all, it was a major and distinctive phase in English literary history.

In the course of time, however, these minority views tended to become those of the majority and to be changed in the process. In part, this may have been due to the enormous expansion of higher education (especially in English literature itself as a university subject) and in consequence to the growth of a sizable reading public of a rather new kind. Behind this educational change, however, must be seen the basic social change which underlay it and brought it about: the advance of the lower middle- and working-classes to positions of much greater relative prominence and influence than ever before in the life of the country. Literature in the years after 1930 tended more and more to be written, and read, by those with a new outlook and interests: both —to speak in very broad terms—more earnest and more straightforward than what had gone before. The reaction against the "ivory tower" of the aesthetic movement, led by Pound and Eliot, became generally accepted. That literature should meet the demands of alert intelligence and common sense and not of relaxed or wild emotionalism also became an accepted view, as did the conception of both fiction and other prose as interesting mainly for the light they shed on moral values and on how to live. These changes, introduced by the writers and critics already mentioned, were powerfully advanced by the critical writings of F. R. Leavis and the magazine *Scrutiny* which he edited from 1932 to 1953. Perhaps it was partly because the changes were already under way, that World War II had no effect on writing in any way comparable with that of World War I. The pervasive literary transformation was already taking place and it was doing so as reflection of a social transformation. But in becoming widely accepted, the newer views and interests lost something of their distinctiveness. Few poets, for example, after Eliot and Pound wrote with anything like their compression and obscurity, or their demands on the reader by way of literary reference. The political poets of the 1930s were lucid, even diluted writers by comparison; William Empson and Dylan Thomas were the only important poets of that decade (and they were contrasting figures) to write difficult verse.

5. Poetry After 1945.—Empson and Thomas, however, represent the recession rather than the continuance of a tradition. Empson was influential on many later poets during the 1950s, but they adapted his laconic wit and nonchalant tone rather than his cramped intense feeling and his immense intellectual ingenuity; and many who seem to have learned something from him (John Wain, Philip Larkin, Donald Davie and D. J. Enright are examples) have drawn as much or more from a writer such as Robert Graves, largely independent of modernism, or from W. H. Auden.

Dylan Thomas influenced many later poets, but they followed his rhapsody and rhetoric rather than the exact craftsmanship and semi-mystical sense of the unity of nature which underlay his rare but outstanding successes as a poet, and their work constitutes a minor current. All in all, by 1960 shrewd common sense, carefully calculated tone, disciplined craftsmanship and sharp intelligence had become dominant in new English verse, even in the case of those like Thom Gunn and Ted Hughes in whom the underlying emotional response seemed strongest. Another significant influence, that of the U.S. poet Wallace Stevens, pointed the same way: in writers like Charles Tomlinson and Elizabeth Jennings whom he influenced, the flamboyance and the obscure philosophizing disappeared, and painterly detail, dry exactness, or limpid rhythm were what remained.

6. Postwar Fiction.—Something similar was true of fiction. Several new novelists, such as Kingsley Amis and John Wain, emerged in the 1950s. Their work commanded considerable interest, and seemed to be the fiction most distinctive of that period. But they wrote in conscious revolt against the avant-garde ambitions of novelists of a previous generation. An underlying intention to provide serious comment on the values of ordinary contemporary life went in their case with hostility to anything that smacked of the "high-brow" and "phoney." In their techniques of narration and characterization they looked back to the 19th-century traditions of English fiction. Strangely, the work of other and in a sense more ambitious novelists showed a similar trend. Lawrence Durrell claimed that his fiction was based on the scientific principle of relativity; but the reader senses little in the books to bear this out. Iris Murdoch, herself a lecturer in philosophy, showed a strong dependence on French Existentialist thought in her first novel *Under the Net* (1954). This seemed as if it might be the herald of a new period of closer contact with developments on the continent; but the impression that she was a cosmopolitan writer was less strong in her later work.

7. Drama.—Drama moved generally in the same direction as verse and fiction. Following his years of contact with the Abbey theatre (*q.v.*) Yeats continued from 1910 until the end of his life to write plays, mainly in verse, for small and sophisticated audiences. In the 1930s Eliot also turned to verse drama (*Murder in the Cathedral, The Family Reunion*), though from the start he wrote with larger audiences and a more general public in mind; and his lead was followed by other writers including Auden and C. Isherwood. After World War II, a significant change set in. Eliot's ambition became, more and more emphatically, to re-introduce verse drama into the popular theatre. To do this he aimed at a diction as near as possible to prose, and sought to adapt the commercial theatre's conventions of locale, staging and construction to his own deeper purposes. His position was thus already remote from that which lay behind his earlier verse; and his later plays (*e.g., The Cocktail Party, The Confidential Clerk*) are better seen as a distinctive kind of work for the commercial theatre, rather than as verse drama in a more ambitious or poetic sense. If so, it is clear that he too, in his work after 1950, illustrated the general movement away from the avant-garde ideas of 1920, and toward a style more traditional in kind and general in appeal.

The most striking theatrical development of the later 1950s, however, was in the direction not of verse drama but prose. A number of new young dramatists, among them John Osborne, Arnold Wesker, Doris Lessing, Bernard Kops and also Harold Pinter, whose work shows traces of the very different influence of Eugène Ionesco, produced in rapid succession a large number of plays which emphatically rejected the witty language, middle-class interests and neat structure of the drawing-room comedy; and depicted, loosely and flexibly, the everyday affairs of the common people of Britain as they seemed after 1945. Some of these plays did little more than transfer some of the methods of a talented if conventional writer like Terence Rattigan to a new social *milieu*. The best of them, however, dealt with their material with a straightforward (if humorous) earnestness and a concern for what can give life significance in contemporary Britain, which gave their work much interest and at the same time made it part of a general trend in writing at the beginning of the 1960s.

Certain writers stand outside this general picture. The strange fiction of the Irish Samuel Beckett, with its deep compassion nightmarish anxiety and sometimes bewildering stylistic experiments, must be seen (like his plays, *e.g., Waiting for Godot*) as a reflection of how this follower of Joyce lived for many years in Paris, and owes more to French influences, and a continental sense of the modern world, than to English. William Golding's use of lyricism and allegory in fiction (especially in his *Lord of the Flies*) and his sense of the importance of the unconscious, give him a distinctive place (though his later work perhaps does so less than his earlier). Hugh Macdiarmid, outstanding figure of the renaissance of Scottish writing in the 1920s and 1930s, continued, in his strange long poem, *In Memoriam James Joyce* (1955), the cosmopolitan—indeed in his case, polyglot—interests of an earlier period. Edwin Muir, until his death (1958), produced a body of work which was memorable in its awareness of all the horror of the atomic age and at the same time of possibilities of innocence and regeneration which look back to Blake and Wordsworth. In his case, too, a less insular kind of awareness was the product of residence abroad and the influence of foreign literature (Muir devoted himself, over a number of years, to translating Kafka). But these exceptions important as they are, do not obscure the general picture of liter-

ary movement in the years between 1940 and 1960; and the fact that it ran in a direction fundamentally different from that of the earlier part of the century. *See* also references under "English Literature" in the Index volume. (J. Hv.)

BIBLIOGRAPHY.—*General Works:* A. W. Ward and A. R. Waller (ed.), *Cambridge History of English Literature*, 15 vol. (1907–27), which although superseded by much later research, remains a standard work; A. C. Baugh (ed.), *A Literary History of England* (1948); T. H. Ward (ed.), *The English Poets: Selections With Critical Introductions*, 5 vol., 2nd ed. (1918); H. Craik (ed.), *English Prose: Selections With Critical Introductions*, 5 vol. (1893–96); F. W. Bateson (ed.), *The Cambridge Bibliography of English Literature A.D. 600–1900*, 4 vol. (1940), *Supplement*, ed. by G. Watson (1957); *Hogarth Lectures on Literature*, ed. by G. Rylands and L. Woolf (1927); D. Daiches, *Literature and Society* (1938); B. Dobrée (ed.), *Introductions to English Literature*, 5 vol. (1938–40); G. Sampson, *The Concise Cambridge History of English Literature* (1941); F. P. Wilson and B. Dobrée, *The Oxford History of English Literature*, 14 vol. (1945 *et seq.*); B. Ford (ed.), *A Guide to English Literature*, Pelican Books (1954–); D. Daiches, *A Critical History of English Literature*, 2 vol. (1960); T. S. Eliot, *The Use of Poetry and the Use of Criticism* (1933); F. W. Bateson, *English Poetry and the English Language* (1934), *English Poetry: a Critical Introduction* (1950); Sir H. J. C. Grierson and J. C. Smith, *A Critical History of English Poetry*, rev. ed. (1947); D. A. Davie, *Articulate Energy: an Enquiry Into the Syntax of English Poetry* (1955); J. W. H. Atkins, *English Literary Criticism* [to 1800], 3 vol. (1943–51). *See* also bibliographies to general articles on literary kinds BIOGRAPHY; DRAMA; NOVEL; etc.

Periodicals. The Year's Work in English Studies (1919–1921); Modern Humanities Research Association, *Annual Bibliography of English Language and Literature* (1920–1921); English Association, *Essays and Studies* (1916–); *The Modern Language Review* (1905–); *The Review of English Studies* (1925–); *Studies in Philology* (1906–); *PMLA* (Publications of the Modern Language Association of America) (1886 *et seq.*); *Essays in Criticism* (1951 *et seq.*); *Scrutiny* (1932–53); *ELH.: a Journal of English Literary History* (1934–). (X.)

Anglo-Saxon Period: R. Wülker, *Grundriss zur Geschichte der angelsächsischen Litteratur* (1885); Alois Brandl, *Altenglische Litteratur* and *Mittelenglische Litteratur* in *Grundriss der germanischen Philologie*, ed. by H. Paul, 3rd ed. (1911); E. Wardale, *Chapters on Old English Literature* (1935); R. H. Hodgkin, *A History of the Anglo-Saxons*, 2 vol., rev. 2nd ed. (1956); C. W. Kennedy, *The Earliest English Poetry* (1943); F. M. Stenton, *Anglo-Saxon England*, 2nd ed. (1947); George K. Anderson, *The Literature of the Anglo-Saxons* (1949).

Early Middle English Period: M. W. Schofield, *English Literature From the Norman Conquest to Chaucer* (1906); J. E. Wells, *A Manual of the Writings in Middle English 1050 to 1400* (1916, with supplements to 1951); W. P. Ker, *English Literature: Mediaeval* (1912); J. J. Jusserand, *A Literary History of the English People*, vol. i, 3rd ed. (1925); R. W. Chambers, *On the Continuity of English Prose* (1932); W. L. Renwick and H. Orton (eds.), *The Beginnings of English Literature to Skelton*, 2nd ed. (1952); R. M. Wilson, *Early Middle English Literature* (1939); R. S. Loomis (ed.), *Arthurian Literature in the Middle Ages* (1959). (C. L. W.)

Chaucer to the Renaissance: J. E. Wells, *A Manual of the Writings in Middle English 1050–1400* (1916, with supplements to 1951); C. Brown and R. H. Robbins, *Index of Middle English Verse* (1943); W. L. Renwick and H. Orton (eds.), *The Beginnings of English Literature to Skelton*, 2nd ed. (1952); D. M. Zesmer and S. B. Greenfield, *Guide to English Literature From Beowulf Through Chaucer and Medieval Drama* (1961); *Oxford History of English Literature*, vol. ii, E. K. Chambers, *English Literature at the Close of the Middle Ages* (1945), vol. iii, H. S. Bennett, *Chaucer and the 15th Century* (1947), and vol. iv, C. S. Lewis, *English Literature in the 16th Century Excluding Drama* (1954); W. P. Ker, *Medieval English Literature* (1912); G. Kane, *Middle English Literature* (1951); D. Everett, *Essays on Middle English Literature* (1955); C. S. Lewis, *The Allegory of Love* (1936); R. Weiss, *Humanism in England During the 15th Century*, 2nd ed. (1957); M. J. C. Hodgart, *The Ballad* (1950); R. S. Loomis (ed.), *Arthurian Literature in the Middle Ages* (1959); E. K. Chambers, *The Medieval Stage*, 3 vol. (1903); K. Young, *The Drama of the Medieval Church* (1933); H. Craig, *English Religious Drama of the Middle Ages* (1955); G. Wickham, *Early English Stages, 1300–1660*, vol. i (1959); R. W. Chambers, *On the Continuity of English Prose* (1932); K. Sisam, *Fourteenth-Century Verse and Prose* (1921); A. W. Pollard, *Fifteenth-Century Prose and Verse* (1903); E. P. Hammond, *English Verse Between Chaucer and Surrey* (1928); H. C. Sargent and G. L. Kittredge, *English and Scottish Popular Ballads* (1904); R. L. Greene, *The Early English Carols* (1935), *A Selection of English Carols* (1962); C. Brown, *Religious Lyrics of the 15th Century* (1939); R. H. Robbins, *Secular Lyrics of the 14th and 15th Centuries*, 2nd ed. (1955), *Historical Poems of the 14th and 15th Centuries* (1959). *See* also bibliographies to ARTHURIAN LEGEND; BALLAD. (N. D.)

Renaissance Period, c. 1510–1660, Poetry and Prose: Sidney Lee and C. T. Onions (eds.), *Shakespeare's England*, 2 vol. (1916); C. S. Baldwin, *Renaissance Literary Theory and Practice* (1939); D. Bush, *The Renaissance and English Humanism* (1939); E. M. W. Tillyard, *The Elizabethan World Picture* (1943); F. P. Wilson, *Elizabethan and Jacobean* (1945); V. de Sola Pinto, *The English Renaissance, 1510–1688* (1951); C. Ing, *Elizabethan Lyrics: a Study of the Development of English Metres* (1951); Hallett Smith, *Elizabethan Poetry* (1952); C. S. Lewis, *The Oxford History of English Literature*, vol. iii, *English Literature in the 16th Century Excluding Drama* (1954); J. Winny (ed.), *The Frame of Order* (1957); J. R. Brown and B. Harris (ed.), *Elizabethan Poetry* (1960); J. Bronowski and B. Mazlish, *The Western Intellectual Tradition*, especially part 1 (1960).

Elizabethan and Jacobean Drama: H. W. Wells, *Elizabethan and Jacobean Playwrights* (1939); A. P. Rossiter, *English Drama From Early Times to the Elizabethans* (1950); L. C. Knights, *Drama and Society in the Age of Jonson* (1951); V. Ellis-Fermor, *The Jacobean Drama*, 3rd ed. (1953); M. Doran, *Endeavors of Art* (1957); I. Ribner, *The English History Play in the Age of Shakespeare* (1957); J. R. Brown and B. Harris (eds.), *Jacobean Theatre* (1960). (P. Dw.)

Early 17th-Century: Prose and Poetry: E. N. S. Thompson, *The 17th-Century English Essay* (1926); H. J. C. Grierson, *Cross-Currents in English Literature of the 17th Century* (1929); D. A. Stauffer, *English Biography Before 1700* (1930); W. F. Mitchell, *English Pulpit Oratory From Andrewes to Tillotson* (1932); B. Willey, *The 17th Century Background* (1934); Joan Bennett, *Four Metaphysical Poets* (1934; rev. ed., 1953); H. C. White, *The Metaphysical Poets: a Study in Religious Experience* (1936); W. Haller, *The Rise of Puritanism* (1938); E. M. W. Tillyard, *The Elizabethan World Picture* (1943); D. Bush, *Oxford History of English Literature*, vol. v, *English Literature in the Earlier 17th Century* (1945); F. P. Wilson, *Elizabethan and Jacobean* (1945); B. Boyce, *The Theophrastan Character in England to 1642* (1947); R. Tuve, *Elizabethan and Metaphysical Imagery* (1947); V. Harris, *All Coherence Gone* (1949); C. V. Wedgwood, *17th-Century English Literature* (1950); T. S. Eliot, *Selected Essays*, 3rd ed. (1951); G. Williamson, *The Senecan Amble* (1951); H. C. Baker, *The Wars of Truth* (1952); L. L. Martz, *The Poetry of Meditation* (1954); M. H. Nicolson, *Science and Imagination* (1956). (D. BH.)

Restoration Period: A. Bryant, *King Charles II* (1931); Gilbert Burnet, *History of My Own Time* (with supplements), ed. by O. Airey and H. C. Foxcroft, 3 vol. (1897–1902); E. T. Campagnac, *The Cambridge Platonists* (1901); G. N. Clark, *The Seventeenth Century* (1929); R. S. Crane *et al.*, *English Literature 1660–1800, a Bibliography of Modern Studies*, from 1926 onward (1950); B. Dobrée, *Restoration Comedy* (1924), *Restoration Tragedy* (1929); C. H. Firth, *Essays, Historical and Literary* (1938); T. H. Fujimura, *The Restoration Comedy of Wit* (1952); D. Nichol Smith (ed.), *Characters of the Seventeenth Century* (1918); A. Nicholl, *A History of English Drama, 1600–1700*, vol. i, *Restoration Drama* (1952); V. de S. Pinto, *English Biography in the 17th Century* (1951), *Restoration Carnival* (1954); L. E. Spingarn, *Critical Essays of the 17th Century*, 3 vol. (1908–09); T. Sprat, *History of the Royal Society*, ed. by J. I. Cope and H. W. Whitmore (1958); D. A. Stauffer, *English Biography Before 1700* (1930); Dale Underwood, *Etherege and the 17th-Century Comedy of Manners* (1959); B. Willey, *The 17th-Century Background* (1934); J. H. Wilson, *The Court Wits of the Restoration* (1948). (V. DE S. P.)

The 18th Century: W. E. H. Lecky, *A History of England in the 18th Century*, 7 vol., 4th ed. (1892); Sir Leslie Stephen, *English Literature and Society in the Eighteenth Century* (1904); S. Johnson, *Lives of the English Poets*, 3 vol., ed. by G. Birkbeck Hill (1905); G. Saintsbury, *The English Novel* (1913); O. Doughty, *English Lyric in the Age of Reason* (1922); M. D. George, *English Social Life in the 18th Century*, 2 vol. (1923); G. M. Trevelyan, *England Under Queen Anne*, 3 vol. (1930–34); S. Morison, *The English Newspaper* (1932); A. S. Turberville (ed.), *Johnson's England* (1933); A. O. Lovejoy, *The Great Chain of Being* (1936); H. N. Fairchild, *Religious Trends in English Poetry*, vol. i and ii (1939–42); B. Willey, *The 18th Century Background* (1940); J. R. Sutherland, *A Preface to 18th-Century Poetry* (1948); D. A. Davie, *Purity of Diction in English Verse* (1952); C. A. Moore, *Backgrounds of English Literature, 1700–60* (1953); A. R. Humphreys, *The Augustan World* (1954); K. Williams, *Swift and the Age of Compromise* (1958); B. Dobrée, *Oxford History of English Literature*, vol. vii, *English Literature in the Early 18th Century, 1700–1740* (1960). (N. C.)

Romantic Period: M. H. Abrams, *The Mirror and the Lamp* (1953); M. H. Abrams (ed.), *The English Romantic Poets: Modern Essays in Criticism* (1960); I. Babbitt, *Rousseau and Romanticism* (1930); J. Barzun, *Romanticism and the Modern Ego* (1943); E. Bernbaum, *Guide Through the Romantic Movement* (1948); C. H. Herford, *The Age of Wordsworth* (1930); G. Hough, *The Romantic Poets* (1953); D. G. James, *The Romantic Comedy* (1948); F. L. Lucas, *The Decline and Fall of the Romantic Ideal* (1936); M. Praz, *The Romantic Agony*, 2nd rev. ed. (1951); C. D. Thorpe *et al.* (ed.), *The Major English Romantic Poets* (1957). (R. P. C. M.)

19th Century: O. Elton, *Survey of English Literature, 1830–80*, 2 vol. (1920); B. Ifor Evans, *English Poetry in the Later 19th Century* (1933); B. Willey, *Nineteenth Century Studies* (1949), *More Nineteenth Century Studies* (1956); G. Hough, *The Last Romantics* (1949); E. Batho and Bonamy Dobrée, *The Victorians and After, 1830–1914*, 2nd ed. (1950); J. E. Baker, *The Reinterpretation of Victorian Literature*

(1950); John Holloway, *The Victorian Sage* (1953); G. M. Young, *Portrait of an Age*, 2nd ed. (1953); W. E. Houghton, *The Victorian Frame of Mind, 1830–70* (1957); P. Appleman, W. A. Madden and M. Wolff, *1859: Entering an Age of Crisis* (1959); David Daiches, *A Critical History of English Literature*, vol. ii (1960). (B. Wy.)

20th Century: David Daiches, *The Novel in the Modern World*, rev. (1961); G. S. Fraser, *The Modern Writer and His World* (1953); Graham Hough, *Image and Experience* (1960); Frank Kermode, *Romantic Image* (1957); F. R. Leavis, *New Bearings in English Poetry*, rev. (1950); Edwin Muir, *Transition: Essays on Contemporary Literature* (1926); V. da Sola Pinto, *Crisis in English Poetry, 1880–1940* (1951); W. Y. Tindall, *Forces in Modern British Literature, 1885–1946* (1947); Edmund Wilson, *Axel's Castle* (1931).

The most comprehensive poetry anthology of the period is *The Faber Book of Modern Verse*, ed. by Michael Roberts (1951); and the social background to the literature of the 20th century is given in D. C. Marsh, *The Changing Social Structure of England and Wales, 1871–1951* (1958). (J. Hy.)

ENGRAVING, LINE, the process of incising lines in a metal plate from which it is intended to take an impression or print. It differs from other print-making processes in three ways: (1) the engraved line is the part that is printed since this is an intaglio printing process as opposed to surface printing which is employed in lithography (*see* LITHOGRAPH) or to relief printing which is employed in woodcut (*see* WOODCUT AND WOOD ENGRAVING); (2) the line is incised without the use of chemical action (*see* ETCHING); (3) the line is incised with a burin or graver.

Technique.—The surface used for engraving is the polished face of a metal plate, usually of copper because it combines strength with pliability, although other metals, such as iron, pewter, zinc or silver, may be used. The burin, or graver (fig. 1), which

FIG. 1.—THE BURIN OR GRAVER USED FOR INCISING LINES INTO THE METAL PLATE

engraves the composition on this plate is a sharp chisellike tool of quadrangular section. In setting to work the engraver places the plate on a cushion or pad so that the plate may be turned in any direction desired. In working he pushes the burin forward on the surface of the metal so that the graver removes a small shaving, leaving the edges of the furrow comparatively clean. In dry point (*q.v.*) the line is scratched by a needle drawn toward the operator, which in its passage leaves minute ridges on either side of the line, called burr. In line engraving the infinitesimal burr that is left by the graver is removed (as in fig. 2[B]) by a sharp instrument called a scraper (fig. 2[C]), whereas in dry point this same burr is left to produce the "furry" effect which is the very essence of the medium. Since clarity and precision are the aims of line engraving and this burr would contain ink that would show in printing, it is always removed. The scraper is also employed for correction in all intaglio processes. To make a correction, the surface of the plate may be cut away to the bottom of the engraved lines which are at fault; the surface is brought level again by hammering the back of the plate and polishing it with an instrument of oval section and polished edge, called the burnisher. When the design has been engraved on the plate, impressions may be taken by running it through a press as is done with an etched plate (*see* ETCHING). Changes in shading, composition, addition of lines, etc., may be made later and the plate reprinted. Each new printing with these variations constitutes a state. A plate with deeply engraved lines may yield a large number of impressions without deterioration (as many, perhaps, as 2,000), while one of delicate lines would print only a small number. In order to prolong their satisfactory use, plates are often treated by a process of electrolysis known as steel facing. It must always be remembered that an impression taken from the original plate—and not the plate itself—is known as an original. Plate mark is the indentation left in the paper by the pressure of the plate in printing.

A counterproof is an impression taken not from the plate itself but from another impression which has not dried: a maculature is an impression taken after the plate has been printed in order to clean it and remove ink that has remained in the engraved lines. (*See* also AQUATINT; GLASS PICTURES AND PRINTS; MEZZOTINT; MONOTYPES; PHOTOENGRAVING; STIPPLE AND CRAYON ENGRAVING.)

Origins of Line Engraving.—There is little doubt that engraving was first practised north of the Alps, but at what place or exact time must remain a matter of conjecture. That it was evolved in the workshops of goldsmiths, later than the more obvi-

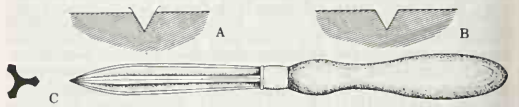


FIG. 2.—ENLARGEMENT SHOWING (A) BURR LEFT BY THE GRAVER; (B) ENGRAVED LINE AFTER BURR HAS BEEN REMOVED BY A SCRAPER; (C) SCRAPER (ALSO SHOWN IN CROSS SECTION)

ous process of woodcut, is generally admitted; a date between 1410 and 1430 and a place somewhere between the Alps and the North sea, in the Rhine basin, may be regarded as probable. The earliest engravings known have generally been regarded to be those of an artist called, from the subject of his most striking work, the Master of the Playing Cards. The place where this engraver, who was almost certainly a goldsmith, worked was probably south Germany or Switzerland. Engraving, in his hands, was already an elaborate technique, presupposing a development of some years. A Flagellation of a Passion series, in the Berlin print room, bears the date 1446, the earliest date on any intaglio print; bestowing on its author the name of Master of the Year 1446. Contemporary with him, or even earlier, is the engraver, working in the Netherlands or Burgundy, known as the Master of the Death of Mary. His ideas of distance and perspective were less advanced than those of his German contemporary, but this may have been an individual weakness not implying an earlier date.

With the master who signed with the initials E. S., some of whose works are dated 1466 and 1467, engraving in Germany reached a further stage of development. The number of his surviving engraved subjects, over 300, indicates that he was primarily an engraver by profession, and the influence of his style and technique on subsequent engraving north of the Alps and even in Italy was great. Martin Schongauer of Colmar (d. 1491) derives from the Master E. S. and may have been his pupil. From E. S.'s still rather tentative technique Schongauer developed a regular system of crosshatching admirably suited to his purposes, and the precision and delicacy of his technique, combined with the distinction and real power of his design, go to make his engravings among the beautiful achievements of the 15th century. The only engraver of the period who is at all comparable to him is the anonymous artist known both as the Housebook Master (from a volume of drawings of daily and domestic usages for the owner of a medieval castle, apparently by the same hand) and Master of the Amsterdam Cabinet (since the largest number of his prints is owned by the print room of the Rijksmuseum in Amsterdam). His highly personal style strongly influenced younger engravers, for example Albrecht Dürer. His technical methods were also original: using a softer metal, perhaps lead or pewter, instead of copper, he drew on it with a dry point needle, utilizing the resulting burr in printing so that his work has more of the character of dry point than of engraving.

The last quarter of the 15th century in Germany was prolific in engravers. Schongauer's followers and imitators were numerous. The course of engraving in the Netherlands after the middle of the century ran more or less parallel with that in Germany, though there were no artists of the importance of E. S. and Schongauer.

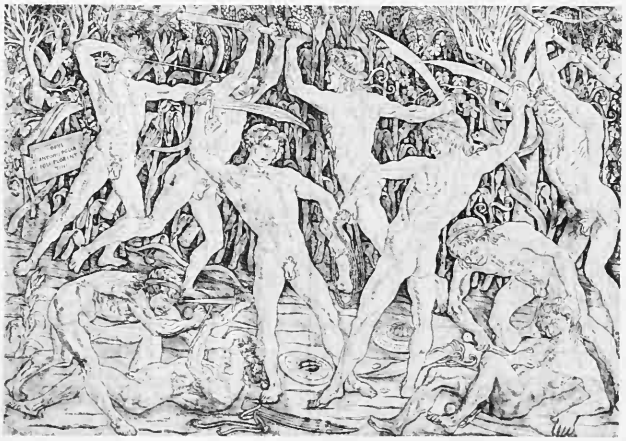
Albrecht Dürer of Nuremberg (1471–1528), although not an actual follower of Schongauer, carried on the Gothic tradition and later developed it, in contact with the Italian Renaissance and with his own fantastic and thoughtful temperament, into that grand and significant style that may be seen in such engravings as the "Knight, Death and the Devil," the "Melancolia" and the "St. Jerome." Dürer was more than an engraver, but it was in this medium, perhaps, that his genius found the most perfect expression; he must be considered, in view of this absolute fitness of means to expression, as one of the greatest, if not the greatest, of the world's engravers.

EUROPEAN ENGRAVINGS OF THE 15TH AND 16TH CENTURIES

Top left: "Saint John the Evangelist on the Isle of Patmos" by the Master E. S. (active *c.* 1440–67), German. *Top centre:* "Madonna and Child" by Andrea Mantegna (1431–1506), Italian. *Top right:* "The Holy Family" by Lucas Van Leyden (1494–1533), Dutch

Right: "Battle of the Naked Men" by Antonio Pollaiuolo (1429–98), Florentine

Bottom left: "The Virgin Mourning Over The Body of Christ" by Marcantonio Raimondi (*c.* 1480–*c.* 1530), Italian. *Bottom centre:* "Saint John the Baptist" by Giulio Campagnola (*c.* 1482–1514), Italian. *Bottom right:* "Saint Sebastian, Saint Anthony and Saint Roch" by Jean Duvet (1485–*c.* 1561), French



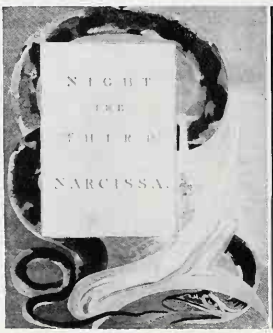
BY COURTESY OF THE ART INSTITUTE OF CHICAGO, FROM THE COLLECTIONS OF (TOP LEFT, TOP RIGHT, CENTRE) MR. AND MRS. POTTER PALMER, (BOTTOM LEFT, BOTTOM RIGHT) CLARENCE BUCKINGHAM, (BOTTOM CENTRE) CHARLES DEERING; (TOP CENTRE) GIVEN IN MEMORY OF CHARLES NETCHER II

ENGRAVING, LINE

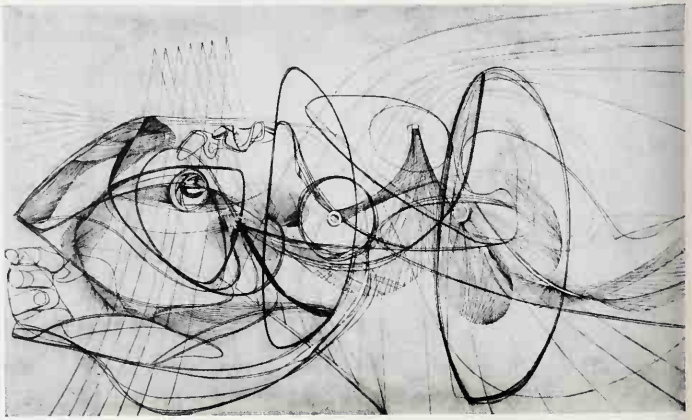
Left: "Sir Thomas Farfax" (Fairfax) by William Faithorne the elder (c. 1616–91), English. After Robert Walker. *Right:* "Nicolas Potier de Novion" by Robert Nanteuil (1623?–78), French

Right: "Les Petits Parains" (Parrains) by Jean-Charles Baquoy (1721–77), French. After J. M. Moreau the younger

Illustration for *The Monikins:* Dr. Reasono. Engraved by Duthie and De Mare from F. O. C. Darley (1822–88), U.S.



Left: Illustration for Young's *Night Thoughts:* Night the Third—Narcissa by William Blake (1757–1827), English. *Right:* Illustration for Shakespeare's *King Richard the Third*, Act IV, Scene III, by W. Skelton (1763–1848), English. After J. Northcote



"Death by Water" by Stanley W. Hayter (1901–), U.S.

BY COURTESY OF (ALL EXCEPT CENTRE LEFT AND CENTRE RIGHT) THE ART INSTITUTE OF CHICAGO AND AS FOLLOWS: (TOP LEFT) THE CLARENCE BUCKINGHAM COLLECTION, (TOP CENTRE) GIFT OF MARY MORLEY SELLERS, (TOP RIGHT) PRINT AND DRAWING PURCHASE FUND, (CENTRE) GIFT OF FREDERIKA SKINNER, (BOTTOM) GIFT OF THE PRINT AND DRAWING CLUB

Lucas van Leyden (1494–1533), in Holland, was in his time placed on a level with Dürer, but posterity will hardly endorse this view, though his earlier engravings, beginning with "Mahomet and the Monk Sergius" (1508), in his 15th year, had a naïve directness and lucidity which were admirable. His engraving further marked an increase in the delicacy of tone values, which even Dürer had not attained.

Vasari's statement that the art of engraving was invented by Maso Finiguerra about the year 1460 has long since been disproved. Not only was engraving known and practised north of the Alps some decades earlier, but even in Italy itself examples dating from at least ten years before Finiguerra's alleged discovery have been found. The earliest Italian engravings, which emanated most probably from the workshops of Florentine goldsmiths, were in marked contrast to contemporary German productions, showing an entirely different—if not much lower—level of technical skill, both in the actual engraving and printing. A first small group of Florentine engravings, about 1450, was succeeded by another group which may, with great probability, be associated with the name of Maso Finiguerra (1426–64). There is every reason to credit the tradition that he was an engraver, and the style of this second group is closely related to that of works in intarsia authenticated as his, and to a large group of works in niello, an art of which he was reputed to be the greatest master. Clearly connected with and probably executed in Finiguerra's workshop after his death are a number of other engravings such as the series of the "Prophets and Sibyls." These, with the engravings actually attributed to Finiguerra, are usually grouped together as Florentine engravings in the fine manner as opposed to a more important group distinguished as in the broad manner. The broad manner group, which are mostly later in date than the fine manner prints, some being copies from the latter, aim at reproducing the effect of pen drawing and derive their inspiration in part from Botticelli.

The first great Florentine engraver actually to use the graver himself was Antonio Pollaiuolo; his only engraving, "The Battle of the Nudes," is stylistically to be included in the group of prints in the broad manner. In Andrea Mantegna (1431–1506), Italy produced its first really great engraver. The "Virgin and Child," probably one of his earlier works, the great "Entombment" and the "Battle of the Sea Gods" were as noble expressions of the great painter's genius as any of his pictures. A whole group of engravers was inspired by Mantegna and worked after his designs. No greater contrast could be found with Mantegna's style than that of Jacopo de'Barbari (d. after 1512), a Venetian, who assimilated the northern technique and used it to express a charming personal quality.

16th-Century Engraving.—There were many engravers working in the north of Italy at the beginning of the 16th century. Benedetto Montagna of Vicenza (1470–1540) continued the style of his more important father, the painter Bartolommeo. In Venice the gifted Giulio Campagnola (about 1482–1514) executed work in a curious dot manner akin to the later stipple that approached in engraving the effects Giorgione was aiming at in painting.

Marcantonio Raimondi (*c.* 1480–*c.* 1530), trained in the workshop of the goldsmith-painter Francia, in Bologna, at first engraved solely after drawings by or in the style of his master. Marcantonio improved his powers and his technique by studying and by copying Dürer and Lucas van Leyden. After a stay in Florence and Venice he went to Rome about 1510. There he came in contact with Raphael, who supplied the engraver with sketches from which to work. It is this contact which has given to Marcantonio his great reputation. Marcantonio's most famous engraving, "The Massacre of the Innocents," can be judged not only on its own merits but in relationship to Raphael's drawings, and it stands the test triumphantly.

The influence exercised by Raphael, especially through the engravings made by Marcantonio and his followers, over European art south and north of the Alps can hardly be exaggerated, though, in Germany, the victory of the Raphaelesque was not so overwhelming or so sudden. The so-called Little Masters—the two Behams, Barthel and Hans Sebald, and Georg Pencz of Nuremberg, together with Jakob Binck and Heinrich Aldegrever—continued the tradition of original engraving, along the lines Dürer had inaugurated, and added something to the sum of the engraver's achievement. In Ratisbon, Albrecht Altdorfer, of a previous generation, was the chief exponent of a quite individual style. However, after the year 1550, engraving, like the other arts in Germany, had passed its prime, and the second half of the century was a decline redeemed by fewer and fewer works of originality.

With the middle of the century engraving in the Netherlands became commercialized. Engravers were employed by publishing firms, like that of Hieronymus Cock at Antwerp, to reproduce works, Netherlandish and Italian alike, for which there appeared to be a demand. Even engravings by Italians were published in the Netherlands, while Netherlandish engravers at the same time worked in Italy, so that the sharp lines formerly delimiting the national styles were blurred. The tendency was for the technique of Italian engraving, which from the first owed something to northern models, to assume an even more northern character; while the influence of Italy was a general one on all northern art and only incidentally on engraving. The output of engraving in what is now Belgium, especially in Antwerp, during the two middle decades of the century was enormous but its character uninteresting; after about 1570 the centre shifted to Holland and it was the school of Hendrik Goltzius, Jacob Matham and Jan Saenredam who continued and exaggerated Lucas van Leyden's later style.

In Italy it was Marcantonio, in his latest and least satisfactory phase, who exerted the greatest influence. By about 1560 the art was mainly represented in the persons of foreign engravers like Cornelis Cort. It was in this school that Lodovico Carracci and other Italian artists were trained.

In France the 16th century produced one outstanding artist who was a remarkable genius, although his technical achievements have always been the cause of dispute: Jean Duvet (1485–*c.* 1561). A group of engravers working at Lyons in the first half of the 16th century produced work of decorative charm, while the portrait engravers foreshadowed the great school of portrait engraving that followed in the 17th century.

17th-Century Engraving.—By the beginning of the 17th century engraving came to be regarded as a reproductive medium. A feature of the 17th century was the increasing popularity of portrait engraving, which became the most important branch of the art.

The engravers formed in the school of Rubens derived technically from Goltzius and his successors. But the engravers were, in general, dull, certainly the one fault of which Rubens could not be accused. Nevertheless, the best of these engravers, Pieter Soutman, Lucas Vorstermans the elder and Schelte van Bolswert, were admirable in their way. Some of these contributed to the famous series known as Van Dyck's *Iconography*. Eighteen of these plates were first etched by Van Dyck himself, to be finished in some examples but not improved by the regular engravers. These elaborated plates suffer in contrast to Van Dyck's masterly etchings.

It was in France, and mainly in the special branch of portrait, that engraving in the 17th century attained its greatest perfection. The 16th-century Netherlandish style of portrait engraving practised by T. de Leu, Leonard Gaultier and others survived into the beginning of the 17th century. It was the influence of painting as practised by Rubens, Van Dyck and Philippe de Champaigne which altered the conception of portrait painting, and consequently of portrait engraving. The earliest to show the new influences at their fullest extent were Jean Morin (d. 1650) and Claude Mellan (1598–1688), though the former's work was mainly etched. Mellan, in particular, founding his technique on that of the engravers of the Italian school, evolved a very distinctive and effective style from which crosshatching was entirely eliminated and which depended on the variation in the breadth of the individual parallel lines. Robert Nanteuil (1623?–78), unlike Morin and Mellan, not only engraved, but himself drew the vast majority of his portraits, which added greatly to their liveliness and lifelikeness. He modeled his faces in a series of small wedge-shaped strokes, cunningly contrived to reproduce the desired texture. Gérard Edelinck (1640–1707), though lacking Nanteuil's extraordinary sincerity, was an engraver of great ability, and his interpretation of pictures

by Lebrun and portraits by Rigaud and Largillière were typical examples of the art of the *grand siècle* at its apogee, while Antoine Masson (1636–1700) may be regarded as the third great portrait engraver of the century.

The 17th century in England produced one engraver of considerable merit: William Faithorne the elder, whose models were Claude Mellan and later Nanteuil. But line engraving in England was superseded in popularity toward the end of the century by the newly invented process of mezzotint.

18th-Century Engraving.—France became the artistic centre of Europe in the 18th century. The technique developed by Gérard Edelinck and Gérard Audran was continued and, if possible, further refined by the Drevet family, until engraving was reduced to an almost mechanical formula, perfect in its way, but without life or individuality.

Portrait engraving is not considered, however, as the most important product of the 18th century. It was in the engraving of society genre that what was newest and most typical of the century was found. A number of skilful engravers were employed by Jullienne, the friend of the painter Watteau, to engrave the master's complete works after his death. These engravers succeeded in rendering in line the shimmer and the grace of Watteau's painting, as far as this was translatable.

This style of engraving was obviously well adapted for illustration which began to have an importance it had not enjoyed since the 16th century. It was the professional engravers in the Watteau tradition who were employed not only to engrave but also to design these illustrations. In many, if not in most cases, indeed, pressure of work made it impossible for them to engrave what they had designed, and they were necessarily forced to employ others. But even in the latter case so close was the connection between the designer, who was an engraver himself, and the executant engraver that engraving ceased to be merely reproductive and again became a medium of direct and original expression. A characteristic of reproductive line engraving from the beginning of the 18th century onward (it was especially important in the case of the French illustrators and of the engravers of the Turner school) was the use of etching for the preliminary laying in of the design subsequently gone over and completed with the graver. The number of able illustrators who worked in France during the latter half of the century was very large and their excellence extremely uniform. Two artists may be singled out for special mention: J. M. Moreau the younger and Augustin de Saint-Aubin.

Italy, in the 18th century, had nothing to show in engraving at all comparable. The most original work was that produced in Venice by Marco Pitteri, in a style resembling that of Mellan. The work of Giovanni Volpato and his famous pupil, Raffaello Morghen, with all its prodigious skill, was decidedly devoid of life and character.

In England excellent work was done in engraving by Sir Robert Strange; by William Woollett, whose landscapes after Richard Wilson formed a real contribution to engraving; and by William Sharp, whose works after English portrait painters were of particular excellence.

19th-Century Engraving.—Illustration in England, though it had never attained the distinction which had marked it in France, was not brought to a sudden end as it was on the other side of the channel by the Revolution. The name that stands out at this time is that of William Blake (*q.v.*), who, trained as a regular engraver, still executed his most characteristic and distinguished work in other media. His engraved illustrations for the *Book of Job* and Young's *Night Thoughts* are wonderful in their conception and power and worthy of his genius. The popularity of Turner was reflected in the very numerous engravings made by engravers such as W. B. and G. Cooke, John Pye, E. Goodall and others. But Turner's obviously untranslatable creations were interpreted with the dullness of entire efficiency. This blight, and the beginning of the use of mechanical means such as machine ruling, lay over all the work produced in England, Europe and America in the middle of the century. Merely reproductive as its purpose had become, engraving obviously had no future in competition with the rapidly perfected photographic processes, and by the end of the century had died out almost completely.

20th-Century Engraving.—Like many other print mediums, engraving in the 20th century lost its utilitarian purpose and was practised by artists for those virtues and peculiarities which are its essential characteristics. The work of Jean Emile Laboureur (1877–), in France, possessed an elucidation and linear simplification which made it highly popular in the 1920s and 1930s. In England the genre subjects of Harry Morley (1881–1943), Stanley Anderson (1884–) and Robert Austin (1895–), executed with painstaking striving for precision, was favoured by collectors who were trying to withstand the surging advance of modern art. Eric Gill (1882–1940) lent his gifts of craftsmanship and decoration to plates for illustration. However, the most vigorous encouragement this classic medium had came from the efforts of Stanley William Hayter (1901–), who by his teaching, original work and guidance of many artists did much to stimulate the use of engraving as an independent medium and in conjunction with other print techniques. Mauricio Lasansky (1914–), also with his work and teaching, introduced the medium to a large number of American artists.

BIBLIOGRAPHY.—*Technique:* A. M. Hind, *The Processes and Schools of Engraving* (1952), an excellent handbook; W. M. Ivins, *How Prints Look* (1943), contains enlarged photographic reproductions of details; S. W. Hayter, *New Ways of Gravure* (1949), a survey, with descriptions of new techniques.

History: A. M. Hind, *A History of Engraving and Etching*, 3rd ed. (1923), the most satisfying general work, with an extensive bibliography, *Early Italian Engraving*, 7 vol. (1938–1949), the definitive work on the early Italian schools; Carl Zigrosser, *The Book of Fine Prints* (1948), embraces all print mediums; Max Lehrs, *Geschichte und kritischer Katalog des deutschen, niederlandischen und französischen Kupferstiche im XV Jahrhundert*, 9 vol. of text, 9 vol. of plates (1927), the definitive work on the 15th century. (H. Es.)

ENGROSSING, in a legal sense, means the copying of a legal document in a large hand (*en gros*), often referred to as making a fair copy. In an economic sense the word refers to rigging the market. In this economic meaning, engrossing is often used in conjunction with the kindred practices of "forestalling" and "regrating." All three terms are now chiefly matters of historical interest.

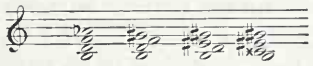
Strictly speaking, engrossing meant buying up supplies, especially crops and foodstuffs, and withholding them from the market in order to drive prices up. Forestalling meant intercepting supplies before they reached the market. Regrating designated yet another method of manipulating prices by buying in order to resell at a profit in the same market. All three practices were regarded throughout the middle ages as monopolistic and reprehensible because they tended to create artificial scarcity, and to increase prices above their natural or competitive level. The prevailing doctrine was that goods, especially foodstuffs, should be brought to the market and sold directly to consumers at prices set by supply and demand without interference from middlemen. This was the policy pursued by nearly all medieval towns, not only in England but everywhere in western Europe and the Byzantine empire. Municipal statutes and guild regulations often threatened engrossers, forestallers and regraters with the pillory or a heavy fine. This legislation can be traced back to the capitularies of Charlemagne which, in 806, censured those who hoarded commodities and refused to sell them at any but exorbitant prices. The scholastic doctors, without a single dissenting voice, regarded any form of monopoly, including engrossing, as sinful and injurious to the community. In canon law, monopoly profits were classified as *turpe lucrum*, or illicit gain, which, like usury, was subject to restitution.

The English common law throughout the middle ages considered engrossing, forestalling and regrating as practices in "restraint of freedom" which were detrimental to the commonweal and, hence, punishable by the courts. Records are full of cases relating to offenders who were either fined or made to suffer worse indignities. Edward VI's often-cited statute (1552) against engrossing, forestalling and regrating enacted nothing new, but simply restated an accepted principle of common law, defined more carefully these three practices, and regulated the penalties. The penalties ranged from two months' imprisonment and forfeiture of the goods to confiscation of all the guilty person's property on the third offense.

Although this statute was modified by subsequent enactments, it remained on the books until 1844 when it was repealed, supposedly because it interfered with the operation of free competition. In the United States also, the famous Sherman Anti-Trust act (1890) did nothing more than reassert the old common-law principle that conspiracies in restraint of trade are illegal. Perhaps it should be added that medieval monopolies were not permanent combinations but usually temporary rings which operated in small, and more or less isolated, markets. (*See* also MONOPOLY.)

See Vernon A. Mund, *Open Markets, an Essential of Free Enterprise* (1948). (R. DE R.)

ENHARMONIC, a musical term of two related meanings. It was originally applied to one of the three genera in ancient Greek music containing intervals smaller than a semitone, termed dieses (*see* GREEK MUSIC [ANCIENT]). In the 16th century Nicola Vicentino (d. 1572) proposed an enharmonic system, modeled on the Greek tradition, in which the interval of the tone was divided into five parts and for which he invented a modified notation. With the general acceptance of equal temperament for keyboard instruments in the 18th century the interval of the semitone became arbitrarily defined so that C ♯ and D ♭, for instance, were one and the same note and the term enharmonic was applied to the notation of these identical notes. This allowed a chord to be notated in several different ways, each of which is seen to belong to a different tonality. In the following example the same chord of the diminished seventh belongs respectively to the keys of C, A, F ♯ and D ♯:



See HARMONY.

ENID, the seat of Garfield county, Okla., U.S., 80 mi. N.W. of Oklahoma City. Located at a favourite watering point on the famous Chisholm cattle trail, Enid was literally transformed overnight from a prairie cattle kingdom to a thriving commercial centre. The Cherokee outlet, a vast hunting preserve under lease by the Cherokees to the Cherokee Strip Livestock association during the 1880s, was reclaimed by the U.S. government and opened to settlement by "run" on Sept. 16, 1893. At the appointed time settlers rushed into the territory and in a few hours Enid (previously chosen as the site of a government land office) was a teeming assortment of tents and shacks. These in a few weeks gave way to frame houses and business establishments, and Enid subsequently became the commercial and cultural centre of northwest Oklahoma. It was named after the Enid in Tennyson's *Idylls of the King.*

Enid's economic base consists of wheat, cattle and oil; its industries include flour and grain milling establishments, meat and dairy plants, oil refineries and oil-field equipment plants. It is the site of Phillips university, a coeducational liberal arts college, chartered in 1906 and supported by the Disciples of Christ; the university and city are joint hosts to the annual Tri-State Music festival. Vance air force base is nearby. The city has a council-manager form of government, effective in 1947. For comparative population figures *see* table in OKLAHOMA: *Population.*

(A. M. G.)

ENISEI: *see* YENISEI.

ENKIDU, a legendary hero originally appearing in Sumerian literary compositions, which were incorporated, with alterations, in the Akkadian epic of Gilgamesh. *See* GILGAMESH, EPIC OF.

ENLIGHTENMENT, a movement of thought and belief, developed from interrelated conceptions of God, reason, nature and man, to which there was wide assent in Europe during the 17th and 18th centuries. Its dominant conviction was that right reasoning could find true knowledge and could lead mankind to felicity.

The ancestral ideas of the Enlightenment reach deep into ancient Greece. There the first philosophers discovered a regularity in the processes of nature and concluded that the ordering principle was intelligent mind. From the 6th century B.C. until Stoicism set the tone for the Roman empire this assumption of a rational and moral order was an enduring guide to thought. With the fall of the Greek city-states, however, speculation had tended to shift to problems of personal salvation, and the eventual result of this turn to religious values was the victory of Christianity. (*See* PHILOSOPHY, HISTORY OF: *Ancient Philosophy.*)

Reason and Faith.—Christianity began early to develop a philosophical content. Its intellectual tools were those inherited from the wreckage of Greek life. With them Christian thinkers hammered out the rigorous system of scholasticism (*q.v.*), which Thomas Aquinas brought to a high development in the 13th century. In Thomas' presentation Aristotle, recently rediscovered by the west, provided the method for obtaining that truth which was ascertainable by reason alone; since Christian revelation contained a higher truth, Thomas placed the natural law evident to reason subordinate to, but not in conflict with, eternal law and divine law.

Imposing as the Thomistic structure was, it did not perpetuate the church's intellectual dominion. First there were deep rents torn in scholasticism by refined disputation and a sapping of the church's prestige through political controversies. Then came the Renaissance with its worldly outlook and after it the Reformation. Both were less movements for intellectual liberty than changes of authority, but, since they appealed to different authorities, they contributed heavily to the breakdown of the community of thought.

The great Dutch jurist Hugo Grotius made a pioneering attempt to fashion a new common denominator. In his *De jure belli et pacis* (1625) he invoked natural law as the principle governing international relations. For 2,000 years no one had denied the existence of natural law, but there was a fresh significance in Grotius' emphasis on its primacy and, above all, on its independence of Christian theology. This law, said Grotius, was a dictate of right reason and its basic propositions were as self-evident as $2 \times 2 = 4$.

It was not by caprice that Grotius appealed to the certainties of an equation. By his time mathematics was developing rapidly and was being used on problems for which Aristotle's answers were no longer satisfactory. The formative influence for the Enlightenment, as in the work of Grotius, was not so much content as method. The great geniuses of the 17th century confirmed and amplified the concept of a world of calculable regularity, but, more important still, they seemingly proved that rigorous mathematical reasoning offered the means, independent of God's revelation, of establishing truth. Hence men as different as René Descartes and Thomas Hobbes, Benedictus de Spinoza and Gottfried Wilhelm Leibniz wrote in the idiom of mathematical demonstration.

If there lingered more faith than certainty, Isaac Newton resolved the issue. Few could understand the mathematics by which in 1686 he stated the law of universal gravitation, but that difficulty did not becloud the great conclusion: Newton had explained the material universe. Newtonianism, then, gave a triumphant vindication to a way of thinking and a fresh incentive to reexamine all problems of human concern.

The most sensitive of those problems had always been religion. The Christian faith, shaken by a growing secularization and by embittered dissensions, had by the 17th century to face reports of peoples in the newly discovered countries beyond the seas who, like the revered men of antiquity, lived in virtue and piety although without the light of Christian knowledge. Men at odds with orthodox faith and yet unwilling to turn atheist assumed that beneath the world's diversities there was a common body of faith, a "natural religion," universally present in the human heart. In 1624, just when Grotius was preparing his case for natural law, Lord Herbert of Cherbury in his *De veritate* gave a description of natural religion which was to pass substantially unchanged into Thomas Paine's *Age of Reason* (1794). Certain beliefs, said Herbert, are so manifest that all reasonable men accept them: the existence of one God who dispenses rewards and punishments, and the obligation resting on men to worship him in repentant piety and virtue. These articles suffice for a religious life in the present and for salvation in the hereafter.

For nearly two centuries the ever-widening empire of scientific rationalism sustained the argument. Reasoning from a mathematically ordered nature to its Architect, the protagonists of natu-

ral religion, the deists, allowed the Architect only those qualities manifest in his handiwork. A widespread battle with Christianity ensued. The deists turned their weapons against the gamut of special beliefs and practices of Christianity, but they pressed most relentlessly against miracles. The marvel, contended the deists, lay in the order and the goodness of the divine creation: it was unthinkable that God would, or could, set aside laws expressive of his own being.

Deism was never an organized movement. Its first home was among English intellectuals, but widening tolerance in England took away the incitement to controversy, so that the issues steadily lost in public interest. In France, on the other hand, the rigid structure of the *ancien régime* provoked angry resentments and Voltaire, aided by numerous lesser *philosophes*, used deism as a rapier for piercing the clerical side of the body politic. The rationalist invasion was less heralded elsewhere, but it persisted during most of the 18th century. There was no census of converts, but certainly deism strongly influenced the upper strata of society all over Europe and in the American colonies.

Christianity's first methods of self-defense were the traditional ones of dogmatic assertion of its truth and legal punishment, but neither proved effective. In a rationalist age supporters of the Christian faith increasingly had to admit the validity of natural religion and to search out reasonable grounds for adding revelation. In the second half of the 17th century John Tillotson, who died archbishop of Canterbury, spoke for many when he held that Christianity was needed to complete the findings of natural religion: its service, he explained, was to test the reasoning process and to give a supplementary motive for a moral life. Tillotson conceded much, but his contemporary John Locke even more when he accepted reason as a judge of the truth of things "above Reason." A generation later, in his *Analogy* (1736), Joseph Butler, himself later a bishop, presented a long-famous restatement of the rationalist defense, but his final contention was that in natural religion and nature there were things just as difficult to believe as was Christian doctrine. (*See* also DEISM; RATIONALISM: *Theological Rationalism.*)

While deists and Christians exchanged volleys, dangers were in preparation for both. One was a revival of materialism (*q.v.*) inspired by the mechanistic explanations of modern science. Francis Bacon leaned toward it and Thomas Hobbes embraced it thoroughgoingly. In France at about the same time Descartes tried to save spiritual values by a theory of dualism, but his mechanically operated material world gave encouragement to bolder men who followed (*see* especially CARTESIANS). By 1750 the bars against extreme opinions had been somewhat lowered and several materialists ventured out into the open. Julien Offray de Lamettrie proclaimed that all things were reducible to matter in motion and he defiantly entitled one of his books *L'Homme machine.* The baron d'Holbach insisted that the cosmos was only matter without freedom or immortality or even first cause, and he looked on religion as the main source of human degradation. Few, however, were prepared to accept such extravagances.

A more subtle danger for Christianity appeared in the skepticism of David Hume. Dissecting with amiable reasonableness, Hume cut away all certainty in human affairs: the soul of man, the powers of reason, the existence of God, cause and effect—all the old anchors tore loose in a great storm of doubt. Immanuel Kant tried to save reason and God from the wreckage, but he was able to do so only by setting radical limits to the domain of reason and by rediscovering God through a nonrational course.

Meanwhile there had been an even more important response to the arguments questioning Christian belief. Religion, it was asserted, did not need to submit to a test of rationality. The mystics and others of unassailable faith went their separate ways in reaffirmed belief. Blaise Pascal, on the other hand, felt agonizingly the tension between the old and the new but found the answer in his wager (*pari*) on the existence of God and in his cry: "The heart has its reasons which reason does not know." A too disputatious theology, with an overintellectualized and dehumanized deity, provided little spiritual nourishment, and a reaction was inevitable. Indeed, at the high noon of rationalism John Wesley brought great masses of Englishmen back to an engrossing concern with sin and salvation through Christ. At the same time Jean Jacques Rousseau preached his own sermons on the religion of the heart. The attempt, therefore, to apply natural-law rationalism to religion provoked a rebellion which led to rationalism's defeat.

Psychological and Ethical Theory.—In their early thoughts about man the rationalists were able to deduce happy conclusions from his relations to God and nature: man was good and reasonable and individual men were essentially equal. With these new convictions they rejected Thomas Aquinas' limits to the powers of unaided reason and scorned the doctrine of the lasting taint of original sin. There was, however, a problem: many men in the flesh were neither rational nor good and they were certainly not equal in matters that counted.

John Locke showed the way to an explanation. His *Essay Concerning Human Understanding* (1690) was a study of the origins of ideas, but its crucial thesis was the shaping power of environment. If men were not rational and good, the fault was not that of inborn defect but of bad surroundings. When reason was not obscured it would tell man what was right and he would do right.

This assumed identity of knowledge and virtue early fell into trouble. Thomas Hobbes had pictured man as moved only by calculations of his own pleasure and pain. Believers in natural goodness indignantly protested, but the ugly portrait could not be destroyed. Certain theologians found it congenial to their doctrine of rewards and punishments and eternal happiness. More important still, a growing number of secular-minded reformers, notably Claude Adrien Helvétius in France and Jeremy Bentham in England, proposed to use pleasure and pain as the levers for reshaping human society. (*See* HEDONISM; UTILITARIANISM.)

A second line of thought broke the link between reason and goodness more insidiously. Even the most exalted rationalists assumed that there were moral promptings in the nonintellectual part of the human being; and two generations and more sought a precise identification of their source. Early in the 18th century the 3rd earl of Shaftesbury thought he found that it is an instinctive affection for the good because of its inherent worth, its beauty and its reflection of nature. Joseph Butler identified it as conscience, the voice of God in human life. Even Hume dropped his skepticism long enough to talk of an innate sentiment of humanity, and his friend Adam Smith, the political economist, gained his first celebrity by a book (1759) which identified sympathy as a directing emotion in conduct.

The discussion raised a question as to the relation between reason and this other ethical faculty. Shaftesbury was persuaded that rationality and moral sense could never be at odds because both were the gifts of nature; Butler declared that in conscience itself there was a power of reasoning. In the second half of the 18th century this comforting assurance began to disintegrate. Adam Smith was unable to say that sympathy and reason would invariably counsel the same decision, and Hume, at his most destructive, declared that reason was incapable of ethical judgment. In Rousseau this separation of reason and feeling turned into warfare, and soon novelists were playing variations on the theme that goodness was merely a matter of unleashing emotion.

These two lines of thought about man, like the long debate on religion, undermined the Enlightenment. In both cases, however, developments were gradual and often for years limited in their effects. Until toward the end of the 18th century there persisted a wide assent to the assumptions and the methods of the natural-law synthesis.

Political Theory.—When the assumptions and methods of the natural-law synthesis were applied to social problems, they required a significant modification. Manifestly there was a distinction between the natural law that kept the planets in their course and the natural law that pertained to man. Hence it was necessary to say that natural law prescribes in certain realms what *is*, in others what *ought to be*. Since in so many respects the human condition fell far short of the ideal *ought*, the Enlightenment became critical, reforming and eventually revolutionary.

When environment was found to be the cause of men's being less than reason and nature intended them to be, a vast program of

change became a moral imperative. A first necessity was intellectual freedom, a hacking-away of those shackles with which for centuries religious obscurantism and intolerance had crippled the human mind. But freedom, however precious, was not enough. Proper reasoning, said a public voice of increasing power, required better and more widespread education.

Subtly all creatures who partook of human nature came to be represented collectively as humanity, and humanity called for love and compassion. At home the sensitized conscience brought before the bar of reason the brutalities of the law and demanded a sympathetic understanding of the mentally ill. Lifting its vision beyond Europe, the humanitarian spirit protested against that most inhumane of practices, human slavery.

Governing these special concerns for justice was the concept of all men linked by nature in universal brotherhood. Standing in the way of the ideal were many grim realities. Most ominous among them was the sovereign state, that insatiable Leviathan which demanded the obedience and the loyalty and, far too often, the blood of its people. Rationalism responded with a defiant call to world citizenship. Almost as if sensing the imminent outburst of modern nationalism, men as varied in temperament as Voltaire and Samuel Johnson and Goethe deplored the irrationality of the patriotic prejudice. War was the great enemy of civilization, and men of the Enlightenment denounced it passionately. Building on plans attributed to Henry IV of France, a number of writers worked at designs for an international federation that might hold the peace.

The application of the rationalist's tools to politics led, as in other applications, to no one fixed and final conclusion. The baron de Montesquieu, analyzing his materials in the questing spirit of a Galileo, found what he believed to be the laws behind the laws by which men were governed. Others, impatient with theory, transmuted reason into reasonableness and were ready to accept despotism provided only that it be enlightened. In its political thought, however, the epoch was too intent on reform to rest with Montesquieu and too much committed to rationalism to share happily with Voltaire the hospitality of the philosopher-king Frederick II of Prussia. The more representative and the more significant course of political thinking, therefore, took another direction.

With the old religious framework broken, the Enlightenment began by considering the nature of man and by inquiring into the circumstances that had produced the state. The essential concepts —the ultimate authority of the people and a sociopolitical contract —were doctrines stemming from antiquity which had been revived and debated vigorously during the period of the Reformation. The first philosopher to manipulate them with the hard logic of rationalism tried to fashion a justification of sweeping absolutism: Thomas Hobbes admitted no place in natural law for the "ought to be"; that law prescribed human behaviour as fixedly as it ordered the seasons. What nature prescribed, according to Hobbes, was self-preservation, and before the organization of society men followed the prescription so ruthlessly that life was "solitary, poor, nasty, brutish, and short." Then, in desperation at this war of all against all, a compact was formed whereby individuals resigned their separate powers of self-preservation to a sovereign outside the contract and beyond restraint.

Locke half a century later accepted Hobbes's strong individualism and his concept of the state as a mere mechanical convenience, but in all other respects his arguments took a contrary turn. According to Locke, the state of nature was a social regime of good will in which natural law had endowed men with rights, cardinal among which were life, liberty and property. There were certain confusions when every man defined his own rights, and minor difficulties rather than desperation inspired the formation of the political community. On the surface, the second of Locke's *Two Treatises of Government* shone with common sense and moral appeal: government was derived from the consent of the members of the state and in practice consent meant rule by an undefined majority and an ultimate right of rebellion. Locke's discussion of the control of the state quickly became the great defense of the Glorious Revolution of 1688 and a Locke so appropriated emerged as the most conservative of revolutionaries. Public opinion was for the most part agreed that England's great problem had been solved and interest in social-contract theorizing declined rapidly. On the other hand, Locke's limitation of the state's functions to the protection of natural rights was to have a different history: in a century of rapid economic expansion the inviolability of liberty and property became a cherished doctrine prejudicial to the old order.

While English politics settled into complacency, in France the decay of absolutism gave a pressing urgency to political thought. Hope lay in borrowing from abroad, and the process reached an important dimension when Montesquieu celebrated the British constitution and when Voltaire introduced Newton and Locke. The Locke so enthusiastically received was stripped of his English moderation and made the source of dogmatic theorizing. Rousseau added complications and confusions. As he took up one by one the standard phrases of his day he distorted them beyond recognition, and in so doing he was a force destructive of the Enlightenment. Yet in one respect he rounded out rationalist thought with a theory of political democracy which Locke had failed to see. In America, in 1776, Thomas Jefferson incorporated the strict Lockean doctrine in the Declaration of Independence as "the common sense of the subject." In 1789, when they wrote their own declaration, the French Revolutionaries borrowed heavily from Locke, but they also made a place for Rousseau's sovereignty of the nation and the "general will."

The main stream of thought readily accepted property as a natural right. In the seething atmosphere of the second half of the 18th century, however, dissenters made their appearance. The abbé de Mably (Gabriel Bonnot), starting with equality as a first principle of natural law, discovered that the golden age had been destroyed by the introduction of private property. His contemporary Morelly held it responsible for the perversion of man's nature, and his successor "Gracchus" Babeuf tried to organize a revolution within a revolution to secure that true equality ordained by nature. In England William Godwin, starting with the immutable laws of the universe, rigorously pushed reason into condemning not only property but even marriage and the state.

For the Enlightenment the law under which men actually lived also required urgent consideration. There could hardly have been a greater contrast between the law of the courts and that which reason perceived in nature. Once Grotius had given his definition, the law of all lands was summoned to the supreme tribunal of reason. Man-made law, said the indictment, was valid only in so far as it was in conformity with nature's law. Even the well-known jurist William Blackstone, steeped as he was in the English common law, accepted the argument. On the continent of Europe, where there was little ground for contentment, legal thought, like political, became doctrinaire: law, said the zealots, was not made but discovered, and the way to reform was through a reasoned code, a summation of legal wisdom which could be administered with mechanical precision. The outcome was partial victory in the Prussian and Austrian codes and in the French *Code Napoléon*.

Economic Theory.—Another problem of human affairs to which the Enlightenment brought its judgment was the economic. The background was a long history of extensive state intervention in economic pursuits and a recent history of middle-class dissatisfactions. In France the wars of Louis XIV had wrought great misery and even before his death critics had dared to propose reforms. Among the first to apply rationalist criteria was P. Le Pesant de Boisguillebert, author of the expression *"laisser faire la nature et la liberté."* In his judgment, state interference did violence to the law of nature which prescribed a harmony of interests; right reasoning under the stimulus of free competition was the proper guide to economic behaviour. A half-century later Louis XV's physician François Quesnay founded the physiocratic school (*q.v.*) on essentially the same assumptions. The physiocrats' economic man was motivated by self-interest but lived under nature's dictate of harmony; hence it followed that unfettered private activity was the reasonable means to the common weal.

Adam Smith's *Inquiry Into the Nature and Causes of the Wealth of Nations,* appearing in 1776 as a declaration of economic inde-

pendence, criticized the physiocrats for their emphasis on the unique importance of land, but otherwise Smith revealed his kinship. He began with the butcher and the baker intent on their own good and went on to demonstrate the merits of "the system of natural liberty." Sharing the British dislike of extremes, he was not dogmatic, but behind his array of observations and arguments there lurked the old rationalist conviction about "the great Director of Nature" leading men "by an invisible hand" to serve a wider cause than their immediate own.

The Idea of History.—The Enlightenment's view of history began with a dilemma. The basic rationalist values were timeless, and the purpose of examining the world's experience was to strip away the accidents of time and place and to lay bare Hume's "constant and universal principles of human nature." Another service, however, was soon required. When real men were so far removed from the "man" of reason and nature, it was the task of the historians to explain the causes of the gap. To their labour most of them brought a point of view which suggested the answers. In consequence the numerous historical works which were so popular after 1750 glossed the theme of the evil consequences of religious and political benightedness.

This quietly intruding sense of the uses of history invited the further question of where the Enlightenment stood with respect to the past and what one could expect from the future. Earlier speculations were not useful. Christian philosophy had discounted the importance of the earthly city, and Renaissance humanism was disposed to see only decay since the high days of antiquity. Francis Bacon was in the vanguard of those who challenged both traditions. The Greeks and Romans, he argued, belonged to the youth of the world; his own generation was that of its old age, the heir of all accumulated knowledge. One could look forward with great hope; good method would mean greater and greater conquests of the secrets of nature.

Literary Aspects.—Before Bacon's death in 1626 an even more explicit comparison between the classical world and the modern began to take shape. At issue were the relative merits of the two literatures. To those bred in reverence for Homer and Virgil it was sacrilege when certain critics preferred authors of their own time. For a whole century Europe resounded with the "warfare of the ancients and the moderns" (*see* ANCIENS ET DES MODERNES, QUERELLE DES). The battle raged on the subjective field of taste, and a clear-cut victory was impossible. When in 1683, however, B. Le Bovier de Fontenelle shifted the ground from literature to science, few could deny the reality of progress. But Fontenelle, like Bacon, was primarily concerned with the future, and he detected a law of progress determining its destiny.

The Idea of General Progress.—As social dissatisfaction mounted, the idea of an intellectual progress became transformed into a belief in the *general* progress of mankind. Hard-headed men like Voltaire saw happy vistas ahead provided only that reason came into its rightful empire, but after 1750 the vision of things to come became too intense to wait upon an if. The young A. R. J. Turgot, the future minister of Louis XVI, surveyed the whole of human history and found it a record of the progress of the race toward perfection; even in ages of apparent barbarism men and society had moved, however slowly, in the right direction and in the future would continue to do so under a law of acceleration. Turgot's friend the marquis de Condorcet converted these calm assurances into fervent prophecy. Hiding in 1793 from the Revolution which he had joyfully served, Condorcet hastily composed his *Esquisse d'un tableau historique des progrès de l'esprit humain.* He saw mankind facing its tenth and culminating epoch, a new day in which the steady march toward perfectibility could not be stopped. The time would come, said this man under sentence of death, when the sun would shine only on free men who knew no master but their reason.

The End of the Movement.—Condorcet summarized the last and most eloquent phase of the Enlightenment. Already the movement was in disintegration. For all its belief in eternal principles, it was from the beginning a flow of ideas in which emphases shifted and conclusions tended to break away from the original synthesis. Religious opinion ranged between the poles of atheistic materialism and of anti-intellectual emotion; the prestige of reason was impaired by Hume's doubt and by Rousseau's hostility; nature broke up into something to be studied with scientific objectivity and something to be enjoyed in romantic indulgence; natural law crumbled under mounting disbelief; man, widely suspected of the selfishness of which Hobbes had accused him, lost his full heritage of dignity and sanctity.

Posterity has heaped both ridicule and scorn on the naïve assumptions of the Enlightenment, but the movement was more than mere self-deception. Its moral significance was high in that it tried to state in a new idiom the oldest values and beliefs in the history of western civilization.

See also references under "Enlightenment" in the Index volume, and biographies of most of the persons mentioned in this article.

BIBLIOGRAPHY.—For a systematic bibliography *see* Preserved Smith, *A History of Modern Culture,* 2 vol. (1930–34); and the series "Rise of Modern Europe," ed. by W. L. Langer, vol. 5–10, namely C. J. Friedrich, *Age of the Baroque, 1610–1660* (1952); F. L. Nussbaum, *Triumph of Science and Reason, 1660–1685* (1953); J. B. Wolf, *Emergence of the Great Powers, 1685–1715* (1951); P. Roberts, *Quest for Security, 1715–1740* (1947); W. L. Dorn, *Competition for Empire, 1740–1763* (1940); and L. Gershoy, *From Despotism to Revolution, 1763–1789* (1944). *See* also Carl L. Becker, *The Heavenly City of the 18th-Century Philosophers* (1932, 1959); G. H. Sabine, *A History of Political Theory* (1950); I. Berlin, *The Age of Enlightenment* (1956).
(D. HA.)

ENNA, a town and episcopal see, since 1926 capital of the province of Enna, Sic., lies 153 km. (94 mi.) S.E. of Palermo by rail and 90 km. (56 mi.) W. of Catania. Pop. (1957 est.) 29,130 (commune). It is situated on a conspicuous plateau (3,110 ft. above sea level) almost in the centre of the island, commanding magnificent views. A Sicel city and centre of the pre-Hellenic cult of Demeter and Kore (Persephone), the ancient Henna early came under Greek influence, first from Gela (7th century B.C.) and later from Syracuse, after it fell into the hands of Dionysius I of Syracuse about 397 B.C. After a brief spell (4th century) of Carthaginian rule it passed to the Romans in 258. In 134–132 it was the headquarters of the great Sicilian slave revolt (*see* SICILY: *History: Roman Sicily*). Of little note in later classical times, its natural strength and strategic position gave it renewed importance in the middle ages. The Saracens held it from 859 to 1087, its medieval name, Castrogiovanni, deriving from the Arabic Kasr-Yani, a corruption of Castrum Hennae. A favourite residence of Frederick II of Hohenstaufen and of Frederick II and III of Aragon, it was among the first cities of Sicily to rally to the Italian cause in 1848 and 1860. In 1927 it resumed its ancient name. An octagonal tower of Frederick II of Hohenstaufen, the cathedral (1307) and the castle, with work of many periods, are the main buildings of interest. The lake of Pergusa, 4 mi. S., is the traditional scene of the seizure of Persephone by Pluto. (J. B. W.-P.)

ENNIS (INIS), the county town of County Clare, Republic of Ireland, lies on the river Fergus, 23 mi. N.W. of Limerick by road. Pop. (1961) 5,678. Incorporated in 1612 by James I, it is now controlled by Ennis urban district council set up in 1898. There is a Roman Catholic pro-cathedral (diocese of Killaloe). The Franciscan abbey founded about 1242 was renovated in 1954 and is now a national monument. The ruins of Clare abbey, founded in 1195, and of Killone abbey are near the town. Ennis is on the main Limerick-Galway road and is the principal rail and road junction of County Clare. Sixteen miles south is Shannon airport, the most westerly airport in Europe. Ennis has flour and whiting mills, textile and imitation jewelry industries.

ENNISCORTHY (*Inis Corthaidh*), a market town of County Wexford, Republic of Ireland, on the river Slaney, 15 mi. N.N.W. of Wexford by road. Pop. of urban district (1961) 5,756. The castle was founded in 1176. Ferns, nearby, was the seat of a former bishopric; its cathedral, ruins of an Augustinian monastery founded about 1160 and ruins of a Norman castle remain. Enniscorthy, which has a cathedral, was incorporated by James I. It was taken by Cromwell in 1649 and burned in 1798 by the United Irishmen, who were subsequently defeated on nearby Vinegar hill. There are breweries, flour mills, potteries, cutlery and bacon factories. Enniscorthy is the centre of an extensive agricultural district, and has livestock fairs.

ENNISKILLEN (Inniskilling), a rural district and municipal borough, the county town of County Fermanagh, Northern Ireland, is located 75 mi. W.S.W. of Belfast by road. Pop. (1961) 7,438 (municipal borough); 15,595 (rural district). It is situated on an island (ancient Innis Cethlen, *i.e.*, Cethlen's island) between Upper and Lower Lough Erne, at a strategic crossing point commanded by an early 17th-century castle. Anciently a stronghold of the Maguires and incorporated by James I, it defeated a force sent by James II in 1689. Enniskillen gained a reputation as a Protestant stronghold. Long a garrison town, it gave its name to two famous regiments of the British army (Royal Inniskilling Fusiliers [27th foot] and the 6th [Inniskilling] Dragoons). Oscar Wilde was a student at the Royal school, founded in 1618. Enniskillen makes nylon hose, bacon is cured and there is trade in agricultural produce. A cattle auction market was built in 1949. Two miles to the north is Devenish Island with the abbey of St. Mary, founded by St. Molaise (Laserian) in the 6th century, and with a round tower 81½ ft. high and a cross.

ENNIUS, QUINTUS (239–169 B.C.), the greatest and most influential of early Latin poets, deservedly called the "father of Roman poetry." Born at Rudiae in Calabria, he was at home in three languages, or had, as he put it, "three hearts": Oscan, though he was of Messapian stock, apparently was his native tongue (his nephew Pacuvius bears an Oscan name); Greek was the language in which he was educated, perhaps at Tarentum; and Latin the language of the army with which he served in the Second Punic War. In 204 he was stationed in Sardinia, whence the elder Cato took him to Rome. In Rome he earned his living as a teacher and by adapting Greek plays. But although his circumstances always remained very limited, he was on familiar terms with many of the leading men of Rome, among them the elder Scipio. A fragment of the *Annales* seems to give, in historical disguise, a self-portrait of the poet and a sketch of such a relationship. His patron was M. Fulvius Nobilior, whom he accompanied on his Aetolian campaign in 189 B.C. Five years later, by being enrolled in a colony founded by his patron's son Quintus, he obtained Roman citizenship, probably taking the name of Quintus on that occasion. Nothing else of any significance is known of his life, except that he continued to reside at Rome and remained productive as a poet until his death. Both Horace's remark that he never worked at his epic unless well in his cups and St. Jerome's statement that he died of gout must be viewed in the light of his own quip in the *Satires* that he never "poetized" unless he had a bout of podagra.

Ennius' main work was the *Annales*, a narrative poem telling the story of Rome from the wanderings of Aeneas to the poet's own day. Begun hardly before 184 B.C., it was apparently edited in sections. About 600 lines have survived, from which it is possible to see how the matter was arranged. Books i–iii contained the origins and the regal period, iv–vi the early history up to and including the war against King Pyrrhus, vii–ix the Hannibalic war (the First Punic War was either sketched very briefly or omitted altogether because Naevius had written of it), x–xii the war against Philip of Macedon and the liberation of Greece, xiii–xv the war against Antiochus the Great and the Aetolian campaign. Here, it seems, with the triumph of his patron Fulvius, the poem was to have ended, but Ennius subsequently added three more books to bring the story up to date. Judged as an epic poem, the *Annales* lacks unity and suffers from the incongruity of mythological elements with eyewitness accounts of contemporary history. But the concept as a whole possesses grandeur, and the execution, though faulty in some details, generally seems to have matched it in vigour and dignity.

Ennius excelled also in tragedy. The titles survive of 19 plays adapted from the Greek, mostly from Euripides; *e.g.*, *Iphigenia (in Aulis)*, *Medea*, *Telephus* and *Thyestes*, the last-named written apparently just before the poet's death. About 420 lines remain, and indicate considerable freedom in the handling of the originals. Remarkable is the strengthening of the rhetorical element and the heightening of the pathetic appeal, largely by skillful use of the Latin features of alliteration and assonance.

Roman plays (*praetextae*) were *Sabinae* and, if they were plays at all and not some other form of composition, *Ambracia* (on the capture of that city by Fulvius) and *Scipio*.

In the *Satires* (*satura*, "medley") Ennius created the only literary genre which Rome could call its own. Written in four books, in a variety of metres on diverse subjects, they were mostly concerned with practical wisdom, often driving home a lesson with the help of a fable. Akin to the Horatian satire, they never apply the lash in the manner of Lucilius. More technically philosophical were two works about which little is known: a *Protrepticus* or *Praecepta*, and an *Epicharmus*, on the theories of the Sicilian poet and philosopher, both written in trochaic tetrameters. There are also fragments of a prose translation of the work of Euhemerus (*c.* 300 B.C.) on the origin and nature of the gods. In a lighter vein apparently was the *Sota*, in Sotadic verse, called after the 3rd-century poet Sotades, and the *Hedyphagetica*, a gastronomical poem based on a work by Archestratus of Gela (4th century). The somewhat experimental hexameters of this work, which can be dated after 189 B.C., seem to provide a *terminus post* for the composition of the *Annales*. Some epigrams, on the poet himself and on Scipio Africanus, are the first Latin elegiac couplets.

Ennius, who is credited also with the introduction of the double spelling of long consonants and the invention of Latin shorthand, was obviously a man of wide interests, conversant with the intellectual and literary movements of the Hellenistic world. Although his critics like to contrast his lack of artistry with the native talent which they all concede him, he was no rude pioneer, trying by the light of his poetic instinct only to transfer Greek classical poetry to Rome. His theoretical studies enabled him to find a way of benefiting from Greek taste and achievement without falling into slavish imitation. He admired the alliance of poetry and scholarship represented by the school of Alexandria, and it is no accident that the famous dream at the opening of the *Annales*, in which Homer on the mountain of the Muses reveals the passing of his soul into Ennius, owes as much to the *Aitia* of Callimachus as to the Pythagorean doctrine of the migration of souls. Ennius seems to have given symbolic expression to his more enlightened attitude by invoking the aid of the Greek Muses rather than of the Italic *Camenae*, to which his predecessors had addressed themselves. It is curious to note that his patron Fulvius at the same time introduced the Muses into the cult of Rome. Of the greatest importance was Ennius' decision in his epic to abandon the primitive Saturnian verse used by Livius Andronicus and Naevius and replace it by the Greek dactylic hexameter. Adapting the Greek line to the nature of the Latin language and fitting the language to the demands of the line, he created and did not fall unduly short of perfecting the mode of poetic expression which reached its greatest beauty in Virgil and was to remain dominant in Latin literature.

Admiration for the work of Ennius, often voiced by Cicero, was universal throughout the republican period, and his *Annales* remained the national poem of Rome until they were eclipsed by Virgil's *Aeneid*. A critical note is first struck by Horace and increases in severity, which reaches its climax in Seneca and Martial. The archaistic age of the Antonines returned to admiration, but from then on Ennius ceased to be read. The loss of his work is to be regretted, both for its own sake and for the light which it would throw on the poetry of Lucretius and Virgil, to mention only those most directly in his debt. "Would that we had him entire," said the great Scaliger, "and had lost Lucan, Statius, Silius Italicus *et tous ces garçons-là*."

Bibliography.—The fragments were ed. by J. Vahlen (1903), and with Eng. trans. in the "Loeb Series" by H. E. Warmington in *Remains of Old Latin*, vol. 1 (1935, repr. 1956). The *Annales* only, with notes, were ed. by L. Valmaggi (1900) and by E. M. Steuart (1925, repr. 1945). The dramatic fragments only were ed. by O. Ribbeck in *Scaenicae Romanorum poesis fragmenta*, vol. 1 (1897). For the fullest and best account of Ennius, *see* F. Skutsch in Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft*, vol. 5 (1905), col. 2589–2629. On Virgil's indebtedness to Ennius *see* E. Norden, *Ennius und Vergil* (1915). *See* also O. Skutsch, *The Annals of Quintus Ennius* (1953). (Or. S).

ENNODIUS, MAGNUS FELIX (474–521), bishop of Ticinum (Pavia), major Latin poet and rhetorician. A member of the important and influential family of the Anicii, he was born

at Arelate (Arles) in Gaul, but lived in Ticinum and Mediolanum (Milan), in this age an important centre of learning. Though his interests were largely secular and literary he was ordained deacon (493) to Epiphanius, bishop of Ticinum. In 507 he was appointed by the pope to compose a panegyric on Theodoric expressing gratitude for the Arian king's tolerance of Catholicism. After a sudden illness he renounced secular pursuits in fulfillment of a vow. Appointed to the see of Ticinum in about 513, he played a not unimportant part in affairs: in 515 and again in 517 he was sent by Theodoric on an embassy to the court of the emperor Anastasius I at Constantinople. Ennodius has been represented as a friend of Theodoric, but his support of him may have been a consequence of the friendship between Theodoric and Epiphanius, and his Anician connection sufficiently explains his ambassadorial employment (Anicia Juliana was an important figure at Constantinople). He died at Ticinum.

Ennodius' literary output is considerable and varied: he composed occasional poems, among them two itineraries, one on a journey from Mediolanum to Brigantium (Bregenz), and another about a journey on the Padus (Po), two poems on works of art and another on a garden; these last two themes he shares with the near-contemporary African poet Luxorius. He also resembles Luxorius (and other poets represented in the *Anthologia Latina*) in the obscenity of many of his epigrams. His *Epithalamium*, a polymetric poem written for Maximus, is so far from being Christian in tone that it contains a complaint by Cupid against the Christian ideal of chastity; this poem is structurally weak and the varied metres used destroy even the appearance of unity. In a different vein is a series of twelve uninspired hymns intended for the Milanese liturgy.

His prose works include a biography of Epiphanius, which throws valuable light on the political activity of the church and is, together with his panegyric of Theodoric, an important source for the historian; *Dictiones*, a collection of model copies of speeches which reveal the continuance of the traditional rhetorical education and give a valuable description of the school of the grammarian Deuterius in Milan; epistles on a wide range of subjects (including some addressed to Boethius, to whom he was related); and the *Eucharisticum de vita sua*, a kind of confession. He also wrote, in a mixture of prose and verse, the *Paraenesis didascalica*, otherwise entitled *Ennodius Ambrosio et Beato*, a didactic treatise on grammar and rhetoric; the claim made by Rhetoric in this work that its servants are masters of the world (*qui nostris servit studiis mox imperat orbi*) illustrates its importance for Ennodius.

Much of his writing shows his devotion to pagan Roman tradition, which was zealously fostered by the Anician family, and like other members of this family he sought to reconcile this tradition with Christianity. The rhetorical bias of his training and interest is reflected throughout his works whose chief concern is with form: his style is affected and excessive, and in overelaboration and diffuseness it rivals that of Cassiodorus.

See editions by W. Hartel, *Corpus scriptorum ecclesiasticorum Latinorum*, vol. 6 (1882) and by F. Vogel, *Monumenta Germaniae historica. Auctores antiquissimi*, vol. 7 (1885). (D. R. Br.)

ENOCH (Henoch in the Douai version of the Bible) was the eldest son of Cain, who named a city after him (Gen. iv, 17, 18). But in the genealogy of the Priestly document (*see* Pentateuch) he appears as the seventh in descent from Adam in the line of Seth. The brief notice in Gen. v, 21–24, where he is identified as the father of Methuselah, is certainly a fragment from an Enoch myth. The fact that his years are given as 365 suggests that he was a solar hero. Enoch appears in the Old Testament also as the name of clans belonging respectively to Midian and Reuben.

ENOCH, BOOKS OF. Three ancient Jewish writings have been preserved under the name of the patriarch Enoch: the First Book of Enoch (I Enoch) or the Ethiopic Enoch; the Second Book of Enoch (II Enoch) or the Slavonic Enoch, also known as the Book of the Secrets of Enoch; and the Third Book of Enoch (III Enoch) or the Hebrew Enoch. Yet another is referred to in the epistle ascribed to Barnabas (2nd century A.D.), but it is no longer extant. The existence of these writings is evidence of the extent to which the figure of Enoch (Gen. v, 18, 21–24) must have dominated certain strands of Jewish tradition. They are among those writings called pseudepigraphal (*see* Apocrypha, Old Testament: *Other Apocryphal Works: Pseudepigrapha*).

I Enoch.—The Ethiopic Enoch (which is regarded as canonical by the Ethiopian Church) has been preserved, as the title indicates, in an Ethiopic version. Three manuscripts of this were brought to Europe in 1773 by the explorer James Bruce, and 29 in all are known. This Ethiopic version is not directly derived from the Semitic original but is based on a Greek translation made probably *c.* A.D. 400. Extensive quotations from the Greek text are made by the early 9th-century Byzantine chronicler George Syncellus, and in 1886–87 substantial fragments of an earlier Greek text were discovered at Akhmim in Egypt. Part of a Latin rendering is also extant. In 1952 a number of Aramaic manuscripts containing parts of the text were discovered at Qumran (*see* Dead Sea Scrolls). Whether the original text of the book was in Hebrew or Aramaic remains uncertain.

Like the Pentateuch, Psalms and Proverbs, I Enoch is divided into five sections.

1. In i–xxxvi a future judgment is predicted, evil is traced back to the fall of the angels (*cf.* Gen. vi, 1 ff.), and a description is given of Enoch's journey on the earth and in the underworld.

2. Ch. xxxvii–lxxi contains the Similitudes or Parables of Enoch, in which the central figure is that of the transcendental Son of Man. It has been argued, but not proved, that the passages which refer to this figure are Christian interpolations; its relation to the development of the messianic hope has been the subject of much controversy. The fragments from Qumran contain nothing from this part of the book.

3. Ch. lxxii–lxxxii is an astronomical section, devoted to the movements of the heavenly bodies.

4. Ch. lxxxiii–xc, a section which seems to have been influenced by the Book of Daniel, describes dream visions in which Enoch sees the various phases in the history of mankind (the nations being presented in animal form) up to the final judgment.

5. Ch. xci–cviii is a varied section, containing a survey of human history (the so-called Apocalypse of Weeks), exhortations and denunciations, a description of miracles at the time of Noah's birth and a concluding exhortation to the righteous.

I Enoch is not one work but a collection of literary material of diverse origin and date. The Similitudes are a fairly obvious unity, and it has been argued that certain passages in which Lamech and Noah are mentioned were originally connected with each other. On any showing, the process by which the collection reached its present form must have been complicated. It has been argued that some parts (viz., vi–xxxvi; xciii, 1–10; and xci, 12–17) are pre-Maccabaean, but this is unlikely; in all probability the material spans the period from the Maccabaean age to the first decades of the 1st century B.C. The problem of dating, however, is complicated by the vexed question of the existence and extent of Christian interpolations in the Similitudes.

I Enoch is a richly stored repository of eschatological teaching. In parts it strikingly exemplifies the characteristics of apocalyptic; *e.g.*, the bizarre symbolism, the presentation of past history as if seen in a vision of the future, the division of history into well-defined periods, and the highly coloured descriptions of the ultimate destiny of the righteous and the wicked (*see* also Apocalyptic Literature). Affinities with the Book of Daniel are evident, not least in the figure of the Son of Man. But whereas in Daniel the Son of Man represents the elect community, in I Enoch he should probably be regarded (in spite of arguments to the contrary) as an individual, a Jewish variant of the conception of the Primordial Man or Anthropos. Some of the language used of this figure in I Enoch suggests a messianic character. The view that the Son of Man is identified with the Suffering Servant of Isaiah xl–lv is probably to be rejected. At all events the supramundane character of the Son of Man in I Enoch has no counterpart in the Old Testament expectation of an ideal king or the prophetic description of the Servant. Its influence on the concept of the Son of Man in the Gospels is widely acknowledged.

In the New Testament Enoch himself is mentioned in Luke iii, 37; Heb. xi, 5; and Jude 14 (a quotation from I Enoch lx, 8). There seem to be reminiscences of I Enoch at various points in Jude 4–15. It has also been claimed that several other passages in the New Testament (*e.g.*, Matt. xix, 28 and xxvi, 24; Luke xvi,

9; John v, 22; I Thess. v, 3; I Pet. iii, 19 ff.; and various parts of the book of Revelation) show the influence of I Enoch. But in none of these passages (with the exception of Jude 4–15) is direct influence demonstrable. The recurrence of similar ideas and phraseology need indicate no more than indebtedness to a common tradition.

II Enoch.—The Slavonic Enoch, which exists only in a Slavonic version of a lost Greek original, is an apocalyptic work based, presumably, on those Jewish traditions about Enoch which also provided material for I Enoch. It begins with a description of a journey, undertaken by Enoch in a visionary experience, through the heavenly realms, in which he sees the various orders of angels and the heavenly bodies. This description represents the cosmological strain in Jewish apocalyptic. In the sequel there is a transition from cosmology to history: a revelation is made to Enoch of the course of history presented in a series of well-defined periods in an artificial pattern, a feature familiar from other apocalypses. Finally Enoch goes to impart to his sons the instruction which he has been given, and then is translated to heaven.

The book exists in two recensions, of which the shorter is the older. Although the substance of the book seems to go back to a Jewish original dating from before the destruction of the Temple in A.D. 70, probably coming from the Greek-speaking Diaspora and possibly from Alexandria, there are extensive signs of Christian influence. That the basis of the work is Jewish tradition needs no demonstration; but it may be argued whether it is a Jewish writing which has been modified by Christian revision, a mainly Christian work on a Jewish theme and incorporating Jewish material or an essentially syncretistic work expressing the themes of an Enoch cult.

III Enoch.—The Hebrew Enoch is a characteristically rabbinic compilation and interpretation of the traditions about Enoch, highly bizarre and speculative in character. It is yet another testimony to the extraordinary strength and pervasiveness of the Enoch legend. It has been assigned to the latter part of the 3rd century A.D., but later dates have also been suggested.

BIBLIOGRAPHY.—*I Enoch:* R. H. Charles, *The Ethiopic Version of the Book of Enoch* (1906), *The Book of Enoch* (1912), *I Enoch* in *The Apocrypha and Pseudepigrapha of the Old Testament,* vol. ii (1913); C. Bonner, *The Last Chapters of Enoch in Greek* (1937); H. Ludin Jansen, *Die Henochgestalt* (1939 [1940]); N. Messel, *Der Menschensohn in den Bilderreden des Henoch* (1922); S. Mowinckel, *He That Cometh,* ch. x (1956); R. Otto, *The Kingdom of God and the Son of Man* (1943); T. W. Manson, "The Son of Man in Daniel, Enoch and the Gospels," *Bulletin of the John Rylands Library,* 32:171–193 (1950).

II Enoch: W. R. Morfill and R. H. Charles, *The Book of the Secrets of Enoch* (1896 [1895]); G. N. Bonwetsch, *Die Bücher der Geheimnisse Henochs* (1922); N. Forbes and R. H. Charles, *II Enoch* in *The Apocrypha and Pseudepigrapha of the Old Testament,* vol. ii (1913); A. Vaillant, *Le Livre des secrets d'Hénoch* (1952).

III Enoch: H. Odeberg, *3 Enoch or the Hebrew Book of Enoch* (1928). (G. W. An.)

ENOMOTO BUYO (Takeaki) (1836–1908), Japanese naval officer, diplomat and statesman. He was descended from a family that had long been attached to the Tokugawa shoguns. He first demonstrated his great energy and patriotism at the age of 18 when he left the shogunal college to make a military survey of Hokkaido and then went to Nagasaki to study with Dutch naval instructors. He was one of 13 young officers sent to the Netherlands to study naval construction. On his return in 1866 he became vice-commander of the shogunate fleet and for the next three years remained loyal to the shogun in opposition to those who favoured restoration of the emperor. He fought to the last against the imperial forces in Hokkaido, which he had hoped to make a republic. After surrendering in 1869 he spent three years under house arrest and on release and pardon became secretary-general for the colonization of Hokkaido. Restored to favour and preferment he began a lengthy ministerial career with appointment as envoy extraordinary to Russia (1873–76), where he concluded the treaty of St. Petersburg by which Japan gave up its claim to Sakhalin island in exchange for the northern Kurils. He next served as navy minister (1876–1882) and then as minister to China (1882–84). Following came the portfolios of communications, education, foreign affairs, and agriculture and commerce. He was created viscount in 1887 and privy councilor in 1892.

Enomoto died in 1908 and was succeeded in the peerage by his son Takenori. (J. A. Ha.)

ENSCHEDE, the largest town in the province of Overijssel, Neth., is the centre of the Twente industrial region and lies about 164 km. (102 mi.) E. of Amsterdam by road, and near the German frontier. Pop. (1957 est.) 118,671 (mun.). Although badly damaged in World War II, Enschede has been largely rebuilt, with a 70 m. (230 ft.) wide boulevard. Principal features are the town hall (1933); the municipal theatre (1955); the Twente national museum, with a unique collection of animal paintings; the natural history museum and vivarium; the Roman Catholic and Dutch Reformed churches and the modern synagogue. In Het Volkspark, the chief public open space, is the war memorial, consisting of six statues. There are good communications by railway, road and canal with the rest of the Netherlands and with Germany; Twente military airport is 3½ mi. N. of the town. Enschede is a major centre of the Dutch cotton industry and its merchants trade extensively in the U.S. cotton market at Memphis, Tenn. Clothing, metal and rubber goods, and tires are also made. It is the meeting place, every three years, of De Kogge, an association of Dutch, Flemish and German writers. About two-thirds of the town was burned down in 1862. In World War II it was occupied by the Germans from May 1940 to April 1945. (H. Hs.)

ENSENADA, CENÓN DE SOMODEVILLA, Marqués De La (1702–1781), Spanish statesman, whose administration was notable for the vigour of its policy of internal reform, was born at Alesanco (Logroño) on June 2, 1702. He began his career as an apprentice of José Patiño at the ministry of marine, where, by 1728, he had risen to the rank of *comisario real.* He was put in charge of work at the new naval arsenal at El Ferrol, and accompanied the expedition which carried out the conquest of Oran in 1732. He organized the expedition of Don Carlos (afterward Charles III of Spain) to Naples in 1736, and was rewarded with the Neapolitan title marqués de la Ensenada. From 1737 he continued his work with the Spanish navy as secretary to the admiralty council and naval intendant. He was sent as secretary of state and war to Don Felipe, duke of Parma, after the renewal of the war in Italy in 1741, and Ensenada was still abroad when, in April 1743 following the death of José del Campillo, Philip V summoned him to Madrid to take office as secretary for war, marine, finance and the colonies. Shortly afterward he was also made secretary of state. Retained in office by Ferdinand VI after his accession in July 1746, Ensenada remained effective prime minister until July 1754.

Ensenada continued the reforms begun by Jean Orry, Cardinal Alberoni, Patiño and Campillo. Like them he encouraged the development of trade and industry, improved the navy, removed abuses in the customs system and sought to reform the methods of tax collection. In particular he boldly tried to extend to the whole country the scheme for the *única contribución* (a single tax to replace the multiplicity of separate taxes then existing) which had been planned by Orry as early as 1702 but which hitherto had been applied only in Catalonia. Ensenada drew up a remarkable series of reports on the finances and general condition of the country which foreshadowed the more searching examinations conducted under Charles III.

Intrigues by Richard Wall (*q.v.*), who disagreed with Ensenada's policy, brought about the latter's downfall in July 1754, when he was banished to Granada. Although again received at court after the accession of Charles III, Ensenada was once more exiled in April 1766, this time to Medina del Campo, where he died on Dec. 2, 1781.

See R. Bouvier and C. Soldevila, *Le Redressement de l'Espagne au xviii^e siècle: Ensenada et son temps* (1941). (B. J. R.)

ENSIGN, a military or naval rank, or a flag, especially the national flag. The term is most often applied to the flag flown at the stern by naval vessels in commission or by merchant vessels. The U.S. navy's ensign is the same as the national flag, but many other navies have distinctive naval ensigns which are "worn" by their war vessels. In the Royal Navy the ensign has a red, white or blue ground with the Union Jack in the upper corner

next to the staff. Until 1864, ships of the Royal Navy were divided into three squadrons and flew the red, white or blue ensign to indicate the squadron to which they were assigned. Since 1864 the white ensign (further distinguished by having a red St. George's cross quartered upon it) has been reserved for use by the Royal Navy and by the Royal Yacht squadron. Passenger liners or other merchant vessels manned by a prescribed percentage of officers and men of the Royal Naval reserve are entitled to fly the blue ensign. Certain other vessels, not of the Royal Navy but owned by the British government, also use the blue ensign. All other merchant vessels display the red ensign. (*See* FLAG: *British and Commonwealth Flags.*) The hydrographic office of the U.S. navy and the British admiralty both issue books which show all naval ensigns and national flags.

In the U.S. navy, ensign has been the lowest commissioned rank since 1862, when it replaced "passed midshipman." The continental army had a grade of ensign during the American Revolution, as did the U.S. army until 1815. The lowest commissioned rank in the British army until 1871 was ensign. As a military rank, ensign has generally been replaced by the rank of second lieutenant. *See* OFFICERS: *Ranks or Grades of Officers.* (J. B. HN.)

ENSILAGE is the preservation of green fodder in a suitable container by a process of controlled fermentation. The final product, a valuable foodstuff for farm animals, is called silage, and the container is known as a silo.

History.—The use of a *sirus* or pit for the storage of grain dates back at least to Roman times and is mentioned by Pliny (A.D. 23–79) though there is no reference to its use for the preservation of green forage at that time. The earliest description of the ensilage process as it is known today dates back to 1842 and is given in the *Transactions of the Baltic Association for the Advancement of Agriculture.* Pits were rapidly filled with fresh grass, as much air as possible being excluded by careful packing and compression. The mass was finally sealed off with boards on which was placed a layer of soil one foot thick. This is essentially the modern process.

There are earlier references to the principle of ensilage, such as that of Professor John Symonds of Cambridge who in 1786 described the storage and preservation of green leaves in wooden casks or pits. Alternate layers of green leaves and of unripe grapes were filled in until the pit was full when it was sealed off. The antiquity of this practice in Italy, where the observations of Symonds were made, is substantiated by quotations from Horace (65–8 B.C.).

The next stage in the development of the practice was in France where ensilage was introduced by A. Goffart and the Vicomte de Chezelles toward the end of the 19th century. They made use of large trenches lined with wood and often roofed over against the rain. Many different crops were used, among them corn or maize.

The practice spread to England and a very full account of the different types of containers then used was given by H. M. Jenkins in 1884 in a comprehensive report that he prepared for the Royal Agricultural society. In Great Britain the ensilage process developed toward the production of a high temperature fermentation using mature herbage. This according to George Fry (1885) was called the sweet ensilage process and did more to retard the development of ensilage in Great Britain than any other factor. High temperatures during ensilage always result in a marked lowering of the digestibility and thus of the feeding value of the resultant silage. Farmers were accordingly unwilling to make use of the process from which they did not get results comparable to those from haymaking.

It was on the North American continent that ensilage, introduced by French enterprise, reached its highest development. The underlying reason for this was the use of corn (maize), a crop which if cut at the correct stage of growth undergoes a rapid and desirable type of fermentation making an excellent silage. Early work was carried out at the land-grant colleges where the principles of the process were soon put into practice and readily adapted to everyday farm use. The dairy farmer's possession of a silo was considered essential. With the increased use of corn silage the cylindrical tower silo was developed, completely replacing the trench type of container. These tower silos, constructed of reinforced concrete, wood and other materials often reached a height of 25 to 50 ft., of variable diameter and holding more than 100 tons of silage.

In Europe and particularly in Britain the next development was the use of tower silos and special mixtures of crops. These silage mixtures combined a cereal such as oats or barley with a leguminous crop such as peas, beans or vetches. Various mixtures were grown for ensilage varying from the simplest, oats and vetches, to a complex one in which all five components were included. If cut at the correct stage, when the cereal was in ear with the grain formed but still soft and while the legumes were just forming seed pods, these mixtures ensiled well and the resultant silage was an excellent animal foodstuff. The crop, however, had to be chopped and elevated to the top of the tower silo, and this meant that the cost of growing the crop together with the high capital cost of a tower silo with a suitable chopper and blower to fill the silo gradually made the process uneconomic and for many years tower silos stood empty.

After the first quarter of the 20th century renewed interest was taken in the ensilage process. In North America corn and, to a lesser extent, sorghum and sunflowers were the staple crops but by the second half of the 20th century, they were being replaced to some extent by grassland herbage including the valuable lucerne or alfalfa crop. This has been brought about by the need to increase the production of protein on the farm and corn, though a valuable foodstuff, is deficient in this respect.

In Europe the need for self-sufficiency, especially in regard to proteins has also resulted in an increased production of silage. The Dutch had for many years made grass silage in simple heaps called clamps, probably leading the world in their knowledge of the utilization of grassland herbage, in which ensilage was an important factor.

The trend has thus been to turn to simple silos and away from the tower, reducing the initial capital outlay. Mechanization of the field harvesting of the crop has also increased and ensilage is now an acknowledged part of farm practice wherever grassland (*q.v.*) is a consideration and where farm animals, particularly ruminants, have to be fed.

The Process.—Any green forage crop after it is cut continues to live, the most noticeable biological process being respiration. This results in the production of carbon dioxide and the formation of heat. Both result from the breakdown of valuable, easily digestible, constituents of the plant, with a consequent loss in the nutritive value of the mass. Control of the amount of air is imperative and for this reason a container is advisable or the losses due to respiration may be excessive. Temperature, too, may have an effect on the feeding value of the resultant silage, as has already been shown. If the mass rises to 120° F., or more, the resultant reduction in digestibility may be serious. Such a silage is easily recognizable because of its deep brown colour, its sweet taste and burnt smell. Silage is sometimes made in a stack or clamp, in both of which the sides are unprotected and air has free access. As a result, the feeding value of silage made in this way must be less than where the air is prevented from penetrating the outer layers freely. The stack and clamp are often used where no silo space is available and is a useful emergency device.

After the main respiratory changes have taken place other processes, due to plant enzymes, result in a breakdown of the proteins in the plant tissue and also in the other constituents of the plant. These processes produce certain organic plant acids and after a while the enzyme action is reinforced by bacteria that were on the plant originally or found their way on to it while the crop was being harvested and loaded into the silo.

The main outcome of all these actions is the formation of acids which in good silage are predominantly lactic acid, with some acetic acid and small amounts of propionic acid. In badly made silage butyric acid is present and may be recognized by a rank smell like that of rancid butter. Where such a fermentation predominates, the protein is also degraded to an undesirable form, much of it being present as ammonia. Where lactic acid formation is rapid the acidity reaches a level of about pH 4.0, where all

other fermentations are kept in check and the undesirable butyric acid formation is controlled.

In practice, therefore, ensilage depends largely upon the control of air in the mass, and on rapid acidification. The first will depend upon the container, the degree of packing of the crop and its protection from the entry of air at the surface. A container that will allow the mass to pack tightly against the sides, preventing the free circulation of air, is desirable. This may be of any shape varying from the cylindrical upright tower silo to the trench either dug in the soil and left with earth sides or lined with a suitable material. It is also common to build two long containing walls above ground level thus making what amounts to an overground trench with airtight sides. In North America this is called a bunker silo. The crop should be packed tightly and the chopping of the material helps considerably. It should also be compacted, especially if it is at all stemmy or of low moisture content. With the modern trench or bunker silo this is easily done by passing the field equipment through the silo or running a tractor backward and forward over the material in the silo. Finally, some sort of cover should be added as soon after filling as possible. This will act as an air seal. Soil is the most practicable cover though many others have been used.

If the crop has been properly filled in and contains enough fermentable carbohydrate material sufficient lactic acid is formed to control all other fermentations and reach the desired acidity of around pH 4.0–4.5. Corn and the silage mixtures of cereals and legumes naturally fall into this category and make good silage. Young leafy green crops and purely leguminous crops tend to be low in fermentable carbohydrates, do not acidify sufficiently rapidly and may make bad silage, with much butyric acid and ammonia. To avoid this some easily fermentable carbohydrate may be added, a molasses solution being the most common. Other materials such as ground cereals, sugar beet pulp, potatoes or mangels have also been used and, if properly admixed with the protein-rich crops, stimulate the desired lactic acid formation.

Since acidity is the controlling factor it was suggested by M. Montanari in Italy in 1890 that a direct addition of acid might be beneficial. This idea was developed by Professor A. I. Virtanen (*q.v.*) of Finland in the A.I.V. process in which a suitable diluted solution of hydrochloric and sulfuric acids is added in such amounts that the desired acidity is reached at once, thus checking all biological changes at an early stage in the ensilage process. Other acids can be used, an organic acid such as the cheap formic acid, for instance. Since the major changes are due to bacteria and plant enzymes it has also been suggested that a sterilizing agent could control the process. Formalin and sulfur dioxide have been suggested and metabisulfites have been used with some success.

Whatever addition is made, the main difficulty is to ensure its even distribution throughout the mass. Molasses, ground cereals or similar materials are favoured since no harm can be done by an excessive use of them. Wilting the crop before ensilage has been widely advocated since this reduces the moisture content and improves chances of lactic acid fermentation. This process is sound but care has to be taken not to exaggerate its use since with overdrying the exclusion of air becomes correspondingly more difficult.

The value of silage depends upon the feeding value of the crop ensiled, since in good silage the composition is little altered and the ensiled product is as digestible as the fresh crop. It is wrong to look upon silage as a substitute for hay. If made from a young, leafy, protein-rich crop silage will have a very high feeding value. If made from good material it is also a valuable source of mineral matter and vitamins, particularly vitamin A or its precursor carotene.

While silage has been largely used in the feeding of dairy cows and young cattle its use is not restricted to this class of stock. Experience shows that good silage made from a crop of high feeding value can be very economically fed to fattening cattle. It can be fed to sheep but should be made from a short, leafy herbage since sheep normally graze such material. The horse can deal with small amounts of silage, of good quality. For pigs and poultry it is of value only as a supplement to their diet, supplying valuable mineral matter and vitamins, but not as a main source of energy or protein. *See* also Crop Drying and Processing; Crop-Processing Machinery; Feeds, Animal.

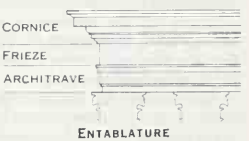
Bibliography.—A. J. G. Barnett, *Silage Fermentation* (1954); E. Crasemann, "Über Grünfutterkonservierung mit besonderer Berücksichtigung der künstliche Trocknung," *Landw. Jb., 31* (1954); N. D. Dijkstra, "Silage Making in the Netherlands," *Emp. J. Exp. Agric., 16* (1948); "What Has the State Agricultural Experiment Station at Hoorn Contributed to Research Into the Feeding Value of Roughage?," *Neth. J. Agric. Sci., 2* (1954); E. S. Hopkins and P. O. Ripley, "Silage Production," Dominion of Canada Department of Agriculture, *Pub. 525,* (1944); J. B. Shepherd *et al.*, "Experiments in Harvesting and Preparing Alfalfa for Dairy Cattle Feed," U.S. Department of Agriculture, *Tech. Bull. 1079* (1954); A. I. Virtanen, "The A.I.V. Method of Preserving Fresh Fodder," *Emp. J. Exp. Agric., 1* (1933); S. J. Watson, *Grassland and Grassland Products* (1951); S. J. Watson and A. M. Smith, *Silage* (1956); T. E. Woodward *et al.*, "The Nutritive Value of Harvested Forages," U.S. Department of Agriculture Yearbook, *Food and Life* (1939). (S. J. W.)

ENSOR, JAMES SYDNEY, Baron (1860–1949), Belgian artist, member of the powerful Postimpressionist generation of artist-innovators, was born on April 13, 1860, in Ostend, where he lived his whole long life. Remarkably precocious, Ensor met quick success with his early landscape, portrait and still-life paintings, two of which were shown at the Paris Salon of 1882, but the brightening of his palette from academic somberness to vivid and luminous colour soon shocked his contemporaries. Critical attack isolated Ensor in his native Ostend where he found a strange and prophetic inspiration in the elaborate carnival celebrations. Out of this local tradition and the masks which his mother sold in the family souvenir shop, Ensor plunged ahead in the late 1880s with phenomenal boldness and energy to create an art whose rich, improvised imagery and exuberant, freely splashed colour astonishingly anticipated 20th-century Expressionism and Surrealism. Ensor's drawings and etchings show the same extraordinary and uninhibited imagination, invention and facility as his painting. Early in the 1900s collections of his work in Germany influenced the development of the German Expressionists and particularly Paul Klee. Unlike such contemporaries as Van Gogh and Seurat, Ensor lived on to enjoy the irony of late recognition and was made a baron by the Belgian king in 1929 when his largest and most famous canvas, "Entrance of Christ into Brussels," was first publicly exhibited. Painted in 1888, it is a brilliant panorama of masked figures paying a carnival homage to the entering Saviour. The scene vibrates with the imminent crucifixion, and it is characteristic of Ensor's pictorial genius as well as the personal sense of outrage expressed in many of his fantastic works. He died on Nov. 19, 1949.

See *James Ensor* for illustrations and bibliography, Museum of Modern Art, New York (1951). (L. Tm.)

ENSTATITE, a rock-forming mineral belonging to the group of orthorhombic pyroxenes (*see* Pyroxene).

ENTABLATURE, in architecture, the assemblage of horizontal moldings and bands, supported by, and immediately above, the columns of classic buildings, or similar forms in any other style (*see* Order).



Entablature

It is usually divided into three main portions: the lowest, called the architrave, originally the beam running from support to support; the central one, called a frieze (*q.v.*), consisting of an unmolded band with or without ornament; and the topmost, called the cornice (*q.v.*), formed of a series of moldings that project from the face of the frieze. Occasionally, in Greek and Renaissance work, the frieze is omitted, as in the caryatid porch of the Erectheum at Athens. In Georgian and American colonial work, the architrave is sometimes omitted, as in many of the designs of Robert Adam.

ENTAIL (Fee Tail), in law, signifies an interest in land bound up inalienably in the grantee and then forever to his direct descendants. A basic condition of entail was that, if the grantee died without direct descendants, the land reverted to the grantor. The concept, feudal in origin, supported a landed aristocracy

since it served to prevent the disintegration of large estates through divisible inheritance or the lack of heirs.

In England in the 13th century, the judicial bias in favour of the free alienability of land caused gifts limited to a grantee "and the heirs of his body" to be interpreted as conditional on the birth of issue, so that once a child was born the land was considered to be held absolutely by the grantee, who could then dispose of it as he wished. In 1285, the statute *De Donis Conditionalibus* declared this practice to be contrary to the intention of the grantor and stipulated that the grantee could no longer alienate an estate, called a fee tail or entail, by disinheriting his heirs or by preventing its reversion to the grantor in the event there were no heirs. In 1472 a fictitious law suit known as the common recovery again permitted the owner to bar the entail and to acquire absolute ownership. Statutory reforms in England now permit the owner to convey the entailed land by a simple deed and even by will.

There were entailed estates in the American colonies, principally in the middle and southern colonies, but almost all the states emulated Jefferson's statute of 1776 for Virginia and abolished entails. *See* Real Property and Conveyancing, Laws of.

(A. Dm.)

ENTASIS, in architecture, the convex curve given to the taper of a column, spire or similar upright member, in order to avoid any appearance of hollowness or weakness. Entasis is almost universal in classic columns. It is exaggerated in Greek archaic Doric work, but grows more and more subtle during the 5th and 4th centuries B.C.

Many attempts have been made to find a mathematical basis for the entasis, and it has been reduced to all kinds of elliptical hyperbolic, parabolic and even cycloidal curves. The immense variety of forms found indicates, probably, that the curve was laid out free-hand and purely empirically. Entasis is occasionally found in Gothic spires and in the smaller Romanesque columns. *See* Greek Architecture; Order.

ENTEBBE, former administrative capital of Uganda, east Africa, is situated at the end of a peninsula jutting into Lake Victoria (*q.v.*), and its altitude (3,760 ft.) gives it a moderate summer climate all the year round. Vividly flowering trees and perpetually green lawns sweep down to the blue lake which is dotted with jungle-clad islands. Population in 1959 was 10,941 (Africans 9,087, Europeans 910, Asians and others 944).

The town is famous for its 70-ac. botanical gardens, which have supplied specimens of tropical trees and other plants to botanical gardens throughout the world. A veterinary research laboratory and a virus research institute are also established there. Entebbe's only outlet by land is a paved road to Kampala, the capital of Uganda, 21 mi. N.N.W. A weekly steamer service to Uganda, Kenya and Tanganyika lake ports calls at Entebbe. The town is also internationally important as the aerial crossroads of east Africa. Its airport, with a runway two miles in length, can accommodate the largest airliners. There is safe landing and harbourage for flying boats on the lake nearby.

Nearly all the inhabitants are engaged either in government service or in shopkeeping. There are no industries, exports or imports, although cotton, coffee and bananas are grown extensively in the surrounding country.

Entebbe was founded as a garrison post in 1893. West of the town is Old Entebbe, where the first European missionaries to reach Uganda landed at the end of the 19th century. (E. A. Js.)

ENTELECHY (Gr. *entelecheia*), a term coined by Aristotle for that which realizes or makes actual the otherwise merely potential. The idea expressed by the term is intimately connected with Aristotle's distinction between matter and form (*q.v.*), or the potential and actual. Briefly, he analyzed each thing into the stuff or elements of which it is composed and the form or the order in which they are arranged. The mere stuff or matter is not yet the real thing; it needs what is variously described as a certain form or essence or function to complete it; only it must be remembered that matter and form are never separated; they can only be distinguished. So, for example, in the case of a living organism one may distinguish the mere matter of the organism (as though it were a mere synthesis of inorganic substances) from a certain form or essence or function or inner activity without which it would not really be a living organism at all; and this "soul" or vital function, as it may be called, is what Aristotle in his *De anima* ii, 1, calls the entelechy (or first entelechy) of a living organism. Similarly, in the case of man rational activity is what makes him really a man, as distinguished from a mere animal, and so on.

Leibniz, following Aristotle, called his "monads" (or spiritual substances) entelechies, in virtue of their inner self-determined activity. Later the term was revived by Hans Driesch in connection with his vitalistic biology to denote an internal perfecting principle which, he supposed, exists in all living organisms.

See also Aristotle; Evolution, Organic.

ENTENTE CORDIALE, a phrase sometimes used to describe the unofficial understanding between Great Britain and France in the years 1831–1846, and more commonly to refer to the friendly relationship established between the two countries by a diplomatic agreement of Apr. 8, 1904.

The most important fruit of the "first entente" was Anglo-French co-operation in the establishment of the independent kingdom of Belgium (1831–39). Disrupted by the mutual animosity of Palmerston and Adolphe Thiers in the Egyptian crisis of 1840, good relations were restored by King Louis Philippe's removal of Thiers and the establishment of a conservative London-Paris axis by Aberdeen and François Guizot. The return of Palmerston to the British foreign office in July 1846 was taken by the French government as signifying the end of good relations, and the first entente ended in the affair of the Spanish Marriages. (*See* France: *History.*)

In common historical usage, the term is more frequently applied to the relationship between Great Britain and France from 1904 to 1914. At the turn of the century Anglo-French relations were strained throughout the world. In April 1904 a series of agreements settled their difficulties over Newfoundland fisheries, Siam, Madagascar and the New Hebrides, modified certain colonial boundaries in Africa, and provided that Britain and France should renounce their rights and interest in Morocco and Egypt respectively in favour of the other. This reconciliation of the two rivals profoundly affected the situation in Europe. In 1905 Germany tested the new entente cordiale by challenging the activity of France in Morocco; but Britain stood firmly by France, offering military help if necessary, and Germany suffered a serious diplomatic defeat.

During the crisis, the British and French general staffs began conversations and ultimately worked out plans which would be put into effect if Britain should decide to support France in war against Germany. This was the second stage of the entente.

In 1911 Germany for a second time challenged French policy in Morocco, and again Britain supported France, a member of the government (David Lloyd George) in a speech warning Germany against war. Once more Germany came off second best in the diplomatic duel.

The final stage of the entente came in 1912, after the rising naval power of Germany had forced Britain to transfer most of its ships from the Mediterranean to the North sea. The French government concentrated its navy in the Mediterranean, but asked for British assurance that the French northern coast would be protected in the event of war with Germany. In November Britain promised that if there was danger of war, the two governments should discuss what they would do, and that if they decided to act, the plans of the staffs would be adopted. This left Britain free to decide, and in August 1914 it took its time about deciding; but a moral obligation was created which in the end proved about as effective as a formal alliance.

See R. Guyot, *La première entente cordiale* (1926); Bernadotte E. Schmitt, *Triple Alliance and Triple Entente* (1934). (B. E. S.)

ENTERIC FEVER is a generic term for typhoid and paratyphoid fevers, the symptoms of which are very similar. *See* Typhoid and Paratyphoid Fevers.

ENTERITIS, a medical term for inflammation of the intestines, especially of the small intestines, generally caused by irritants, poisons, pathogenic organisms or unknown factors. *See* also Gastrointestinal Tract, Diseases of; Pig. (Ch. B. C.)

ENTERTAINMENTS (ADMISSIONS) TAX: *see* AMUSEMENT TAX.

ENTHYMEME, in traditional logic, the name of a syllogistic argument which is incompletely stated. (*See* SYLLOGISM; LOGIC, HISTORY OF: *Ancient Logic.*) Any one of the propositions may be omitted, but in general it is that one which is most obvious or most naturally present to the mind. In point of fact the full formal statement of a syllogism is rare, especially in rhetorical language, when the deliberate omission of one of the propositions has a dramatic effect. Thus, the suppression of the conclusion may have the effect of emphasizing the idea which necessarily follows from the premisses. Far more common is the omission of one of the premisses which is either too clear to need statement or such that its omission is desirable. This use of the word differs from Aristotle's original application of it (in his *Prior Analytics,* ii, 27) to a syllogism based on "probabilities or signs"; *i.e.,* on propositions which are generally valid or on particular facts which may be held to justify a general principle or another particular fact.

ENTOMOLOGY is the science that treats of insects. In its zoological aspects it is concerned principally with the taxonomy, morphology, physiology and ecology of insects as animals; in its applied phases, with the importance of insects to the health and welfare of mankind, including the control of noxious species and the encouragement and management of beneficial ones. Insects as organisms, including their anatomy, physiology, habits and classification, are considered in the article INSECT and in other articles on important or well-known insects and on individual insect orders.

This article is divided into the following sections:

I. History of Entomology
  1. Early Greek and Roman Studies
  2. After Pliny
  3. Revival of Learning
  4. 18th Century
  5. Entomology as a Distinct Science
  6. Contributions of Entomology to General Biology
II. Modern Entomology
  1. Taxonomy (Classification)
  2. Morphology
  3. Physiology
  4. Ecology
III. Applied Entomology
  A. Medical and Veterinary Entomology
    1. Development
    2. Arthropods as Vectors of Animal Disease
    3. Control of Arthropod-Borne Diseases
  B. Agricultural and Forest Entomology
    1. Development
    2. Progress in the Control of Plant Pests
  C. Principles of Insect Control
    1. Cultural Methods
    2. Physical and Chemical Methods
    3. Biological Methods
    4. Integrated Control
    5. Legislation

I. HISTORY OF ENTOMOLOGY

Although insects have long been associated with man, the earliest written records concerning them are only a few thousand years old. These relate to beekeeping, silk culture, wax production and similar utilitarian aspects; to their alleged medicinal qualities; to their religious significance; or to their role as pests. A relief in the temple of Neuserre at Abu Sir, in Egypt, which dates from about 2600 B.C., illustrates beekeeping techniques; and a papyrus of about 1500 B.C. gives formulas for preparing insecticides and repellents against lice, fleas and wasps. Other records appear in the Old Testament and in early Arabic, Sanskrit and Chinese works. In the eastern world the foundations of animal classification were laid by Charaka between 500 and 600 B.C., and a classification of insects was provided by Umasvati in the 2nd century A.D. In the west the Greek philosopher Aristotle (384–322 B.C.) organized the knowledge of insects into the beginnings of a science.

1. Early Greek and Roman Studies.—Aristotle, who named these animals *Entoma,* noted that the insect body is divided into three major parts, the head, thorax and abdomen. He clearly distinguished insects from other so-called "exsanguineous" or "bloodless" animals and described much of their natural history. He recognized the presence in insects of an alimentary canal, a heart and senses of sight, smell and hearing. Aristotle distinguished winged from wingless insects, and those with hard front wings (elytra) from those with two or four similar wings. He also observed the two main types of insect mouth parts, biting and sucking. Metamorphosis was observed by Aristotle; he recognized that some insects, such as locusts and cicadas, laid eggs from which new generations arose, and that butterflies came from caterpillars after these had passed through a chrysalis stage. (However, the chrysalis was mistaken for some form of egg, and the real egg, though observed, was not recognized as such.)

Aristotle also fostered the then prevalent notion that most insects arose by some form of spontaneous generation from filth, carrion, soil, hairs, feathers and galls. Nevertheless, within the mass of sometimes fanciful descriptions and, to us, pseudoscientific hearsay on insects, the early Greek naturalists sprinkled some very sound generalizations that laid the groundwork of the modern science of entomology.

Of the Roman schools of philosophers, Pliny the Elder shed further light on the subject of insects, while repeating much that was fable. He maintained that insects breathed and contended that they had "moisture" in place of red blood, which placed them in a group not truly exsanguineous. He also described a number of species not found in Aristotle's works, and thus qualifies as a significant contributor to the entomological knowledge of the time.

2. After Pliny.—From the time of Pliny (A.D. 23 or 24–79) more than 1,000 years of intellectual darkness passed in the west without any recorded advance in entomological knowledge. The spirit of inquiry flagged, stifled by religious tenets and the belief that the knowledge of the ancients was complete and thus fresh inquiry needless. The few works of the middle ages that dealt with natural history, such as the anonymous *Physiologus,* the bestiaries and the *Origines sive Etymologiae* of Isidore of Seville, are uncritical compendiums based on earlier writers and are restricted to insects affecting man or to moralizing homilies and animal symbolism.

3. Revival of Learning.—Early in the 16th century, with the widespread rebirth of learning, fresh writings on biology appeared, and in 1602 the first substantial entomological work, *De Animalibus insectis,* by Ulisse Aldrovandi, was published. His work describes and illustrates many insects for the first time. Various categories are defined and separated by a key, or identification guide. His classification was based mainly upon habitat: major divisions were land insects and aquatic insects, and subdivisions were based on the presence or absence and the number and nature of legs and wings. Aldrovandi's use of the term insect, however, included most arthropods and also echinoderms and worms. About 30 years later appeared Thomas Moffett's rather similar *Insectorum sive minimorum animalium theatrum,* containing collected observations made over a period of years by Edward Wotton, Konrad von Gesner, Thomas Penn and Moffett himself. It yielded new information and illustrated many species for the first time. These works marked the first real progress beyond the stage reached by Aristotle and Pliny.

In the 17th century the widespread use of printing with type aided the dissemination of knowledge, and political stability and religious freedom stimulated great discoveries that had repercussions on entomology. New worlds opened up by the voyagers brought new insects to be described; new realms of thought created by such scholars as Sir Francis Bacon and René Descartes led to new methods of scientific approach. Observation and experimentation became the building blocks of science.

The advent of the microscope led to the greatest immediate advances in entomological knowledge, and the greatest contributors were Jan Swammerdam, Anton van Leeuwenhoek, Marcello Malpighi and Robert Hooke, the so-called "minute anatomists." Hooke's interest was to discover rather than to interpret what could be seen by the use of his microscope; nevertheless he produced figures illustrating hitherto unknown details of insect anatomy, such as the scales on the wings of moths. Malpighi was primarily a biologist; from his study of the anatomy of the silkworm he described the spiracular (breathing) system of insects, silk glands, Malpighian (excretory) tubules, reproductive organs,

the main features of the central nervous system and the development of the wings. Swammerdam's discoveries were equally fundamental, but more familiar. He demonstrated, in his *Biblia naturae*, the presence of the butterfly's wings already formed within a caterpillar about to pupate; he corrected Malpighi's conception of the insect brain and nervous system; he opposed William Harvey's oddly inconsistent interpretation of insect metamorphosis (Harvey's own dictum, *"omne vivum ex ovo"*; that is, "all life from the egg," was denied applicability to the insects by Harvey himself). Swammerdam accurately described and illustrated the life history and anatomy of the cabbage butterfly, honeybee, may fly, dragonfly and many others. His observations on the development of insects led him to classify them in four major divisions, according to the nature of their transformations. Three of these correspond more or less to the modern divisions Apterygota (primitive wingless insects with little or no metamorphosis), Heterometabola (insects with simple metamorphosis and externally developing wings) and Holometabola (insects with a complete metamorphosis). The fourth division (now untenable) contained primarily flies that, like the bluebottle, pupate within the last larval skin. Leeuwenhoek was the first to recognize the occurrence of viviparity (live birth) in aphids (leading 50 years later to Charles Bonnet's proof of parthenogenesis, or reproduction without males); the insect origin of the sweet substance honeydew and of the dye cochineal; and the eggs and larvae of fleas. In 1668 Francesco Redi, by careful experiment in the rearing of blowflies from eggs laid on flesh, finally refuted the Aristotelian belief that "flies" could arise out of the putrefaction of flesh, thus disposing of spontaneous generation as an acceptable tenet.

4. 18th Century.—The outstanding insect biologist of the first half of the 18th century was R. A. F. de Réaumur. When he was 51, and had already attained considerable achievement in other scientific fields, he published the first volume of his *Mémoires pour servir à l'histoire des insectes* (1734). Five additional volumes were published in his lifetime, and though the work was never finished, it forms a milestone in entomological history. Réaumur's Swedish counterpart, Carl de Geer, whose seven similar *Mémoires* were the last works of their kind to appear, devised a system of classification, based on the wings and the mouth parts, which was a great advance and has stood the test of time. This period, characterized by enormous accessions to knowledge of the bionomics of insects, boasted other pioneers also. Among them were: Jan Goedart, Maria Merian, Roesel von Rosenhof, Johann Frisch, Moses Harris, Eleazer Albin, William Gould and Francis Willughby.

Development of Insect Classification.—John Ray introduced a new epoch in entomology, that of the taxonomic classifiers or systematists. Ray's own classification combined the metamorphotic system of Swammerdam with the morphological and ecological systems of earlier authors, and at the lower level he gave a more precise definition of species, using descriptive phrases (*nomina specifica*) in association with names of generic significance. However, these phrases were restricted to 12 words, and hopeless difficulties arose when it became necessary to compare the works of different authors.

These obstacles were gradually overcome by Linnaeus in the various editions of his *Systema naturae* published from 1735 onward. In Linnaeus' earlier editions the class, order, genus and species were proposed as categories and defined. Later, in 1758, when he first applied consistently his binomial nomenclature system to the whole of the animal kingdom, Linnaeus conferred upon entomologists a far greater boon than on any other group of zoologists; the system at that time allowed for the convenient handling of about 2,000 species. It has since been universally accepted and has proved invaluable in the orderly arrangement of the more than 800,000 different species of insects (over three-fourths of all the animals known) now named and described. Another great step forward was Linnaeus' limitation of the class Insecta by strict definition to what is now generally called the phylum Arthropoda, rejecting the quite unrelated animals (echinoderms, worms, etc.) that had till then been classed as insects. This simultaneous simplification of nomenclature and classification provided a new mechanism just when it was most needed to enable the rapidly expanding knowledge of the insect world to be conveniently organized.

The period from 1758 to Charles Darwin's *Origin of Species* (1859) was a golden age in the study of the morphology and taxonomy of insects, for few entomologists yet had any doubts about the fixity of species. Great argument occurred as to the precise definition of the orders of insects and their affinities, the wingless creatures such as lice and fleas, spiders, mites, centipedes and others that formed the Linnaean order "Aptera," proving particularly troublesome. Various systems of classification were elaborated in succession to the alary system (based on wings) of Linnaeus, such as the cibarian system (based on mouth parts) of Linnaeus' disciple Fabricius, the curious circular system of MacLeay (supposed to express affinity and designed to overcome the restrictions of a linear system), the philosophical system of Oken, and so on. But it is to the celebrated French entomologist P. A. Latreille that the credit really belongs for fusing the best elements of all these classification schemes into a natural system. Latreille also expanded the Linnaean systematic categories by means of intermediate groups, such as the family, to contain closely related genera, a taxonomic concept that has proved to be invaluable.

5. Entomology as a Distinct Science.—By the first half of the 19th century entomology had become a science in its own right. Handbooks such as those of W. Kirby and W. Spence, H. Burmeister and J. O. Westwood were published; faunal works dealing comprehensively with the insects of a whole country, such as J. Curtis' *British Entomology* began to appear; treatises on single orders, such as J. W. Meigen's *Classification of the Diptera* and J. Hübner's works on Lepidoptera became available. Entomological societies were founded (Paris, 1832; London, 1833; Stettin, 1837) as well as independent entomological journals. By the middle of the century so many insects had been named and described that the British museum undertook to publish lists of them. Eighty volumes (20,000 pages) appeared, yet these still did not cover the entire field.

6. Contributions of Entomology to General Biology.—The study of such a vast assembly of insects could not fail to increase the doubts already widely current concerning the fixity of species. Darwin's theory of evolution through natural selection, quickly found champions in such entomologists as A. R. Wallace, H. W. Bates, Roland Trimen and others. The evidence they brought forward and the generalizations they made about protective coloration and mimicry provided valuable proof of the reality of evolution. To entomology, genetics owes one of its greatest discoveries, sex linkage, the outcome of breeding experiments with the currant moth, *Abraxas grossulariata*. The study of hybrids, of polymorphism and other forms of variation in insects, notably in the vinegar or so-called fruit fly, *Drosophila melanogaster*, also has led directly to many other important advances in genetics. Insect psychology drew during this period such distinguished workers as Sir John Lubbock, A. H. Forel, E. Wasmann and J. H. C. Fabre and led to the remarkable discoveries of K. von Frisch concerning communication and behaviour of honeybees. Developments in the field of applied entomology have stimulated research and experiment in insect physiology and ecology to such an extent that there are probably few creatures in existence that are better known in respect of the whole of their "natural history" than are, for example, some of the major crop pests and insect vectors of disease. However, both in applied and in pure science, insects still offer limitless fields for exploration.

II. MODERN ENTOMOLOGY

The early science of entomology, as we have seen, was mainly concerned with attempts to classify insects and arrange them in systems according to their structure. The modern science of entomology is firmly grounded in insect taxonomy, morphology, physiology and ecology. Further, these phases of the modern science provide much of the data for applied entomology and contribute the principles from which sound programs of pest control can be developed.

1. Taxonomy (Classification).—This study is concerned

with the classification of insects. It attempts to systematically summarize and give meaning to everything that is known about insects and is thus, by nature, the most comprehensive subdivision of the science. The principal sources of its data, once limited to external structures, include internal anatomy, embryology, cytology, biochemistry and serology, ecological and behavioural characteristics, geographical distribution and other comparative features.

In order to accomplish its basic objectives taxonomy must provide names and means of recognition for all the species of insects, living and fossil, that can be found and characterized. This task is far from complete, although it has been aided by the efforts of thousands of entomologists during the past 200 years. It has been estimated that more than 2,000,000 species of insects remain to be named, and but a small fraction of those already known have been adequately described in their various forms and developmental stages. Properly performed, the descriptive aspect of insect taxonomy is an exacting task requiring the use of a wide variety of techniques and analytical methods that provide an unlimited challenge to the modern student of the subject.

Descriptive taxonomy, although it has its own rewards and compensations, is a means to a greater end, the development of an orderly system of classification that appears to reflect the natural relationships and the presumed evolutionary history of insects. This aspect has attracted the attention of some of the most brilliant entomologists and, as a result, much progress has been made. The species now known are more than enough to indicate the diversity of insect types and to provide data from which a higher classification may be derived. The present classification is under continuous scrutiny and is subject to revision as more data become available and new interpretations are made. As shown earlier, its basic framework was provided by Linnaeus (1758), its first clear outlines by Latreille (1831). At its upper levels (orders) it makes use of fundamental structural and developmental features (wings, mouth parts, metamorphosis, etc.) that broadly adapt insects for various ways of life and that, by their very nature, have long persisted with relatively little change. At the intermediate levels (family) the groups exhibit less broadly adaptive characters and a more restricted way of life.

From the standpoint of applied entomology, the contribution of insect taxonomy is twofold. First, it provides a gigantic filing system for the data, raw and refined, that has been accumulated about insects. Second, by means of the classifications that it provides, generalizations can be made as to the probable habits, importance and means of control for newly discovered pests; leads can be obtained as to the probable home of introduced pests whose natural enemies are being sought; and a sound basis can be provided for plant quarantine procedures and pest control legislation.

2. Morphology.—This branch of entomology is concerned with the structure of insects, external and internal, macroscopic and microscopic. Although it provides the greatest single source of features by which insects can be recognized and classified (insect taxonomy) and contributes to a fundamental understanding of how insects live (insect ecology) and function (insect physiology), insect morphology is a distinct subdivision of the science and has developed its own body of principles and methods.

Like a coin, insect morphology has two sides that are quite different and, at the same time, closely related. One aspect is "classical morphology," which has enjoyed a continuous development from the days of Malpighi (1628–94) and Pierre Lyonnet (1707–89) to the present time. The other aspect might be termed the "new morphology." While the new morphology has its roots in the past, it has been greatly influenced by developments in other sciences and by recent technological innovations.

Classical v. the New Morphology.—Classical morphology has concerned itself primarily with the description and interpretation of the gross external and internal anatomy of insects. The classical morphologist has always been concerned with homologies and phylogenetic relationships. However, this aspect of morphology has probably passed through its "Golden Era"; many of its enduring facts and generalizations are contained in the great textbooks and reference works of A. Berlese (1909–25), C. Schröder (1912–28), H. Weber (1933), R. E. Snodgrass (1935), P. P. Grassé (1949) and G. Grandi (1951). Nevertheless, the facts and concepts of classical morphology are not complete, and valuable additions to this phase of entomology are being made each year. Classical morphology is important because it provides basic information useful to the insect taxonomist, physiologist, ecologist and toxicologist.

The new insect morphology is in its formative stages, and its outlines are just beginning to emerge. Apparently it will deal with the finer levels of insect structure, including the ultramicroscopic study of cells; it tends to be experimental, whereas classical morphology remains primarily descriptive. The new morphology requires the integrated use of a variety of techniques, some of which are derived from the fields of chemistry and physics. Furthermore, its complexion is likely to change drastically from time to time, with the introduction of new techniques. Finally, it should be noted that insect morphology is merging with the main stream of contemporary research on the cell and cell interrelationships. Therefore, major contributions to the morphology of insects are likely to be of general interest to other biologists. Classical morphology tends to emphasize the uniqueness of insects; the new morphology frequently deals with problems that are peculiar to insects, but, more often, it deals with basic problems of structure and organization that insects share with other organisms.

Other Morphological Disciplines.—The following types of investigations may be included in the new insect morphology: histological and cytological studies, insect histochemistry, electron microscope studies and experimental morphology.

Contributions have been made to descriptive insect histology and cytology ever since the second half of the 19th century, but information on the subject has accumulated so haphazardly that the subject has never been monographed in a definitive way.

Histochemistry is a borderline discipline that can best be described as a wedding of histological technique and analytical chemistry. Histochemistry contributes both to physiology and microanatomy because it makes intracellular localizations of chemical substances possible.

Electron microscopy, in a very short period of time, has made numerous important contributions to an understanding of cellular structures. Among the most exciting contributions are those that deal with the ultramicroscopic structure of mitochondria, Golgi apparatus and other cytoplasmic structures. These ultrastructure studies demonstrate the basic similarity of insect cells to cells of other animals.

Experimental morphology includes a diversity of approaches to the problem of interpreting form and function. Examples include regeneration experiments, the extirpation and transplanting of endocrine glands to determine the effects of such operations on a system and other activities in the borderline area between insect morphology and insect physiology.

3. Physiology.—In its modern phase insect physiology involves primarily experimental studies of the activity of the living insect body and its organ systems and tissues. Its aim is to reveal the dynamic complexity of individual insects, which, in turn, determines both structure (morphology) and interaction with the environment (ecology). Comparatively little work was done before 1900, but by 1950 several hundred investigators and many research centres were devoted to the study, the most prominent centre being that of V. B. Wigglesworth at Cambridge, England. Since 1945, studies in insect physiology have been stimulated by interest in the toxic action of synthetic organic insecticides.

Insect physiology is perhaps more complex than vertebrate physiology. The small size of insects often makes experimentation technically difficult. The methods and apparatus required are usually on an ultramicroscopic scale. Fewer than 20 insect species have been extensively studied. These were selected for large size (*e.g.*, the cockroach *Periplaneta*), ease of rearing in the laboratory (*e.g.*, the mealworm *Tenebrio*), economic importance to man (*e.g.*, the honeybee *Apis*, the silkworm *Bombyx* and various mosquitoes), or exceptional suitability for certain experiments (*e.g.*, the vinegar fly *Drosophila*, in genetics; the housefly *Musca*, in toxicology; the bug *Rhodnius*, in endocrinology).

The present picture of the activity of living insects is an imperfect patchwork of mechanisms assembled from studies of many different species. Only three special areas have been intensively studied: (1) susceptibility to insecticides, and physiological mechanisms involved (enzyme inhibition, detoxication). (Some of this work is useful in insect control.) (2) The chitinous exoskeleton and ecdysis (molting), metamorphosis and hormonal control. (The intrinsic fascination of this phenomenon has attracted many brilliant investigators.) (3) Nutrition. Their size, rapid growth, and ease of rearing make many insect species ideal experimental animals, and knowledge of their dietary needs is now exceptionally thorough. (The infinitely varied dietary adaptations and rapid proliferation of insects makes them an incessant danger to man and his food supply, and knowledge of nutrition and growth can be applied to many entomological problems.)

Abnormal physiology is the primary concern of the field of insect pathology, which, in its applied aspects, contributes to the biological control of insects through the use of microbial organisms (*see* below).

4. Ecology.—This study is concerned with the environmental relations of insects, including their habits and behavioural adaptations. Just as insect taxonomy draws on broad areas of knowledge to place insects in a comprehensive system of classification, insect ecology utilizes data from physiology, genetics, biochemistry, morphology, psychology, pathology, geography, geology, paleontology, meteorology, climatology and soil science to place insects relative to their environment. Because of these interrelationships, most investigations of living organisms have their ecological aspects. Ecological research is specifically directed toward the solution of problems of environmental relationships, abundance and distribution.

The insect ecologist, like the physiologist, is concerned with function. The physiologist investigates function at the level of the cell, organ, tissue or individual, whereas the ecologist is concerned with individuals, groups, local populations, communities and ecosystems (more or less stable groups of interacting communities). In addition, insect physiology is primarily a laboratory science, insect ecology primarily a field science.

Studies of the relationships of individuals or particular species to their environments are usually classified as autecology, those restricted to behaviour as ethology and those concerned with distribution as biogeography. Investigations involving food chains, predator-prey relations, parasite-host relations, population growth and natural control can be categorized as biodemography or population ecology. Studies of still higher levels of organization, such as the community and the ecosystem, are referred to as synecology and biocoenology, respectively. These branches of ecology are not distinct units, as they contribute to one another and all have a common denominator—the study of population dynamics.

The methods of insect ecology are diverse and depend in large part upon the complexity of the population, community or ecosystem under investigation. One of the interesting points in the analysis of environmental interaction is that while the environment of a population can be considered as the sum of the environments of the individuals, the environment of any particular individual includes the other members of the population, and thus the environment of the population includes the population itself.

Every insect has certain basic requisites—food, shelter, space and physical conditions—for its survival and reproduction. The supply of these requisites varies greatly in quality and quantity and in space and time. Within areas where species can exist, the abundance and pattern of distribution is determined by the supply of requisites, by other animals competing for these same requisites, and by natural enemies (parasites, predators and pathogens). The insect ecologist determines the nature of these relationships, measures them, and analyzes their short-term and long-term effects.

The great variety and diversity of insects, their small body size, ability to fly, high reproductive potential and significance to man make them valuable and interesting subjects for ecological research. In recent years, there has been a renewed interest in insect ecology because it has become clear that if man is to manipulate and control insect populations to his advantage he must understand how they are controlled in nature. Thus economic entomology in the modern sense becomes applied insect ecology.

III. APPLIED ENTOMOLOGY

Applied entomology is concerned with insects in relation to man, his domestic and game animals, his crops, his manufactured products and possessions of all kinds, and with the means by which the activities of insects can be curbed or increased for man's benefit. Traditionally, the scope of applied entomology has been broadened to include not only the insects that affect man and his interests but also the economically important terrestrial arthropods, in particular, mites, ticks and other arachnids; isopod crustaceans; and centipedes, millepedes and symphylids. It is convenient to consider applied entomology under two major headings: medical and veterinary entomology, and agricultural and forest entomology.

A. Medical and Veterinary Entomology

1. Development.—Medical entomology is that branch of applied entomology concerned with the insects and other arthropods harmful to the health of man, either by their direct action or by virtue of their ability to transmit the causative agents of many animal diseases. It is vitally concerned with the public health and welfare. The World Health organization has estimated that 50% of all human deaths are caused by arthropod-borne diseases. Typhus, plague, malaria, yellow fever, sleeping sickness and filariasis are among these; without the vectors to transmit them, these diseases would cease to exist.

Veterinary entomology is the closely related field concerned with the role of arthropods in the health of domestic animals and, by extension, of game animals also. It is largely an agricultural science, but the groups of organisms involved draw little distinction between man and other animals; thus it is convenient to consider the medical and veterinary aspects together.

Early man, like many primitive peoples today, probably showed little concern for most of the flying, buzzing and crawling insects that civilized man today regards as intolerable. In medieval times even parasitic forms such as lice and fleas were accepted without complaint; in fact, the ability to tolerate them was sometimes associated with piety. The fact that fleas were involved in the plague epidemics of the 17th century was unrecognized, and some of the means then adopted for the control of the disease no doubt contributed to its dissemination instead.

The first demonstration of an insect-borne human disease was provided by Patrick Manson in 1878. He proved that filariasis, caused by the parasitic worm *Wuchereria bancrofti*, was carried by a mosquito, *Culex*. When the mosquito bites a person infected with filariasis, the tiny worms are drawn into the gut where they continue to develop; after a period of incubation the worms may be transmitted to another person at a subsequent feeding. Manson also suggested that malaria might be mosquito-borne, and in 1898 Ronald Ross demonstrated these parasites (*Plasmodium*) in *Anopheles* mosquitoes. By the turn of the century it had been established that yellow fever, plague, sleeping sickness, filariasis and Texas cattle fever were carried by arthropods, and medical entomology was beginning to assume its role as a major phase of applied entomological science.

The problems of medical entomology are most numerous in tropical regions and in undeveloped areas. In the tropics favourable conditions for breeding and for arthropod activity usually exist throughout the year. In temperate regions lessened activity during the winter months reduces the number of potential insect-borne diseases because of the inability of some of the parasites to maintain themselves without continuous cyclical transmission. Although the incidence of insect-borne disease may vary seasonally, many insects that are human or domestic parasites (lice, fleas, bedbugs) are able to maintain themselves in microclimates that are to a considerable extent independent of the outside climate; insects so adapted are able to continue their activity the year round.

The incidence of insect-borne diseases is greatest in those areas that offer the greatest opportunity for contact between man and an insect vector. Even in tropical regions, where man is exposed to insects throughout the year, the disease problem is most

serious where the population is poorly housed and where sanitation is lacking.

World War II experience demonstrated that insect-borne diseases can be minimized, even in the worst areas; but in the long run the greatest hope for decreasing the incidence of insect-borne diseases involves raising the standard of living in the world's underdeveloped areas in order to take fullest advantage of the knowledge and techniques of medical entomology. During World War II in certain tropical areas where major diseases were effectively controlled, more troops suffered from chronic septic sores following insect bites than from any other disease.

Individual reaction to bites is complicated. When first exposed to a new biting species there is generally little reaction, but soon sensitization occurs, with results ranging from persistent local irritation (stimulating scratching, skin damage, and the ingress of bacteria) to edema and even tissue death (necrosis). Further exposure to bites usually yields, sooner or later, a relatively insensitive state, so that those habituated to an area may be hardly aware of bites (their immunity to attack is more apparent than real). There is evidence that in at least some instances arthropods infected with a disease agent (*e.g.*, the mite *Trombicula*, with scrub typhus) cause a more severe local reaction than uninfected arthropods.

Although bee and wasp stings are painful to most persons, beekeepers usually become tolerant of them. However, a few handlers become hypersensitive and may react fatally, not to the pain-producing substances in the venom, but to some body protein of the insect. Similarly, sensitivity to insect bites is not a simple reaction to digestive enzymes in the saliva, but to some more generally distributed and specific protein substance. The bite of certain spiders (*e.g.*, the black widow, *Latrodectus mactans*) may be fatal. In this case the venom acts directly on the victim and does not depend on any reaction due to an acquired sensitivity.

Dermatitis is frequently caused by arthropods. The bristles of many lepidopterous caterpillars rubbed onto the skin cause severe discomfort; no previous sensitization is necessary to obtain the full reaction. Exposure to insect and arachnid pests of stored products may cause sensitization in a minority of handlers after a period of years.

Some parasites of man have a relatively permanent and intimate association. Ectoparasites such as lice cause irritating infestations in addition to any possible pathogens that they may carry. Endoparasites range from the mite *Sarcoptes*, which lives within the skin and causes scabies, to the maggots of flesh-eating flies, which destroy tissue and cause myiasis.

2. Arthropods as Vectors of Animal Disease.—When an arthropod carries a disease, transmission may be mechanical or biological (cyclical). Thus typhoid bacilli may be carried mechanically from excrement to food on the feet of houseflies, or the trypanosomes (blood parasites) causing sleeping sickness on the mouth parts of a *Stomoxys* fly. Where pathogenic organisms are resistant (*e.g.*, bacterial spores or the eggs of some worms), accidental mechanical transmission is possible over a longer period.

Biological transmission occurs when an important part of the life history of the parasite takes place in the body of the vector. The simplest case concerns worms like *Wuchereria*, which metamorphose but do not multiply. Other parasites, such as the rickettsiae of typhus, reproduce inside the host, but no morphological changes have been recognized. In still others, such as the plasmodia of malaria, the parasite has a complex process including sexual reproduction and the production of infective forms quite unlike those that entered the mosquito when it ingested its blood meal. As a rule, where biological transmission occurs, the parasite can develop only in some specific host (or a small number of hosts). When a *Culex* mosquito ingests blood containing malarial parasites, the latter are destroyed and no transmission takes place. Even among the usual vectors, *Anopheles*, not all individuals exposed to infection become carriers, for the physiological state of the vector and the climatic conditions to which it is exposed have important effects.

Hereditary transmission of pathogens has not been demonstrated in insects, but it does occur regularly in arachnids (*e.g.*, the rickettsia of scrub typhus in the mite *Trombicula*, the spirochete of relapsing fever in the tick *Ornithodorus*), where the organism is passed via the egg through many generations. Certain pathogenic organisms, particularly viruses, previously thought to be mechanically transmitted, have been shown to have a closer relationship to the vector than earlier workers realized, and thus the distinction between biological and mechanical transmission is not so clear-cut as had been imagined.

For a list of the more important diseases carried by insects and other arthropods, see the articles INSECT; PARASITOLOGY. For a discussion of particular arthropod-borne diseases consult them by title, as MALARIA; YELLOW FEVER; TYPHUS FEVER; KALA AZAR; SLEEPING SICKNESS; TYPHOID AND PARATYPHOID FEVERS; PLAGUE; ROCKY MOUNTAIN SPOTTED FEVER AND OTHER SPOTTED FEVERS; etc.

3. Control of Arthropod-Borne Diseases.—Arthropod-borne diseases involve man, the insect or other vector and the pathogenic organism. As a result, they may be combated in several ways. The vector has generally been regarded as the most accessible, if not the weakest, link in the chain. During and after World War II, powerful new synthetic insecticides offered great hope for vector control or elimination. However, widespread resistance to these chemicals has appeared among arthropod vectors, and their control no longer appears so easy as was at first imagined. As a second method of attack, the pathogenic organism can be treated by prophylactic drugs, for example, quinine, atabrine, paludrine (chlorguanide), pyrimethamine and chloroquine in the case of malaria. Finally, a barrier placed between man and the vector will either prevent the insects from becoming infected or prevent infected insects from passing on the pathogenic organisms. As emphasized above, a rise in the standard of living automatically does much to eliminate or minimize arthropod-borne diseases. On the other hand, in underdeveloped areas the diseases reduce efficiency and tend to prevent the rise in the standard of living that would contribute to disease reduction.

B. Agricultural and Forest Entomology

1. Development.—Agricultural entomology is that branch of applied entomology concerned with the relations of insects to man's sources of food, fibre and other products derived from plants and animals. For convenience, those aspects concerned with the production and preservation of wood and wood products are classed as forest entomology. The two fields, although united in principles and related in much the same way as medical and veterinary entomology, have certain basic differences that characterize them.

Although the penetration of civilized man into new areas of the world and the raising of living standards in underdeveloped areas tends to minimize the problems of medical entomology, this is not necessarily true of agricultural entomology. In temperate as well as in tropical regions the introduction of modern methods of agriculture has usually increased the need for insect control measures. The United States is a case in point. The Indians of eastern North America practiced a limited agriculture involving relatively few crops based on native plants of the western hemisphere; these crops included beans, squashes, pumpkins, Indian corn (maize), melons, potatoes, tobacco, sunflower, cotton, etc. Although these crops were attacked by native insect pests that no doubt caused damage, it is likely that the scattered nature of the agriculture, the low standards of quality and the use of primitive hand-picking control methods minimized these losses, except for periodic outbreaks of such general feeding pests as grasshoppers and locusts.

The European immigrants adopted most of the native Indian crops and, as they pushed westward or southward, grew them wherever the environment would permit. The native insect pests naturally spread with the crops. At the same time, encroaching civilization reduced the numbers of insect-eating birds or drove them from the farm areas. The corn earworm followed maize wherever it was grown, the Colorado potato beetle spread east-

ward and occupied all of the potato-growing areas east of the Rocky mountains and the cotton boll weevil invaded Texas and the cotton-growing areas of the south.

The early settlers also brought with them plant crops from the old world. Most of these plants came from temperate regions and had close relatives in the new world. The new world plants were hosts for numerous insects that were able to transfer to their old world counterparts, for example, most kinds of deciduous fruits and nuts. This provided a source of specific native pests for introduced crops. Further, the settlers inadvertently brought with them agricultural pests from the old world. These included not only such well-known pests as the horn fly, gypsy moth, asparagus beetle, codling moth, etc., but also species that, though minor pests in their native land, became major pests in America, having left behind their own specific parasites, predators or diseases. The re-establishment of the natural control of such species is a major objective of the biological method of insect control (*see* below).

The speed with which a newly arrived foreign pest can spread under the conditions of modern agriculture was graphically illustrated by the dispersion of the spotted alfalfa aphid. An old world species, it first became established in the southwestern U.S., invading southern California in 1954, and within four years it had spread throughout the length of that state, inflicting over $35,000,000 in direct crop damage and in costs of control.

In addition to the mechanical damage they wreak, many plant pests serve as vectors of serious plant diseases: Dutch elm disease is spread by certain bark beetles; fire blight of apples and pears in part by bees; and many viral diseases by aphids, leafhoppers and others. (*See* also PLANT DISEASES.)

2. Progress in the Control of Plant Pests.—As man brought his ingenuity to bear on the problem of controlling insect pests and began to achieve some success, higher standards in agricultural products were demanded. Laws were enacted to penalize growers, shippers or processors of foodstuffs contaminated with insects or insect parts and to confiscate such foodstuffs. The housewife demanded higher standards in frozen fruits and vegetables than she had insisted on in the fresh product. Thus, previously ignored or minor pests of crops and processed foods provided another major source of insect problems affecting agricultural production.

Finally, the very methods most widely adopted to combat agricultural pests (insecticidal applications), have in some instances been the source of new problems or have aggravated some of the old ones. As mentioned earlier many insect pests have developed or are developing resistance not only to specific insecticides but in some cases to many, and are now more difficult to control than previously. The specificity of modern insecticides sometimes brings about outbreaks of a secondary species following the treatment for a primary pest. The unfortunate elimination of a pests' natural enemies during treatment may result in a rapid resurgence of the pest, requiring repeated treatments in order to accomplish control. Chemicals toxic to insects are often toxic to man and other animals and therefore must be used in such a manner as to provide no hazard to human health, to domestic animals or wildlife or to honeybees and other pollinating insects.

Thus, the impact of civilization and modern technology has greatly increased the problems of agricultural entomology. Nevertheless, tremendous progress has been made in developing ways and means of contending with insect pests (*see* below).

Forest entomology is most highly developed in the northern hemisphere, especially in Scandinavia, Germany, the U.S.S.R., Canada and the United States. Its methods have differed historically from those of agricultural entomology because the greater time necessary for the crop to mature and the lower annual financial yield per acre have dictated low-cost protective measures. As a result, less expensive indirect control methods have generally been sought, especially those which alter the environment so that it becomes unfavourable to the injurious species. This has necessitated an ecological approach and an understanding of the dynamics of pest populations. Emphasis, until recently, has largely been placed on silvicultural and biological methods of control. Since World War II, the availability of synthetic insecticides and improvement of aerial methods of application have stimulated the use of chemical control, especially against defoliators. Wood preservative technology has become highly developed, and a number of useful chemicals are available for the protection of forest products.

C. Principles of Insect Control

In insular areas or continental areas with pests newly established, total destruction (eradication) of a particular pest is sometimes feasible, as has twice been done with incipient infestations of the Mediterranean fruit fly in Florida. However, under most circumstances the objective of insect control is to reduce the population of a pest species to a level where it is no longer causing significant economic loss or adversely affecting health or welfare. This point varies widely from crop to crop and from place to place depending upon complicated economic and social factors. In any event a thorough knowledge of the ecology of the pest and associated species is necessary if a sound control program is to be developed.

Control procedures fall into a number of categories; for convenience these will be discussed under cultural, chemical and physical, biological, legislative and integrated methods.

1. Cultural Methods.—Cultural methods usually involve an alteration in existing agricultural or forest practices, and have the advantage of being preventive rather than remedial in effect. By employing such measures as crop rotation, special types of cultivation, destruction of crop residues and alternative host plants, pruning or thinning and adjustment of the time of planting or harvesting, it is often possible to avoid or minimize damage to crops by insect pests. The successful planning of such measures depends upon detailed study of the habits of the pest and the range of its food plants, and of the factors that affect both. Insect injury may be reduced by improving the health of the crop plant, as by the use of fertilizers. Plant tolerance of, or resistance to, a pest in some instances is an inherited characteristic, and selection or breeding of resistant varieties is an extremely important method of indirect control, especially of crops in which the low return per acre precludes the use of chemicals.

Noteworthy examples of cultural methods of insect control include the following. In North America losses from the destructive western pine beetle (*Dendroctonus brevicomis*) have been greatly reduced by selective logging of high-risk trees. In Europe early and late plantings of cereals are utilized to control the Hessian fly, whose greatest numbers occur in mid-season. Attack by the wheat bulb fly (*Hylemyia coarctata*) may be avoided if the crop does not follow a summer fallow or early potatoes, since the eggs are laid in bare soil. In Asia flooding of stubbles helps to control rice borers by drowning the resting stages, and in America flooding of cranberry bogs and asparagus fields is practiced for control of pests of these crops. Manuring with nitrogen facilitates the recovery of tea from attacks of the shot-hole beetle in Ceylon, and phosphatic manures appear to hasten the maturing of barley ears to a condition that renders them less liable to attack by gout fly. The application of nitrogenous fertilizers to peach trees in the spring tends to reduce infestation by the shot-hole borer (*Scolytus rugulosus*), a bark beetle, in America. The use of American grape rootstocks resistant to *Phylloxera* has long been practiced; similarly, apples grown on certain Northern Spy stocks are not liable to attack by the wooly aphid. Efforts to produce varieties of maize, wheat, and alfalfa that are resistant to the attack of various insects in the United States have yielded encouraging results, while in south and east Africa the selection of hairy varieties of cotton, which are resistant to attack by leafhoppers, has been an outstanding success.

2. Physical and Chemical Methods.—Physical control methods include the exclusion of pests by banding, screening and other types of barriers, the use of traps and mass collecting devices, hand destruction and certain of the cultural procedures mentioned above. The production of sterile males by irradiation (X-rays and gamma rays from radioactive bodies) shows promise as a means of insect control. It is hoped that by reducing the

number of fertile eggs the population of a particular insect pest may be reduced below a certain critical number and that the population will thus continue to decline naturally.

Chemical methods are the most widely utilized means of the control of insect pests that exceed the economic-injury level or threaten health or welfare. The scientific study of insecticides and their modes of action is the province of insect toxicology. Chemical control agents (insecticides, acaricides, etc.) are readily available in most agricultural and urban areas, directions for their use are generally specific, and machinery for their application, when required, can be rented or purchased, or the services of a commercial applicator contracted. Chemical repellents may be used to keep noxious or disease-bearing insects away from man or domesticated animals and to a lesser extent from various agricultural products. Chemical attractants have been developed to lure insects to poisonous baits and into traps.

The use of chemical control methods in a given case generally is based upon the assumption that the cost of the chemical treatment is less than the crop damage or injury that would have occurred without it. However, in some situations the utilization of chemicals may have far-reaching undesirable effects that are not always immediately obvious. As a result, most responsible persons recommend their application only when seriously needed, at the threshold of economic injury, and deplore their use on a routine basis as a kind of crop or health insurance.

Types of Insecticides.—The older insecticides were generally classed as stomach poisons, contact poisons or fumigants, depending upon their principal mode of action. Stomach poisons were frequently applied as sprays or dusts to the plants on which a particular insect fed; they were also incorporated in baits scattered over the infested areas. The oldest and most widely used stomach poisons were compounds of arsenic, including sodium arsenite, paris green, lead arsenate and calcium arsenate. A group of fluorine compounds, including sodium fluoride, sodium fluoaluminate (cryolite) and others, was much used, and also compounds of mercury and zinc. Although all these compounds had the advantage of retaining their toxicity indefinitely, their use was restricted because of their danger to man and animals and because they could affect only chewing insects.

Contact poisons have been useful for controlling sucking insects, for killing adult flies, mosquitoes and similar flying species and for destroying insect eggs and pupae. Nicotine sulfate and some oils with volatile constituents are among the older contact poisons. The plant-derived insecticides are an extremely important group that includes derris (the chief poisonous constituent of which is rotenone) and pyrethrum (which contains pyrethrins and cinerins). Pyrethrins are practically nontoxic to warm-blooded animals but have a very rapid "knock-down" effect on many insects; under normal conditions they are volatile and break down under the influence of light, and their effect is thus fugitive. They can be stabilized, however, and their effect intensified (synergized) by the addition of piperonyl butoxide and other compounds. The use of such synergists, which, although having only slight insecticidal properties, are able to greatly intensify those of other compounds, is of growing importance.

Fumigants are used to kill insects in granaries, warehouses, ships' holds, greenhouses and in the soil, and to treat plants and commodities under tents. Fumigation of grain is usually preceded by reducing the air pressure, in order to assist penetration by the fumigant. Hydrogen cyanide is still used for the control of insects on citrus trees, in greenhouses and in storage enclosures. Mixtures of carbon disulfide, ethylene oxide, or ethylene dichloride with carbon tetrachloride furnish nonflammable fumigants for the treatment of stored products; methyl bromide is also used for this purpose. Naphthalene and paradichlorobenzene are used for protecting clothing and museum collections against pests. D-D mixture is a liquid fumigant that is injected into the soil to kill wireworms before crops are planted.

The new synthetic insecticides, which have revolutionized the control of insects since World War II, cannot be so readily classified according to their mode of action. The compound now familiarly known as DDT was the first of a rapidly expanding series of synthetic insecticidal materials. Its special qualities include an extremely high contact toxicity to many kinds of insects, chemical stability and low volatility; DDT thus can be used at low concentrations to give deposits with long-lasting insecticidal effectiveness. DDT is used to control a wide range of insects attacking crops. Many other chlorinated hydrocarbons have been discovered, some of which, like toxaphene and dieldrin, have properties resembling those of DDT, and are used for similar purposes, whereas others have been put to special uses. BHC is used to control root-feeding pests, locusts and grasshoppers; the crude product is a mixture of isomers, some of which impart a taint to the crop treated, but the purified gamma isomer (lindane) is practically odourless. Aldrin has similar uses, and chlordane is of special value in controlling ants.

Even more potent than the chlorinated hydrocarbons is the group of organic phosphorus (organophosphorus) compounds that includes among many others malathion, parathion, TEPP, schradan and systox. The latter two compounds, called systemic insecticides, have the valuable property of being absorbed into the sap and translocated when applied to either the leaf or root; the treated plant may thus be rendered toxic for several weeks. The systemic insecticides are of particular value against sap feeders such as aphids or coccids, and can aid in preventing the spread of virus infections by these insects. As a group, the organophosphorus insecticides have an extreme toxicity for warm-blooded animals; they are dangerous to apply but decompose rapidly enough within the plant to leave no toxic residues at harvest. Another class of insecticides, the carbamate esters, are increasing in importance because of their low toxicity for mammals.

The widespread use of insecticides has resulted in the development of resistant strains of some pests. These have arisen by natural selection of genetically controlled characters that enable the individuals carrying them to withstand treatment; the mechanisms involved are still obscure, but involve enzyme inhibition and ability to detoxify the insecticide. Resistance develops where successive generations of a substantial part of the population are treated.

Application of Insecticides.—Insecticides are applied to the surfaces of plants as sprays or dusts. The insect comes into contact with them while moving about or ingests them while feeding. Dusts, which consist of the insecticidal chemical mixed with inert materials, or carriers (*e.g.*, kaolin, pyrophyllite or talc), are compounded by milling the ingredients together or by spraying the dissolved toxic material onto the carrier. Dusts are applied in an air blast generated by a hand-duster or a power-driven fan, and their physical properties must be adjusted so that they do not settle too rapidly or drift too far. Sprays are usually solutions of the insecticide, emulsified with water, or water-stable suspensions of particles of the insecticidal substance; wetting, spreading or adhesive agents are usually added.

The older types of insecticides were applied at low concentrations in considerable quantities of water, often amounting to several hundred gallons per acre. To supply and cart water in such quantities was expensive and, particularly in drier and less-developed countries, impracticable; dusts then had to be used. After 1945 more powerful toxicants were developed; these are used at relatively high concentrations in small quantities of water, in conjunction with fine-aperture nozzles that emit enough small droplets to give adequate coverage.

Application equipment ranges in size from knapsack sprayers or dusters to power-operated, tractor-mounted or tractor-trailer machines. Aircraft are used for dusting crops and plantations, and helicopters have been found especially useful for applying low-volume sprays to crops, particularly in rough terrain. *See* also Sprays and Dusts in Agriculture.

3. Biological Methods.—The use of specific parasites, predators or pathogens for the control of pests involves the application of principles of biological control. Biological control is a phase of natural control, the combination of environmental factors that maintains the fluctuating populations characteristic of all species. Pest populations cause economic injury only when natural control fails to maintain the population below certain levels. Parasites,

predators or disease may be utilized like chemicals, to meet a local situation, or they may be introduced into the general environment to fill deficiencies in the existing natural control. Most modern efforts in applied biological control have involved the latter method.

The most spectacular successes of applied biological control have involved introduced pests. Most introduced pests leave all or many of their natural enemies behind, and the subsequent introduction of these enemies sometimes permanently lowers the level of the pest population below the point at which the pests are economically important. More commonly, however, the response of the biological control agents to fluctuations in the density of the pest population is such that the pest is still capable of causing periodic or local damage, in much the way that a native pest with its indigenous natural enemies may do. In both instances the biotic agents become a permanent part of the agricultural or urban environment and cognizance must be taken of this fact if alternative control methods are utilized.

A classic example of applied biological control is the destruction of the introduced cottony-cushion scale (*Icerya purchasi*) of citrus in California by the Australian ladybird beetle (*Rodolia cardinalis*). Other examples include the control of the sugar-cane leafhopper (*Perkinsiella saccharicida*) in Hawaii by a bug (*Cyrtorhinus mundulus*) that sucks its eggs; of the small white butterfly (*Pieris rapae*) on cruciferous crops in Australia and New Zealand by hymenopterous parasites obtained from the butterfly's native Europe; and of the coffee mealy bug (*Planococcus kenyae*) in Kenya by hymenopterous parasites (*Anagyrus* species) obtained from Uganda, whence the mealy bug is thought to have originally come.

Examples of insects that have become pests in their native countries and have been controlled by imported parasites include a beetle (*Promecotheca reichei*) and a moth (*Levuana iridescens*) attacking coconut in Fiji and controlled, respectively, by a chalcid wasp from Java and a tachinid fly from Malaya. In large continental areas with less sharply defined environmental regions, biological control has not yet proved so successful. The reasons for this are complex and not well understood. In Canada the European wheat-stem sawfly (*Cephus cinctus*) appears to have been suppressed by hymenopterous parasites introduced from England. But although numerous insect enemies of the gypsy moth, European corn borer and Japanese beetle have been introduced into the United States, results have been less striking.

The term "microbial control" is applied to that phase of biological control concerned with the role of microorganisms in the reduction and regulation of the number of insects in a particular area or in a given population. It is an applied phase of insect pathology. The microorganisms involved are, in general, pathogenic only for insects, and thus have the advantage of not leaving residues that are toxic to man and animals. Potentially, representatives from each of the major groups of microorganisms may be involved: bacteria, fungi, protozoa and viruses. Nematodes (roundworms) pathogenic for insects may also be used. As yet there are only a few instances in which these pathogens have been used to control insects, but the field is receiving much attention, and it is likely that increased use will be made of them both as microbial insecticides and as introduced natural enemies of insect pests.

A spore-forming bacterium, *Bacillus thuringiensis*, is an example of an insect pathogen that, used as a microbial insecticide, is capable of effectively controlling certain butterfly and moth pests; and *Bacillus popilliae* has been introduced in certain areas of the United States to control the Japanese beetle. A polyhedrosis virus has been used as an insecticide to control the alfalfa caterpillar (the larva of the orange sulfur butterfly, *Colias philodice eurytheme*) and another polyhedrosis virus has been introduced from Europe into Canadian populations of the European pine sawfly (*Neodiprion sertifer*), bringing about effective control.

4. Integrated Control.—Chemical and biological control methods have been largely developed as separate procedures. However, they are not necessarily alternative methods. They can complement or augment one another. Since the control of pest populations is a complex ecological problem, the success of any method involves an understanding of the ecosystem—the living and nonliving environment of which the pest species is a part. In most agricultural or forest ecosystems, populations of potentially harmful organisms are held at subeconomic levels by natural controlling forces. Other pests may be under satisfactory biological control in most areas for most of the time but periodically reach the level of economic injury in limited places or at certain times of the year. The careful integration of chemical control in such situations may permit retention of the existing biological control by reducing the population below the economic level and holding it until such time as the natural control can again become effective. In some instances this can be accomplished through the use of selective insecticides that shift the balance in favour of the natural enemies. In other cases proper timing of the insecticide application can produce a selective action.

In any event, integrated control requires an intimate knowledge of the pests and their natural enemies, a means of sampling the pest populations and the knowledge to predict those situations in which chemical control is necessary. Such procedures are being successfully applied to a number of orchard and field crops, and have been particularly effective in the supervised control programs employed by alfalfa and cotton growers in California.

5. Legislation.—In most countries laws have been enacted to control the spread of insect pests. These regulations are designed not only to prevent the introduction of pests from foreign countries but to eradicate or retard the spread of a native species that evades local quarantines and to permit the use of government funds for the suppression of pests that breed in areas remote from those where they cause economic losses. In some countries related legislation prevents the adulteration or misbranding of insecticides.

In the U.S. registration is required for pest-control chemicals that move in interstate commerce or are offered for sale in certain states. Further, the U.S. Food and Drug administration establishes tolerances for the various insecticidal chemicals, specifying maximum amounts of these materials that are permitted in the form of residues on foods and food products.

Since the regulation of insect pests is often a matter of world interest, many plant quarantine agreements have been drawn up between countries. Several international agencies, among them the Food and Agriculture Organization of the United Nations, have instituted projects aimed at the control of crop pests. *See* also PLANT QUARANTINE; FOOD AND AGRICULTURE ORGANIZATION.

BIBLIOGRAPHY.—*General:* A. Berlese, *Gli Insetti* (1909–25); D. J. Borror and D. W. De Long, *An Introduction to the Study of Insects* (1954); J. H. Comstock, *An Introduction to Entomology* (1940); G. Grandi, *Introduzione allo studio dell' Entomologia* (1951); P.-P. Grassé, "*Insectes*" in vol. 9 and 10 of *Traité de Zoologie* (1949–51); A. D. Imms, *A General Textbook of Entomology* (1957); Alexander B. Klots and Elsie B. Klots, *Living Insects of the World* (1959); H. Oldroyd, *Insects and Their World* (1960); H. H. Ross, *A Textbook of Entomology* (1956); C. Schröder, *Handbuch der Entomologie* (1928); H. Weber, *Lehrbuch der Entomologie* (1933).

History: F. S. Bodenheimer, *Materialen zur Geschichte der Entomologie bis Linné* (1928–29); E. O. Essig, *A History of Entomology* (1931); L. O. Howard, *A History of Applied Entomology* (1930); H. F. Wilson and M. H. Doner, *The Historical Development of Insect Classification* (1937).

Insect Ecology, Morphology, Physiology and Taxonomy: H. G. Andrewartha and L. C. Birch, *The Distribution and Abundance of Animals* (1954); G. P. Baerends, "Ethological Studies of Insect Behaviour," *Ann. Rev. Ent.*, 4:207–234 (1959); C. T. Brues, A. L. Melander and F. M. Carpenter, *Classification of Insects* (1954); C. Elton, *Animal Ecology* (1951); C. Elton, *The Ecology of Invasions by Animals and Plants* (1958); G. F. Ferris, *The Principles of Systematic Entomology* (1928); O. A. Johannsen and F. H. Butt, *Embryology of Insects and Arthropods* (1941); A. Macfayden, *Animal Ecology: its Aims and Methods* (1957); R. Matheson, *Medical Entomology* (1950); R. E. Snodgrass, *Principles of Insect Morphology* (1935); K. D. Roeder, *Insect Physiology* (1953); E. A. Steinhaus, *Insect Microbiology* (1946); E. A. Steinhaus, *Principles of Insect Pathology* (1949); M. J. D. White, "Cytogenetics and Systematic Entomology," *Ann. Rev. Ent.*, 2:71–90 (1957); V. B. Wigglesworth, *Principles of Insect Physiology* (1956); also the Annual Review of Entomology (1956 *et seq.*), serial publications of local and national entomological societies, such as the Royal Entomological Society of London and the Entomological Society of America, and periodicals dealing with special

subjects, such as the *Journal of Insect Physiology* (1957 *et seq.*) and the *Journal of Insect Pathology* (1959 *et seq.*).

Applied Entomology: J. R. Busvine, *Insects and Hygiene* (1951); A. S. Chandler, *Introduction to Parasitology* (1949); W. Ebeling, *Subtropical Fruit Pests* (1959); H. T. Fernald and H. H. Shepard, *Applied Entomology* (1955); S. A. Graham, *Principles of Forest Entomology* (1952); W. B. Herms, *Medical Entomology* (1950); H. Martin, *The Scientific Principles of Crop Protection* (1959); C. L. Metcalf and W. P. Flint, *Destructive and Useful Insects* (1951); E. A. Parkin, "Stored Product Entomology," *Ann. Rev. Ent.*, 1:223–240 (1956); W. S. Patton and F. W. Cragg, *A Textbook of Medical Entomology* (1913); T. V. Ramakrisna Ayyar, *Handbook of Economic Entomology for South India* (1940); J. Smart, *A Handbook for the Identification of Insects of Medical Importance* (1948); U.S. Department of Agriculture, *Insects: The Yearbook of Agriculture* (1952). Serial publications include *Review of Applied Entomology* (1913 *et seq.*); *Chemical Abstracts* (1907 *et seq.*); *Journal of Economic Entomology* (1908 *et seq.*); *Bulletin of Entomological Research* (1910 *et seq.*); United Nations Food and Agriculture Organization, *Plant Protection Bulletin* (1952 *et seq.*).

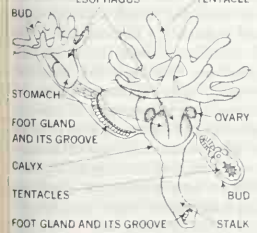
Insect Control: R. E. Balch, "Control of Forest Insects," *Ann. Rev. Ent.*, 3:449–468 (1958); J. L. Brann, Jr., "Apparatus for Application of Insecticides," *Ann. Rev. Ent.*, 1:241–260 (1956); A. W. A. Brown, *Insect Control by Chemicals* (1951), "Spread of Insecticide Resistance," *Adv. Pest Control Res.*, 2:351–414 (1959); C. P. Clausen, "Biological Control of Insect Pests in the Continental United States," *U.S. Department of Agriculture Tech. Bull. 1139* (1956); D. E. H. Frear, *Chemistry of the Pesticides* (1955); W. M. Hoskins and H. T. Gordon, "Arthropod Resistance to Chemicals," *Ann. Rev. Ent.*, 1:89–148 (1956); A. W. Lindquist, "New Ways to Control Insects," *Pest Control* vol. 29, no. 6 (1961); A. Mallis, *Handbook of Pest Control* (1945); W. A. McCubbin, *The Plant Quarantine Problem* (1954); R. L. Metcalf, *Organic Insecticides* (1955) and "Carbamate Insecticides," *Agr. Chem.*, vol. 16, no. 6 (1961); R. H. Painter, *Insect Resistance in Crop Plants* (1951); H. H. Shepard, *The Chemistry and Action of Insecticides* (1951); M. E. Solomon, "Insect Population Balance and Chemical Control of Pests," *Chem. and Indus.*, pp. 1143–47 (1953); E. A. Steinhaus, "Potentialities for Microbial Control of Insects," *Agr. Food Chem.*, 4:676–680 (1956); V. M. Stern, R. F. Smith, R. van den Bosch and K. S. Hagen, "The Integrated Control Concept," *Hilgardia*, 29: 81–101 (1959); H. L. Sweetman, *The Principles of Biological Control* (1958); Y. Tanada, "Microbial Control of Insect Pests," *Ann. Rev. Ent.*, 4:277–302 (1959); F. E. Weick and G. A. Roth, "Aerial Application of Insecticides," *Ann. Rev. Ent.*, 2:297–318 (1957). Several articles on insect control may be found in the U.S. Department of Agriculture's *Insects: the Yearbook of Agriculture* (1952). The agricultural departments of most countries and many of their subdivisions publish information on the control of injurious insects. (E. G. L.)

ENTOPROCTA is the name of a group of tiny, sessile aquatic invertebrate animals less than one centimetre tall and of no known economic importance. Some are solitary, others form colonies.

Authentic fossil entoprocts are unknown, but species of the recent epoch are widely distributed. Only two species live in fresh waters; the rest are marine or brackish forms. They inhabit all oceans and seas from the poles to the tropics, from shore level to depths of 900 ft.

Structure and Function.—Entoproct individuals are called zooids or polyps. The body consists of a stalk bearing a tentacle-crowned head, the calyx, on a fleshy arm, called the lophophore. The calyx is soft; the stalk soft or hard, sometimes chitin-reinforced, depending upon age and species.

The basal end of the stalk attaches to some object, such as an alga, rock, shell or body of another sessile or slow-moving sea animal. The association is not parasitic or obligatory but apparently accidental or commensal.

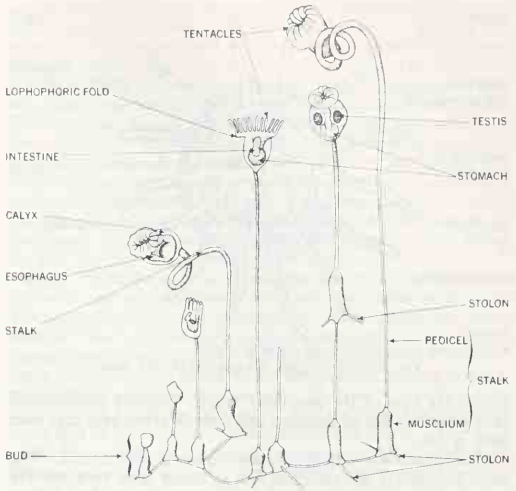


AFTER ATKINS. FROM "MCGRAW-HILL ENCYCLOPEDIA OF SCIENCE AND TECHNOLOGY," 1960. REPRODUCED BY PERMISSION OF THE PUBLISHERS

FIG. 1.—LOXOSOMA SINGULARE ZOOID WITH TWO BUDS

The free end of the stalk either expands very gradually (as in *Loxosoma*) or enlarges very suddenly (as in *Barentsia*) to form the tentacled calyx.

The calyx contains the organs of digestion, excretion and reproduction as well as the main ganglia and nerves. The pseudocoel, or space between the calyx body wall and the organs, contains a very diffuse tissue of mesenchymelike cells. Entoprocts



AFTER ROGICK. FROM "MCGRAW-HILL ENCYCLOPEDIA OF SCIENCE AND TECHNOLOGY," 1960. REPRODUCED BY PERMISSION OF THE PUBLISHERS

FIG. 2.—BARENTSIA LAXA COLONY: ZOOIDS OF VARIED AGE AND SIZE, SOME WITH EXPANDED TENTACLES, OTHERS WITH TENTACLES CURVED INTO CALYX.

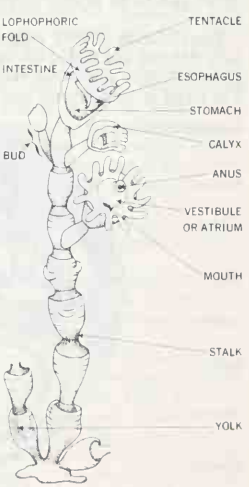
of the families Pedicellinidae and Urnatellidae may shed or lose their calyces in times of adversity. The stalks, however, may regenerate new calyces.

Eight to 36 ciliated tentacles, depending upon the species, adorn the head and encircle the mouth and anus. The tentacles aid in attracting food, repelling unwanted objects and dispelling fecal wastes and larvae. They are not retractile into the body cavity but merely curve inward over the mouth and anus. The gut is U-shaped. Above and about the stomach are located the main ganglia, the excretory protonephridia and the testes, ovaries, gonoducts and brood chamber.

Individuals reproduce sexually and by budding. Zooids may be protandric (alternate the sex), hermaphroditic (contain organs of both sexes) or dioecious (contain organs or one sex, male or female). Sometimes whole colonies are of one sex. Fertilization is internal. Ciliated larvae develop in a brood chamber. When released, the larvae swim about for a brief time before metamorphosis and attachment to the substratum.

Buds originate from different areas of the body wall, depending upon the species. In the family Loxosomatidae new zooids bud from the calyces. In Pedicellinidae new zooids generally bud from stolons, partitioned tubes of varying length and diameter that sprout from the stalk bases and that anchor the stalks to the substratum. In Urnatellidae, and sometimes in Pedicellinidae, new zooids bud from the stalk proper.

Smooth muscle fibres in the stalk, calyx, lophophore and tentacles enable a zooid to nod its head, wave its stalk, flick or curl its tentacles and tighten the lophophoric sphincter that sur-



AFTER ROGICK. FROM "MCGRAW-HILL ENCYCLOPEDIA OF SCIENCE AND TECHNOLOGY," 1960. REPRODUCED BY PERMISSION OF THE PUBLISHERS

FIG. 3.—URNATELLA GRACILIS COLONY, CONSISTING OF A FLEXIBLE STALK WITH URNLIKE SEGMENTS

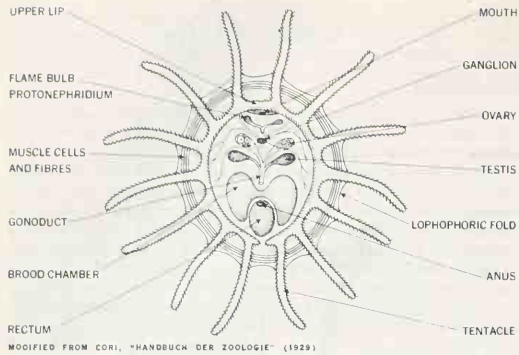


FIG. 4.—PEDICELLINA CERNUA CALYX, TOP VIEW

rounds the base of the tentacular crown. Solitary loxosomatids can change location by creeping, and their detached buds may even swim a bit.

Classification.—Entoproct classification is troublesome. The name Entoprocta did not come into existence until 1869, but the animals that were so designated had been known long before. Between 1774 and 1847 they were known as zoophytes because of their tentacled polyps. Some workers put entoprocts into the Polyzoa of J. V. Thompson (1830) or the Bryozoa of C. G. Ehrenberg (1831). The terms Polyzoa and Bryozoa have become synonymous; although Polyzoa has priority, it is too vaguely defined to be acceptable to all specialists.

In 1869 H. Nitsche divided the Bryozoa into Entoprocta and Ectoprocta, on the basis of the location of the anus with respect to the tentacular crown. In Entoprocta both anus and mouth are located close together within the tentacular circle. In Ectoprocta the anus is outside the tentacular circle. Further study led B. Hatschek to elevate Entoprocta to phylum rank (1888), but the rank did not gain universal favour.

In 1921 A. H. Clark renamed the entoprocts "Calyssozoa," then C. Cori (1929) renamed them "Kamptozoa"; but these new names also failed to gain support because they seemed unnecessary.

Ernst Marcus (1939) argued that entoprocts and ectoprocts were Bryozoa and should not be placed in separate phyla. L. H. Hyman (1951) considered entoprocts entitled to phylum rank. The views of both Hyman and Marcus have adherents.

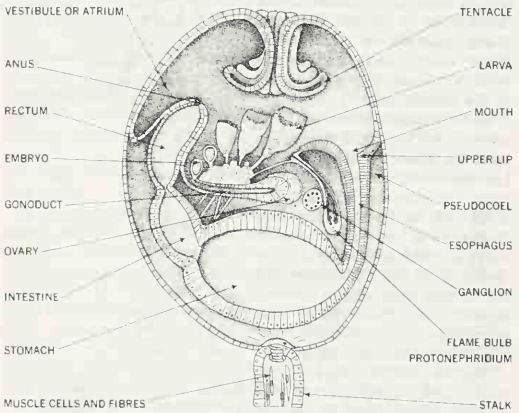


FIG. 5.—PEDICELLINA CERNUA CALYX, WITH TENTACLES INCURVED, INTERIOR OF RIGHT SIDE. LARVAE AND EMBRYOS ARE ATTACHED TO EMBRYOPHORIC EPIDERMIS

Entoprocta comprise three families, 12 genera and about 90 species. The families differ mainly in budding methods, growth habit (solitary or colonial), stalk anatomy and the individual zooid's ability to shed its calyx.

Family Loxosomatidae contains four genera (*Loxocalyx, Loxomespilon, Loxosoma, Loxosomella*) and about 58 solitary marine species. Family Pedicellinidae contains seven genera (*Barentsia, Chitaspis, Coriella, Loxosomatoides, Myosoma, Pedicellina, Pseudopedicellina*) and about 30 colonial species, either marine or brackish water. Family Urnatellidae contains one genus, *Urnatella*, and two colonial, freshwater species.

See also BRYOZOA.

BIBLIOGRAPHY.—L. H. Hyman, *The Invertebrates*, vol. 3, pp. 521–554 (1951); E. Marcus, "Briozoarios marinhos brasileiros, III," *Univ. São Paulo Bol. Faculd. Filosof. Ciênc. e Letr, XIII, Zoologia*, 3:111–354 (1939); M. Prenant and G. Bobin, "Bryozoaires, Première Part," *Faune de France*, ed. by P. Lechevalier, 60:1–118, 336–397 (1956); M. Rogick, "Studies on Marine Bryozoa, II," *Biol. Bull.* 94, 2:128–145 (1948). (M. D. Ro.)

ENTRECASTEAUX, ANTOINE RAYMOND JOSEPH DE BRUNI, CHEVALIER D' (1739–1793), French navigator who contributed much to knowledge of the Australian islands, was born at the Château d'Entrecasteaux, Var, and entered the navy at the age of 15. In the war of 1778 he commanded a 32-gun frigate and successfully convoyed a fleet of merchant vessels from Marseilles to the Levant, though they were attacked by two pirate vessels, each of which was stronger than his own. In 1785 he was put in command of the French fleet in the East Indies, and in 1787 he was made governor of the Mascarene Islands. While in command of the East India fleet he made a voyage to China, and in 1791 he was given command of an expedition to search for J. F. La Pérouse, lost since 1788. The expedition comprised the "Recherche" and the "Espérance." No tidings were obtained of La Pérouse, but in the course of his search Entrecasteaux made important geographical discoveries. He traced the outlines of the west coast of New Caledonia and touched at several places on the south coast of New Holland. Visiting Tasmania, he surveyed and named D'Entrecasteaux channel, the entrances to the Huon and Derwent rivers, Bruny Island and other places. The ships then sailed for the East Indies, and near Java Entrecasteaux died of scurvy on July 20, 1793.

See E. P. E. de Rossel, *Voyage de Dentrecasteaux* (1808).

ENTRE DOURO E MINHO (popularly called MINHO), the old province of Northwest Portugal, lies between the Minho and the Douro rivers. It comprises the administrative districts of Viana do Castelo, Braga and Oporto, and after 1936 was divided into the provinces of Minho and Douro Litoral (*qq.v.*). Pop. (1950) 2,053,079. Area 3,137 sq.mi. This region is the historic unit, first developed in the 9th century A.D., from which Portugal evolved. It was referred to in 841 as "Provincia Portucalense" and in 883 as "Portugal." Since the 16th century it has maintained almost constantly about one-fifth of the total population of Portugal and is proverbially the most intensely cultivated. (J. M. Ho.)

ENTREPRENEUR, a term used in economics to denote one who assumes the risk and management of a business. The factors of production were generally divided by the earlier economists into land, labour and capital; these were identified with the economic classes—landlord, worker and capitalist. The capitalist was the person who provided capital for the business and also directed the business. For these activities the capitalist received interest and profits. Profits covered many different types of activity, among them incomes received by management of special ability, net incomes from the ownership of businesses and returns from assuming the non-insurable risks of business undertakings. It also included the rewards, if successful, for starting some new venture or bringing into the market a new process or product. It is in this last sense that the capitalist becomes a true entrepreneur.

The entrepreneur is thought of as being the dynamic element in the capitalist order in his role as an innovator; he makes his entrepreneurial profits by seizing upon an opportunity to secure a temporary monopoly for himself by successfully changing mar-

ket products and processes. However, in an attempt to share the entrepreneur's good fortune, other businessmen will follow his lead, and his profits will tend to diminish in the face of competition. Thus the entrepreneur imparts to the capitalist process the dynamic quality needed for its continued progress. (L. N.)

ENTRE RÍOS, a province of eastern Argentina, is separated from Uruguay by the Uruguay river. Area 29,427 sq.mi. Pop. (1960) 803,505. As its name implies, it is located between two rivers, the Paraná and the Uruguay, and it is sometimes referred to as Argentine Mesopotamia. The surface is an undulating plain, interrupted in the north and along the eastern and western margins by low hilly ridges. The province tapers in the south into the Paraná deltaic lands of the northwest of the Plate estuary. Marked differences in the appearance of the landscape are provided by the vegetation, which varies from forests through park country to extensive grasslands. Rainfall is usually well distributed throughout the year and the summers are hot and humid.

Some settlements in the province were founded in early colonial times, Paraná being established about 1588, but the area did not receive provincial status until 1814. During the dictatorship of Juan Manuel de Rosas (from 1829 to 1852) the province became a centre of opposition based on antagonism to his control of river trade. Gen. Justo José de Urquiza, one of its leaders, contrived Rosas' overthrow. The new Argentine confederation, established in 1853, had its capital at Paraná, in Entre Ríos, which Urquiza governed dictatorially. Dissension in 1861 between Urquiza and his provincial rival, Santiago Derqui, led to Urquiza's downfall (and assassination by another Entre Ríos *caudillo* López Jordán) and the transfer of the capital to Buenos Aires once more. In the second half of the 19th century the province became a great receiving area for immigrants, which led to rapid agricultural development and economic expansion. (*See* also ARGENTINA: *History.*)

A quarter of the cultivated land is devoted to wheat, and other important products are corn, oats, alfalfa, flax (for linseed), grapes, barley and peanuts. Its pastoral significance is, however, still considerable and the earliest meat preserving and packing plants in Argentina were established in the province. Refrigeration plants, breweries, distilleries, flour mills and consumer goods industries are numerous. The capital, Paraná, and the other principal urban centres are located on or near the two rivers, Paraná and Uruguay, the limits of ocean navigation being Paraná and Concepción del Uruguay respectively. Gualeguay on the river of the same name, the most important tributary of the Paraná in the province, is a pastoral centre with an airport. Its port, Puerto Ruiz, is accessible by coastal vessels.

Although rivers provide the most important means of communication, Entre Ríos has several thousand miles of roads, built chiefly between 1928 and 1949, and 1,400 mi. of railways, making it easily accessible from the rest of the Argentine *pampa.*

(G. J. B.)

ENUGU, the capital of Eastern region, Nigeria, lies 151 mi. N.N.E. of Port Harcourt by rail. The municipality comprising Iva Valley, Ogbete, Ogui, Asata and Uwani had an estimated African population of 82,594 in 1960. The town owes its existence to the discovery of coal in 1909. With the completion of the railway from Port Harcourt in 1916, it developed rapidly, becoming a township in 1917 and a municipality in 1956. The principal streets are High street, Ogui road and Abakaliki road. Except for the Old Secretariat (1929) and the Governor's lodge, the most important buildings date from 1950. They include the house of assembly, government offices, the College of Technology (1955), national archives and regional library buildings (1959), municipal centre, African Continental bank, a government women's training college, Queen's school and the College of the Immaculate Conception. In 1960 the new University of Nigeria was opened at Nsukka, 30 mi. N. Trunk roads link Enugu with other regional centres. The railway from Port Harcourt links the town also with the north and west. The airport at Emene, 7 mi. E., is served by internal air services. The coal industry employs more than 8,000 workers and annual production exceeds 800,000 tons. There is a railway workshop, a cement plant at Nkalagu, 33 mi. distant, and an electricity supply from the Oji river power station. Other industries include modern pottery and a sawmill. Coal is the only important export, and imports include motor vehicles, textiles and building materials. (JA. C. M.)

ENURESIS is a medical term that usually means bed-wetting, but it has been used as a label for any apparently involuntary urination after an age when control should have been achieved. Some authorities think this should be accomplished by age three, but many children do not achieve it until five or six. About 5% of all children wet the bed occasionally after age five, and the incidence decreases until puberty. A few cases persist into adult life. Boys are affected twice as often as girls. A small percentage wet during the day as well as at night, and a few wet only during the day.

About 5% of the cases of bed-wetting are due to anatomical defect or disease or are deliberate while appearing to be involuntary. The cause of the other 95% is obscure. Some believe that these cases of "true" enuresis are caused by a persistence of the physiological incontinence of infancy due to inadequate or poor training. Some believe that it is caused by a retarded maturation or developmental defect of the nervous system. Others believe that enuresis is a result of psychological factors, that the child has emotional needs that he cannot satisfy during the waking hours; the enuresis might be a substitute for sexual gratification, an expression of deep-seated anxiety and fears or a disguised expression of hostility toward a parent which cannot be expressed openly.

The treatment of enuresis also is a matter of controversy. A great variety of methods, drugs and manipulations have been used, all with some success, and yet no single method has been entirely successful. Treatment should begin with a physical and emotional evaluation and a study of the environment. Whether emotional factors are primary or the result of the enuresis, the enuretic needs psychological help. Proper training methods can help prevent enuresis. No advantage is gained by beginning training for nocturnal bladder control before the age of two. The discomfort of being wet aids the development of internal inhibitions. Taking the child to the toilet at intervals at night may keep him dry, but it also trains him to urinate during sleep. (J. F. KD.)

ENVER PASHA (1881–1922), a controversial figure in Turkish politics of the period between the Young Turks' revolution and the collapse of the Ottoman empire at the end of World War I, was born on Nov. 23, 1881, in Istanbul, the son of Ahmet Bey, a member of the Sultan Abdul-Hamid II's retinue who later became a pasha. Leaving the military academy as a staff officer with the rank of cavalry captain, Enver served in the III army corps in Salonika, at that time a chief centre of revolutionary activity against the sultan's regime. Eventually he himself joined the Young Turkish secret society *Ittihad ve Terraki* ("Union and Progress") and, in 1903, went also to lead a Turkish *bashi-bazouk* band (irregular mounted troops) operating in Macedonia against Bulgarian and Greek guerrilla forces. He was one of the organizers of the revolution of 1908 and accompanied Mahmud Shevket (*q.v.*) in his march on Istanbul in April 1909, which led to the deposition of Abdul-Hamid II. After this he was appointed military attaché in Berlin, where he was impressed by German military might.

Enver left Berlin in 1911, when the Italo-Turkish War broke out, to organize the Ottoman resistance; and he held out in Cyrenaica till the armistice (his account of his African campaign was published in German as *Um Tripolis*, 1918).

Enver could not get home from Africa for the first of the Balkan Wars of 1912–13, but was in Istanbul in time to lead the *coup d'état* of Jan. 23, 1913, that put the Young Turks in power. He was then appointed chief of the general staff of the Turkish army; and in June a German general, Otto Liman von Sanders, became head of a military mission to Turkey. On July 22, 1913, Enver recaptured Edirne (Adrianople) from the Bulgars and became a national hero. In Feb. 1914 he was appointed minister of war. Thenceforward, with Talat Pasha and Jemal Pasha, he formed a triumvirate that virtually ruled the country. In March 1914 he married the princess Naciye, a niece of the reigning sultan Mohammed V.

When Turkey entered World War I on the German side in Oct. 1914, Enver became the virtual commander in chief of the army. His military plans, however, were coloured by fantastic Pan-Turanian or Pan-Turkic schemes, aimed at uniting the Turkic peoples of Russian central Asia to the Ottoman Turks. This led him to a disastrous campaign in Dec. 1914 in which he lost a whole Turkish army corps in the region of Sarikamis (between Erzurum and Kars), but he was able to recover some of his prestige when the British had to withdraw from Gallipoli (*see* DARDANELLES CAMPAIGN), though much of the real credit for the Turkish victory belonged to Mustafa Kemal, who became more and more his hated rival.

The Russian Revolution of 1917 and the consequent weakening of Russia enabled Enver to try his eastern plans again. In the autumn of 1918 he succeeded for a time in occupying Baku, but the defeat of Germany brought Turkey down too, and after the armistice at Mudros in October he fled to Germany. There he saw some of Russia's Communist agents, such as Karl Radek. He thought of assisting Soviet Russia against his most hated enemy, Great Britain, by making trouble in India and the middle east; with this in mind he flew to Moscow and interviewed Communist leaders. He also formed the idea of overthrowing Mustafa Kemal's regime in Turkey with Soviet aid, but this plan received no support from Moscow. Though the Russian leaders became suspicious of him, Enver was nevertheless allowed to go to Turkistan with a view to helping to organize the central Asian republics. In 1921, however, the revolt of the *basmachis* in Bukhara against the Soviet regime flared up, and Enver joined the insurgents. He was killed in action against the Red army near Baljuvan, in Tajikistan, on Aug. 4, 1922. (M. P. P.)

ENVOY: *see* AMBASSADOR; DIPLOMACY.

ENYALIUS AND ENYO, the masculine and shortened feminine forms of the name of a primitive Greek war deity. The form Enyalius ("warlike") occurs in Homer's *Iliad* either as a separate name or as an epithet of the god Ares (*q.v.*). Some late traditions made Enyalius Ares' son. The word is an epithet of Ares in the oath taken by youths at Athens. In Argos, Enyalius was worshiped as a separate deity; but in Sparta "Enyalius" was Ares' usual cult title.

Enyo, "destroyer of cities," appears first in the *Iliad,* once in conjunction with Athena, once with Ares. Generally she is much less prominent in literature than Enyalius, and in cult is found only occasionally, in association with him or with Ares. Her name was revived by the Romans as a translation of that of their war-goddess Bellona.

See M. P. Nilsson, *Geschichte der griechischen Religion,* 2nd ed., vol. i, p. 519 (1955). (H. W. Pa.)

ENZELI: *see* PAHLAVI; PAHLAVI LANGUAGE.

ENZINAS, FRANCISCO DE (FRANCIS DRYANDER) (1520-1553), Spanish theologian, who translated the New Testament into his native tongue, was born at Burgos of a wealthy and aristocratic family. He was sent abroad to be educated and seems to have matriculated at Louvain university in 1539. He moved to Wittenberg in 1541 and there began a close association with Philipp Melanchthon, who advised Enzinas to undertake the translation of the New Testament. This was published in 1543 and dedicated to the emperor Charles V, but the emperor's confessor found heretical indications in the work and consequently Enzinas was imprisoned, at Brussels. He escaped to Antwerp and, in 1545, to Wittenberg. He went to England, bearing letters of recommendation from Melanchthon to Thomas Cranmer and the lord protector in 1548, and was given the chair of Greek at Cambridge. Ill-health obliged him to return to the continent in 1550; he died of plague at Strasbourg in 1553.

Enzinas used various translations of his surname (Sp. *encina,* "evergreen oak"); the Greek version, "Dryander," is that by which he was best known in England.

ENZIO (or ENZO; It. for Ger. HEINZ, diminutive of HEINRICH) (*c.* 1220–1272), titular king of Sardinia, was a natural son of the Holy Roman emperor Frederick II. In 1238, he married Adelasia, widow of Ubaldo Visconti, ruler *giudice* of Torres and Gallura, two of the four *giudicati* of Sardinia, despite the veto of Pope Gregory IX, who claimed Sardinia as a papal fief. Enzio assumed the title of king of Torres and Gallura, or king of Sardinia. After his second excommunication in 1239, Frederick appointed Enzio his legate-general for the whole of Italy. Enzio then played a prominent part in the campaigns against the Guelphs in northern Italy. Defeated by the Bolognese near Fossalta on May 26, 1249, he was taken prisoner and remained a captive at Bologna until his death on March 14, 1272. His marriage with Adelasia had been annulled by the pope; in 1249, shortly before his capture, he had married a niece of Ezzelino da Romano. Perhaps Frederick's favourite son, he wrote poems in Italian, as also did Frederick's other natural sons, Frederick of Antioch and Manfred.

BIBLIOGRAPHY.—F. W. Grossmann, *König Enzio* (1883); H. Blasius, *König Enzio* (1884); L. Frati, *La prigionia del re Enzo a Bologna* (1902). (N. R.)

ENZYMES are substances that cause and direct the numerous chemical reactions that occur in living organisms. Many chemical reactions will not proceed at an appreciable rate except in the presence of minute amounts of certain substances called catalysts. Enzymes are the catalysts of living cells. In order to understand the full significance of this definition, it is necessary to grasp the fact that the living cell or the living organism may be regarded as a system of chemical reactions, beautifully integrated and co-ordinated, of extraordinary complexity, but nevertheless amenable to experimental investigation and to some measure of comprehension or understanding. Most of these chemical changes that occur in living cells are caused and controlled by enzymes. The study of the mode of action of the enzymes is therefore one of the principal approaches to the study of what goes on in a living cell.

Examples of Enzyme Action.—Some enzymes are discharged from the cells that form them and therefore function outside the living cells. The secretions of the digestive tract of animals contain many such extracellular enzymes. Thus the enzyme ptyalin, secreted by the cells of the salivary glands, is discharged into the mouth, where it acts on the starch of the food, breaking down the large starch molecules into the simpler sugar, maltose. The cells of the stomach secrete the enzyme pepsin, which acts on proteins to degrade them to polypeptides and amino acids.

The process of degradation is called hydrolysis, since water is used in the chemical reactions involved. Ptyalin hydrolyzes starch and, similarly, pepsin hydrolyzes protein. Trypsin and chymotrypsin, which are present in the secretions of the pancreas gland, also act to hydrolyze protein. All of these enzymes of the digestive tract act to convert the complex molecules of the food into smaller simpler molecules. This degradation, which involves the conversion of solids into water-soluble substances, and of large water-soluble molecules into smaller ones, is the essence of the process of digestion. The secretions of the digestive tract are not the only place where enzymes that hydrolyze proteins are found; they also occur within the cells of animal tissues, and in plants and microorganisms as well. Papain, a hydrolytic enzyme obtained from the latex of certain plants, has been used commercially as a meat tenderizer. The meat thus treated can be regarded as partially predigested.

Mainly because of their relatively easy accessibility, the extracellular enzymes were among the first to be recognized and studied. Most enzymes, however, are not secreted by the cells that form them, but normally remain inside the cell. It is these intracellular enzymes that direct the processes of cellular metabolism. The complex synthetic reactions that require energy are under enzyme control, just as well as the degradative processes that yield energy. In fact, much modern research on life processes consists of the study of enzyme reactions and of the interrelationships between them. (*See* BIOCHEMISTRY.)

It sometimes happens that the enzyme activity of an extract increases on standing. This increased activity may be due to the presence of a proenzyme, or zymogen—an enzyme precursor which, of itself, has no enzyme action. The classical example is pepsinogen, the inactive precursor of pepsin. Pepsinogen is converted to pepsin by exposure to acid or to another proteinase such as trypsin. The acid of the gastric juice can cause formation of the

active enzyme. Pepsin itself also converts pepsinogen to pepsin. The conversion has been shown to consist of the splitting of a peptide fragment from the pepsinogen molecule. Trypsin and chymotrypsin are also secreted originally in the form of inactive precursors, trypsinogen and chymotrypsinogen.

History.—The results of the action of enzymes were observed long before the nature of the responsible agents was recognized. Thus the ability of gastric juice to dissolve chunks of meat was observed by Lazzaro Spallanzani in 1765. In 1785 William Irvine noted that aqueous extracts of sprouted barley caused a liquefaction of starch. In 1833 Anselme Payen and Jean François Persoz made a more detailed study of the process of starch solubilization, or liquefaction, and showed that the responsible agent, which they called diastase, was destroyed or inactivated on boiling, that it acted catalytically (*i.e.*, without undergoing permanent change itself a small amount could cause the conversion of a large amount of starch to sugar) and that it could be concentrated and purified by precipitation with alcohol. This work on diastase is generally regarded as the first clear recognition of an enzyme, though the word enzyme (meaning "in yeast") was first introduced by the German physiologist Wilhelm Kühne in 1878.

Many of the great chemists of the 19th century were interested in the problem of the nature of enzyme action. It was not until 1926, however, that James B. Sumner succeeded in purifying and crystallizing urease, an enzyme that causes a hydrolysis of urea to ammonia and carbon dioxide. After obtaining the urease from jack beans, which are rich in this enzyme, he showed that the pure crystalline enzyme is a protein.

The significance of this finding was not fully realized or acknowledged until other enzymes were also purified and found to be protein in nature. In 1930 John H. Northrop and M. Kunitz succeeded in preparing pure crystalline pepsin and trypsin and in showing that these enzymes are also proteins. After hundreds of enzymes had been purified and no exceptions were found, it became reasonably safe to assume that all enzymes are proteins. In 1946 Sumner and Northrop received the Nobel Prize in Chemistry for their work on enzymes.

The first successful studies of intracellular enzymes were an outgrowth of studies of the process of fermentation. When a yeast cell causes sugar to ferment, *i.e.*, converts it to alcohol, the cell absorbs sugar from the liquid medium outside the cell and secretes the alcohol and carbon dioxide products back into the liquid medium. Under these circumstances, no amount of chemical analysis of the components of the medium can give detailed information about how the cell makes alcohol from sugar. The dawn of a new era of enzyme study began with the demonstration that fermentation can occur in the presence of cell extracts.

Though the fermentation of sugar to alcohol is generally carried out with live yeast, the presence of intact yeast is not necessary, an extract or press juice from the yeast being sufficient, as was first demonstrated in 1897 by the German chemist Eduard Büchner, who won the 1907 Nobel Prize in Chemistry for this discovery. Subsequently the study of extracts provided the most important means of acquiring information about the nature of the chemical reactions that occur in living matter. The extracts of yeast convert sugar to alcohol, just as living yeast does, and in a similar manner extracts of muscle can convert sugar to lactic acid, just as living muscle does.

Studies of such yeast and muscle extracts demonstrated that the processes of alcohol and lactic acid formation are both due to about a dozen different enzymes, each acting in sequence on the successive products of the reaction. The two processes are almost identical except for the last steps in the reaction sequences. The reaction catalyzed by each individual enzyme is simple; but the overall effect is a complex change.

Cellular respiration (the oxidation, within the cell, of a food molecule such as glucose, to give carbon dioxide and water as end products) is another biological process involving a sequence of enzyme-catalyzed reactions. The study of the enzyme systems which bring about respiration went hand in hand with the study of fermentation. The oxidoreductases necessary for these processes were characterized chemically in the 1930s largely through the work of the German cell physiologist Otto Warburg, who thereby laid the foundations for present-day knowledge of the chemical mechanism of enzyme action.

Properties.—Since enzymes are proteins, they exhibit the physical and chemical properties of proteins and are studied with the same tools and techniques as those employed for studying protein molecules. A characteristic property of proteins is their large molecular size. Enzyme molecules range in molecular weight from about 10,000 to several million, and they cannot pass through semipermeable membranes that readily allow the passage of water, sugar and other substances of small molecular size. Undenatured proteins are characterized by their instability to heat. The inactivation of enzymes by heating often begins at about 120° F (50° C); almost all enzymes are completely inactivated when held in boiling water for a few minutes. This sensitivity to heat and the inability to pass through a semipermeable membrane are regarded as almost definitive. That is, if a catalytic action of an extract of biological origin disappears on heating and is retained on dialysis (fails to pass through a semipermeable membrane), the action is regarded as probably being attributable to the action of an enzyme system.

Small changes in acidity or alkalinity often have a large effect on enzymes. Few enzymes can survive treatment with strong acid or alkali. Substances toxic to living organisms also cause the destruction or inactivation of enzymes. Some of these effects are relatively specific; *i.e.*, a particular toxic agent may affect a particular enzyme or group of enzymes. For example, the poison hydrogen cyanide is thought to act mainly by inhibiting the cytochrome oxidase, an enzyme necessary for the normal consumption of oxygen in the respiratory activities of most cells that exhibit aerobic metabolism.

Coenzymes and Vitamins.—Some enzymes contain no organic substances except amino acids as the building blocks of their molecules. The hydrolytic enzymes pepsin and trypsin belong in this category. In other cases, specific organic compounds other than amino acids are firmly bound to the enzyme. These specific compounds are known as prosthetic groups and can be shown to play a part in the operation of the enzyme as a catalyst. Several of the more important prosthetic groups are derivatives of vitamins. For example, the enzyme carboxylase, which splits carbon dioxide from certain organic acids, contains a phosphate of thiamine (vitamin B_1) as a necessary constituent of the molecule. Another vitamin, B_2 or riboflavin, is a constituent of a group of enzymes called flavoproteins, which function in respiratory metabolism to transfer hydrogen and/or electrons from food molecules to oxidants such as molecular oxygen. Metals such as iron and copper also are built into these flavoproteins, whereas the metal magnesium is a constituent of the thiamine-containing carboxylase. Still other enzymes contain a derivative of vitamin B_6, pyridoxine. All enzymes of this group catalyze chemical transformations of amino acids. The different enzymes are specific, however, for different reactions and also for different amino acids. These differences in specificity must be determined by the protein component of the enzyme, since the prosthetic group is the same.

In some instances an enzyme may operate with the aid of an organic substance of low molecular weight that is not attached firmly to the enzyme molecule. Such a substance is called a coenzyme. An example is cozymase (also called coenzyme 1 or diphosphopyridine nucleotide). The vitamin niacin or nicotinamide is a constituent part of cozymase, which acts in conjunction with a group of enzymes termed dehydrogenases. These enzymes function in the oxidation of sugar and of many other compounds important in intermediary metabolism.

Mechanism of Enzyme Action.—Enzymes show a remarkable degree of specificity for the substances on which they act (substrates). A particular enzyme catalyzes only one kind of reaction and acts only on one kind of substance or group of closely related substances. This high degree of specificity is one of the characteristics that sharply differentiates enzymes from other catalysts. The first step in the action of an enzyme is a combination of the substrate molecule(s) with the enzyme at the so-called active site. A chemical change then occurs, and the final step is the dis-

sociation of the product(s) of the reaction from the enzyme. The high specificity of the enzyme is related to the fact that the structurally complex protein can combine in a particular way necessary for the catalytic action only with molecules that fit very particularized structural requirements.

Like other catalysts, enzymes do not appear to undergo any net change during the reactions which they cause. This apparent absence of change is of course due to the fact that the reaction sequence involves a cyclic regeneration of the original enzyme, so that one molecule of enzyme reacts over and over again within a brief space of time, each time with a different molecule of substrate. The activity of a purified enzyme is sometimes expressed as its turnover number, the number of molecules of substrate acted upon by one molecule of enzyme in one minute. Most enzymes have turnover numbers higher than 1,000 and some have turnover numbers larger than 1,000,000. The enzymes choline esterase and catalase have the highest turnover numbers known—about 18,000,000. Catalase, an enzyme present in red blood cells, causes the decomposition of hydrogen peroxide to oxygen. When hydrogen peroxide is mixed with blood a foam is formed due to the rapid release of the oxygen gas in the protein-containing solution.

If an enzyme is present in low concentration relative to the concentration of its substrate—the usual situation—then the enzyme does not contribute to the net energy requirements of the reaction catalyzed. Only those reactions that are energetically possible without the enzyme can occur in its presence. The reaction effected by an enzyme is catalyzed in a reverse as well as in a forward direction, and if the energy change is suitable, the reaction proceeds in both directions simultaneously until an equilibrium is reached.

Enzymes may contain prosthetic groups (nonprotein substances). In some cases it has been possible to determine how the prosthetic group functions in the reaction. In the flavoproteins, for example, the flavin prosthetic group can exist in two forms, a yellow oxidized form and a reduced colourless form. The reduced form contains two more hydrogen atoms than the oxidized form. When the flavoprotein glucose oxidase acts on glucose, two hydrogen atoms are removed from the glucose and added to the flavin prosthetic group, thereby reducing it. The reduced flavin then passes the hydrogen (in the form of hydrogen ions and electrons) on to oxygen, after which the regenerated oxidized prosthetic group can act on another molecule of glucose. Glucose oxidase is used in the dried egg industry to remove glucose, which causes the development of the undesirable odours and flavours of the dried egg product during storage. When glucose loses two hydrogen atoms under the influence of glucose oxidase, gluconic acid is formed. The latter substance has no undesirable effects on the egg powder.

As already mentioned, the enzyme molecule is large—usually much larger than the substrate molecule on which it acts. One might therefore suppose that one enzyme molecule could combine with and act on many substrate molecules simultaneously, but this is not the case. Current data show that there are only one or a few sites on the large enzyme molecule where the reaction with the substrate takes place.

One of the problems of enzymology under intense investigation is the chemical identity of the active sites of the enzyme. If a prosthetic group is present and demonstrably functional, it is clear that the prosthetic group marks the site of action. But if the enzyme contains no prosthetic group, the problem of locating the active site is more difficult. One approach to this problem depends on the use of specific enzyme inhibitors. Many poisons and toxic substances exert their effect on the living organism by combining with and inactivating enzymes which are essential for normal physiological function. Thus the so-called nerve gases, which are compounds of phosphate, combine with an active site of the enzyme choline esterase, the proper function of which is essential for normal transmission of nerve impulses. Because the inhibitor is bound at the active site, the normal substrate is prevented from reaching the enzyme.

Choline esterase contains no prosthetic group, only amino acids, as building blocks of its molecule. The inhibited enzyme can be degraded and the fragments examined to see which amino acids have been attacked by the inhibitor. This approach can give information about the identity of amino acids present at the active site; it does not, however, define the geometrical arrangement of the active site. Since the activity of the enzyme is generally lost when the protein molecule is disorganized, the correct geometrical arrangement of the structural components of the enzyme is obviously essential for enzyme function, but the precise characterization of the geometry of an active site has not been elucidated for any enzyme.

The determination of the complete chemical structure of an enzyme is beset by the same very great difficulties that attend the similar investigation of other proteins. By 1963 the covalent chemical structure of bovine pancreatic ribonuclease had been determined. This enzyme has the unusually low molecular weight of 12,700, and consists of a chain of 124 amino acids looped and folded into a compact structure held together by four disulfide bridges. It is anticipated that success in structural determination will lead in time to the complete chemical synthesis of enzymes; meanwhile, all available enzymes are products of living cells.

Enzymes in Genetics.—In view of the importance of enzymes in intermediary (cell) metabolism, it is not surprising that the enzymes should determine many of the physiological and morphological differences among animals and plants. The genes (heredity units) exert their effect, in many instances, by directing the synthesis of proteins. Since most tissue proteins are enzymes, the synthesis of enzymes in a cell is under the control of the genes of that cell. The relation of genes to enzymes is seen, for example, in the hereditary condition in human beings which is manifested by a sensitivity to milk. The condition is due to the loss of an enzyme necessary for the degradation of galactose, a constituent of lactose, or milk sugar. The enzyme loss results from a mutation (change) in the genetic makeup of the cell.

The genes controlling enzyme synthesis are themselves composed of deoxyribonucleic acid (DNA). The sequence of nucleotides in this substance determines the sequence of amino acids in the enzyme proteins. The manner in which this is accomplished chemically was under active investigation in the mid-1960s, and constituted one of the most rapidly developing areas in the biological sciences (*see* NUCLEIC ACIDS).

Nomenclature and Classification.—Names such as ptyalin, pepsin, trypsin, chymotrypsin, and papain were assigned to enzymes before 1961, the year in which the Commission on Enzymes of the International Union of Biochemistry issued its recommendations for nomenclature and classification. The commission proposed that enzymes be classified and named on the basis of the chemical reaction catalyzed. Six main classes were defined, as follows:

1. Oxidoreductases catalyze oxidation-reduction reactions.
2. Transferases catalyze transfer reactions.
3. Hydrolases catalyze reactions involving cleavage of a molecule with the uptake of water.
4. Lyases catalyze the removal of groups from their substrates, leaving double bonds; or, conversely, add groups to double bonds.
5. Isomerases catalyze isomerizations.
6. Ligases, or synthetases, catalyze the joining together of two molecules coupled with the breakdown of a pyrophosphate bond in adenosine triphosphate or a similar nucleotide.

At the time of the commission's report more than 650 different enzymes were known, and it was anticipated that the number would continue to increase. Many old names are retained, but the recently discovered enzymes have, for the most part, been named according to the above system.

See also DIGESTION; FERMENTATION; NUTRITION; PROTEINS; VITAMINS; and references under "Enzymes" in the Index.

BIBLIOGRAPHY.—A. v. Szent-Györgyi, *On Oxidation, Fermentation, Vitamins, Health and Disease* (1939); O. H. Warburg, *Heavy Metal Prosthetic Groups and Enzyme Action* (1949); J. B. Neilands and P. K. Stumpf, *Outlines of Enzyme Chemistry*, 2nd ed. (1958); P. D. Boyer, H. Lardy, and K. Myrbäck, eds., *The Enzymes*, 2nd ed. rev., 8 vol. (1959–1963); M. Dixon and E. C. Webb, *Enzymes*, 2nd ed. (1964).

A book of review articles, *Advances in Enzymology*, appears annually.

(B. V.)

EOCENE AND PALEOCENE. The Paleocene is the first epoch of the Tertiary (*q.v.*) period in the geologic timetable; it began about 70,000,000 years ago and lasted probably 10,000,000 to 15,000,000 years. The Eocene, the second epoch of the Tertiary period, lasted probably 20,000,000 years. In these comparatively minor segments of geologic time the mammals and flowering plants made substantial progress on the evolutionary road to eventual prominence.

Two conditions favoured the evolution of the warm-blooded animals: the predatory dinosaurs had died out in the Cretaceous (*q.v.*) period, before the Paleocene; and the widespread but locally varied warm climate during the Paleocene brought an abundance of vegetation anticipating the modern flora. "The extraordinary way in which mammals and birds both took advantage of the great opportunity is one of the marvels of geological history." W. B. Scott comments in *A History of Land Mammals of the Western Hemisphere.*

The percentages of living species represented by fossils in each series of strata were the basis of the divisions of the Tertiary period, first proposed by Charles Lyell in his *Principles of Geology* (1833). He called the oldest division Eocene, from the Greek *eos* ("dawn"), *kainos* ("recent"), the fossils of which in his opinion included only 3½% of still existing species—an estimate that was later only slightly modified. This Eocene in the dawn of modern life, he said, might be further subdivided into older, middle and newer parts, as was subsequently done, the oldest part being called Paleocene, from the Greek *palaios* ("ancient"), by W. P. Schimper in 1874, the middle part remaining typical Eocene and the youngest part being called Oligocene, from the Greek *oligos* ("few"), by H. E. Beyrich in 1854. This article deals with the Paleocene and the Eocene, in that order; the succeeding Oligocene is dealt with in a separate article (*see* OLIGOCENE).

The fossiliferous clays, sands, limestones and lignites deposited during the early part of the Tertiary were first studied at their conveniently accessible outcrops in the regions surrounding Paris and London, which thus comprise the type areas. These areas were parts of a broad depression that was repeatedly invaded by a gulf of the ancient North sea and in consequence received a complex series of interfingering marine and fresh-water deposits containing many plants and animals. Adjacent to this body of water to the north and northwest was a land mass that included Scandinavia and the western British Isles and probably Iceland, Greenland and North America. To the south and east was the central European land mass, but there were no Alps, that part of Europe and northern Africa being portions of another large basin with many small proliferations, occupied by Tethys, a great Mediterranean sea that extended eastward across Asia Minor, India, southeastern Asia and the East Indies. Parts of Australia, New Zealand, Argentina, southeastern North America, the Pacific coast of North America and Greenland were sites of embayments similar to that of the Paris-London basin, but large areas in the interior regions of the continents were, with few exceptions, not invaded by seas and received only nonmarine sediments containing remains of land animals and plants.

Geologic Time Chart

System and Period	Series and Epoch	Distinctive Records of Life	1,000 Years
CENOZOIC ERA			
Quaternary	Recent	Modern man	11
	Pleistocene	Early man	1,000
Tertiary	Pliocene	Large carnivores	
	Miocene	Whales, apes, grazing forms	
	Oligocene	Large browsing mammals	
	Eocene	Rise of flowering plants	
	Paleocene	First placental mammals	70,000
MESOZOIC ERA			
Cretaceous		Extinction of dinosaurs	130,000
Jurassic		Dinosaurs' zenith, primitive birds, first small mammals	160,000
Triassic		Appearance of dinosaurs	200,000
PALEOZOIC ERA			
Permian		Reptiles developed, conifers abundant	235,000
Carboniferous			
Upper (Pennsylvanian)		First reptiles, coal forests	260,000
Lower (Mississippian)		Sharks abundant	285,000
Devonian		Amphibians appeared, fishes abundant	320,000
Silurian		Earliest land plants and animals	350,000
Ordovician		First primitive fishes	400,000
Cambrian		Marine invertebrates	500,000
PRECAMBRIAN TIME			
		Few fossils	3,500,000–4,000,000

The following discussion of the Paleocene and Eocene epochs indicates how they were identified through study of marine fossils in European occurrences as distinctive from each other and from other epochs, describes the Paleocene and Eocene in North America and the mammals characteristic of each stage and indicates other occurrences and the physical characteristics of these first two epochs of the Cenozoic era. For discussion of the study of rocks and fossils the reader is referred to the article GEOLOGY. *See* also FOSSIL; PALEOBOTANY; PALEONTOLOGY.

PALEOCENE

Paris-London Basin.—The chalky deposits of the Danian stage in the Paris basin, long considered to be the latest stage of the Cretaceous period of the Mesozoic era, are now considered to be the earliest stage of the Paleocene, containing the nautilus *Hercoglossa danica,* distinctive foraminifers and no rudistids. A regression of the sea permitted erosion of its exposed surface on which, with readvance of the sea, were deposited the conglomerates, chalk and glauconitic sands that characterize the early part of the following Montian stage. Brackish-water marls at Meudon, near Paris, and the chalk at Mons, in Belgium, completed the stage, which, apparently, was not represented by deposits in the London area. The Thanetian stage began with a widespread fossiliferous deposit of glauconitic and calcareous sand, called the Thanet sands in England and the Bracheux sands in France. Limestone marls and algal travertines were simultaneously deposited in fresh-water lakes or sloughs near Rilly and Sézanne, about 60 mi. E. of Paris.

The Thanetian marls of Rilly contain fossil snails, insects, fishes, turtles and crocodiles, and the travertines preserve abundant fossil ferns and flowers, fruits, seeds and leaves of dicotyledons. These deposits and fossils early received attention, the plants in particular being described by Gaston de Saporta in 1868, and those of the same age from the marls and sands at Gelinden, Belg., by Saporta and A. F. Marion in 1873 and 1878. Reviewing this flora in 1874, Schimper noted that it differed significantly from the known Cretaceous and Eocene floras and therefore proposed that this earliest portion of Tertiary time be called Paleocene.

The term Paleocene attracted little attention until Victor Lemoine in 1878 described some mammalian remains from the conglomerate at Cernay, near Reims, and A. von Koenen in 1886 reported a considerable marine fauna from the vicinity of Copenhagen. Thereafter Paleocene strata were identified, chiefly by remains of primitive mammals, in many parts of the world, and the term came into general use. In 1939 it was officially recognized by the United States geological survey.

The Paleocene in North America.—About the time that the Paleocene strata of the Paris-London basin and neighbouring regions were being studied, the geologists of the United States and Canada were exploring the vast coal-bearing terrain of the western interior of North America. This area surrounds what is now the continental divide, which roughly follows the core of the Rocky mountains. In Late Cretaceous time and for a long time thereafter the site of the Rocky mountains was occupied by a chain of relatively low hills with wide flood plains that bordered on embayments of a sea. On the east side of the Rockies, from Canada to New Mexico, the low plains became a long, broad, troughlike, unstable area that intermittently subsided and in the course of millions of years accumulated a thick series of lignitic deposits. In western North Dakota and eastern Montana a general section of these deposits seen on a clear day displays several series of strata contrasted by dark and light colours.

These coloured zones of the Late Cretaceous and Early Tertiary have been named and, in part, mapped. Field experience has shown that the colour feature is unreliable for defining these zones very far beyond their original locale. Nevertheless, to a practised eye

*Generalized Table of Paleocene and Eocene Chronology**

Epoch	European stages	England	France	United States			American vertebrate ages
				Southeastern	Western interior		
Eocene	Ludian	Headon beds	Montmartre gypsum	Jackson	Duchesne river		Duchesnean
	Bartonian	Barton clay	St. Ouen limestone		Uinta		Uintan
	Auversian	Upper Bagshot sands Upper Bracklesham beds	Beauchamp sands	Claiborne	Bridger		Bridgerian
	Lutetian	Middle Bagshot sands Lower Bracklesham beds	Calcaire Grossier		Green river		
	Ypresian	Lower Bagshot, Alum bay sands London clay Oldhaven beds	Cuise sands	Wilcox	Wasatch: Wind river		Wasatchian
	Sparnacian	Woolwich, Reading beds	Soissons clay, lignite		Wasatch: Indian Meadows		
Paleocene	Thanetian	Thanet sands	Rilly chalk, travertine Cernay conglomerate Bracheux sands	Midway	Fort Union: Sentinel Butte	Clark Fork	Clarkforkian
	Montian		Mons chalk Meudon marl		Fort Union: Tongue river	Tiffany	Tiffanian
					Fort Union: Lebo	Torrejon	Torrejonian
	Danian	Chalk	Chalk		Fort Union: Cannonball	Dragon	Dragonian
					Fort Union: Tullock	Puerco	Puercan

*Absolute agreement in cross-correlation is not intended.

the latest Cretaceous strata have a brownish cast that distinguishes them from the smoky-hued shales and coals and the whitish to yellowish sandstones of the overlying Paleocene sequence. Between the lowermost Paleocene zones there are at some localities in North Dakota tongues of sand and clay of brackish, estuarine origin containing oysters and other mollusks. These tongues are continuous eastward into a marine formation deposited in what was called the Cannonball sea, whose further connections, whether with the Gulf of Mexico or the Arctic ocean, are unknown. At many localities where the outcrops of coal beds have burned, the overlying strata have been baked, sometimes fused, forming red clinker, which adds vivid splashes of colour to the picturesque badlands.

At first this sequence of sandstones, shales, clays and coal seams, typically exposed in the badlands, valleys, buttes and plains of Canada, North Dakota, Montana, Wyoming, Colorado, New Mexico and Utah, was called the Great Lignite and, on the basis of abundant fossil plants, was considered to be Tertiary. Presently, dinosaurian remains were discovered in the basal strata, which then became known vaguely as the Laramie formation, tentatively assigned to the Cretaceous; and the overlying strata became the Fort Union formation of the Eocene. Because at most exposures no recognizable unconformity breaks the sequence of Cretaceous-Tertiary beds at and above the critical zone containing the last dinosaurs, the placement of the Mesozoic-Cenozoic boundary in the western interior became the subject of long debate. However, beginning in 1880 mammalian remains began to be taken from variegated clays and sands in the San Juan basin, New Mexico, called the Puerco formation by E. D. Cope, but later separated into the Puerco and Torrejon formations, of Early and Middle Paleocene age, respectively. In 1901 the first mammals were found in the Fort Union formation on the east side of the Crazy mountains in Montana, and subsequent exploitation of those strata yielded a large fauna of Middle Paleocene age. Other localities in the Rocky mountain states added to these remains, thus, together with the fossil plants, producing overwhelming evidence for a substantial flora and fauna significantly different from that of the typical Cretaceous and Eocene. Consequently, the Mesozoic-Cenozoic boundary can be placed with confidence, although somewhat arbitrarily, at the horizon above the last dinosaurian remains and just below the zone marking the resumption of coal deposition after a long, coal-barren, late Cretaceous interval.

Paleocene Mammals.—The striking change from the Cretaceous was the absence of dinosaurs and appearance of such new mammalian lines as primates, taeniodonts, creodonts, condylarths and pantodonts. The multituberculates, marsupials and insectivores continued from the Cretaceous. (*See* also MAMMAL: *Evolution and Relationships.*)

Beginning with the oldest, the nonmarine American Paleocene is subdivided on the basis of the fossil remains of mammals into the Puercan, Dragonian, Torrejonian, Tiffanian and Clarkforkian, each of these referring to the respective area yielding the fauna of that name.

The Puercan refers to the fauna from the dark, purplish or greenish shales and sandstones of the Puerco formation that crops out in the San Juan basin, northwestern New Mexico, particularly in the badland exposures of the dissected escarpment extending southeastward for 30 mi. from a point just east of Ojo Alamo. The fauna is large, consisting of fishes, turtles, crocodiles and archaic mammals among which are distinctive species of the primitive ungulates *Carsioptychus*, *Conacodon*, *Ectoconus* and *Oxyacodon*, the insectivorelike *Onychodectes* and the herbivorous *Taeniolabis*. The Dragonian is the fauna from the variegated clays of the Joes valley member of the North Horn formation in Dragon canyon, Sections 7 and 8, Township 19 South, Range 6 East, in the central plateau of Utah. The fauna contains characteristic species of the ungulates *Anisonchus, Ellipsodon* and *Dracoclaenus.*

The Torrejonian was originally included in the fauna from the Puerco formation, but subsequent study revealed its distinct nature, whereupon the variegated shales and sandstones containing it were called the Torrejon formation. These strata overlie the Puerco formation in the same general region of the San Juan basin, New Mexico. Characteristic mammals of this fauna are species of the primitive insectivore *Deltatherium*, the insectivorelike *Conoryctes* and *Psittacotherium*, the ungulates *Mioclaenus* and *Periptychus* and the herbivore *Ptilodus*.

The Tiffanian fauna is named from the varicoloured shales that crop out near Tiffany station in Section 20, Township 33 North, Range 6 West, in the San Juan basin in southwestern Colorado. It contains characteristic species of the earliest ungulates of large size (amblypods), the eight-foot *Barylambda* and the *Titanoides*, the herbivore *Ectypodus*, a primate, *Labidolemur*, and the carnivorous *Phenacodus*.

The Clarkforkian refers to the fauna taken from the gray shales exposed in the divide between the Clark Fork and Big Horn basins, northwest of Ralston, Wyo. The characteristic mammals are species of three carnivores, *Didymictis*, *Phenacodus* and *Thryptacodon*, the herbivores *Ectypodus* and *Esthonyx* and the lemurlike primate *Plesiadapis*.

Other Occurrences and Characteristics.—The Paleocene of the Gulf embayment of southeastern North America is comprised in the marine sands, clays and limestones of the Midway group of formations whose correlative in the western interior is the Cannonball formation, a basal member of the Fort Union in North Dakota. On the west coast the Martinez formation of California represents a Paleocene embayment.

Exposures at Atanekerdluk, on the west coast of Greenland, yielded abundant fossil plants, which, although originally identified by Oswald Heer as Miocene, were subsequently shown by Saporta to be clearly of Paleocene age.

At least parts of the Sysran strata in the Volga basin, U.S.S.R., are Paleocene. The American Museum of Natural History expedition to Mongolia in 1925 found a formation called Gashato that contains mammals probably of late Paleocene age. The Río Chico formation in southern Argentina is the most probable Paleocene formation in South America.

The marine life of the Paleocene resembled that of the Late Cretaceous but lacked ammonites and rudistids. The nautiloids *Hercoglossa ulrichi* and *H. vaughani* replaced *H. danica*. Among the gastropods *Calyptraphorus* was a newcomer. Foraminifera were numerous but included no conspicuous nummulites.

Like the fauna, the flora also, after losing the Mesozoic cycadeoids, began to assume the beginning of a modern aspect and included ferns, cycads, conifers, palms and dicotyledons. Species of hazelnut (*Corylus*), birch (*Betula*) and viburnum (*Viburnum*) are almost indistinguishable from living species. These, as judged by the requirements of the living species, suggest temperate conditions, but the presence of palms in the Yellowstone valley of Montana and of many broad-leaved dicotyledons in Greenland points to warmer conditions than prevail in those areas today.

Except for the crustal movements that caused oscillations of the strand line in the embayments just discussed, the only other evidences of tectonic activity during the Paleocene are a few thin lenses of conglomerate among the otherwise fine-grained sediments of the western interior, suggesting that sporadic slight uplifts of the Rocky mountain land mass caused rejuvenation at the heads of streams.

The Paleocene epoch amassed a series of 5,000 ft. or more of strata without duplicating synchronous deposits. Some volcanic activity occurred during the epoch, as is shown by bentonitic deposits in Montana and lava flows and dikes in Colorado and New Mexico. The principal economic products of the Paleocene epoch are lignite, clay and bentonite.

EOCENE

The Eocene in Europe.—As with the Paleocene, the account of the Eocene properly begins with the European areas, the northern or Paris-London basin and the southern or Tethyan basin, where the deposits were first studied intensively. The European stages, without multiplying synonyms, are, beginning with the oldest, Sparnacian, Ypresian, Lutetian, Auversian, Bartonian and Ludian.

Conglomerates, sands, clays and lignites comprise the Sparnacian stage in the Paris-London basin. The clays and lignites of Soissons in France are contemporaneous with the estuarine Woolwich and Reading beds in England. This stage, on the assumption that a stage is the succession of deposits during a complete cycle involving advance and regression of the sea, has been regarded widely as latest Paleocene. As the mammals found in the Cernay conglomerate of the previous Thanetian stage point to very late Paleocene age, however, the Sparnacian is most likely the introductory stage of the Eocene and is so accepted by many paleontologists.

A widespread marine invasion, carrying the disk-shaped foraminifers *Nummulites elegans* and *N. planulatus*, inaugurated the Ypresian stage that includes the Cuise sands of the Paris basin, the sandy Oldhaven beds, the famous gray-brown London clay and the Alum bay and Bournemouth leaf beds in the Lower Bagshot sands in the London and Hampshire areas. Although the London clay contains a marine fauna it also preserves, at some localities, remains of land animals and plants. The latter consist principally of carbonized and pyritized fruits, seeds and twigs—a flora whose affinity appears to have been with the living Indo-Malayan tropical rain forest.

The Lutetian stage, which marked the greatest advance of the Eocene sea, is noted for its beds of limestone—an excellent and much-exploited building stone—called the Calcaire Grossier. The weathering of this limestone into forested plateaus dissected by steep-sided valleys accounts for the attractive scenery of the Paris basin, as compared with the almost treeless plains developed on the late Cretaceous chalks and sands that crop out at the perimeter of the basin. The Calcaire Grossier is distinguished by the presence in great numbers of the foraminifer *Nummulites laevigatus*, which may have reached the basin from Tethys by a seaway through the Loire or Rhône valley. The sands and clays of the Middle Bagshot sands and the Lower Bracklesham beds in England likewise contain *N. laevigatus*.

The Auversian stage, containing *Nummulites variolarius*, is represented by the Beauchamp sands in the Paris basin and the Upper Bagshot sands and Upper Bracklesham beds in England. Overlying these in the Bartonian stage are the Barton clay and sands with *N. prestwichi* in England and the St. Ouen fresh-water limestone in the Paris basin. Finally, the Ludian stage, with the Headon brackish- and fresh-water beds in England and the Montmartre gypsum, probably deposited when the Paris basin was temporarily cut off from the sea, completed the Eocene record in the Paris-London embayment. Mammalian remains, including those of the large, three-toed, horselike ungulate *Palaeotherium*, from this gypsum were the basis on which Georges Cuvier founded the study of vertebrate paleontology. Adjacent coastal regions in Denmark, Scandinavia and northern Germany have somewhat similar Eocene strata, and in central Europe occur isolated fresh-water deposits of sands, clays and limestones containing Eocene mammalian remains. Southwest France and northeast Spain had embayments from the Atlantic ocean that received sediments which at some localities, as Biarritz, now form prominent cliffs.

The Tethys sea, a much expanded Mediterranean, with its many embayments in southern Europe, northern Africa, Asia Minor, India and Burma, was the site of deposition of thousands of feet of sediments—limestones in the deeper portions and conglomerates, sands and marls near shore or in subsidiary border basins. The limestones are hard and very fossiliferous but for the most part the fossils consist of the disk- or coin-shaped foraminifers of the genus *Nummulites*, whence the name nummulitic limestone. This limestone, comparable in many ways to the fusulina limestone of the Pennsylvanian, was quarried near Cairo, Egy., in middle Eocene Mokottam strata and comprised most of the material used in building the pyramids. It contains *Nummulites gizehensis*, but the Fayum region southwest of Cairo also yielded the earliest remains of primitive elephants (*Moeritherium*) and whales (*Protocetus*).

The clastic sediments around the borders of Tethys are sometimes called the Flysch but, since they contain few fossils, are difficult to correlate. Most of the sediments were subsequently greatly folded, uplifted and eroded, thus becoming the basis for much of the fine scenery high in the Pyrenees, Alps, Carpathians and Himalayas.

North America.—In North America embayments and inundations of both the Atlantic and Pacific coasts produced a succession of fossiliferous marine and nonmarine strata. The glauconitic sands, marls and clays of the Atlantic border are referred to the Aquia, Nanjemoy and McBean formations. Farther south the advance of the Gulf of Mexico over a wide front, with a lobe reaching to southern Illinois, outlined what is now the Gulf coastal plain, developed from the conglomerates, sands, clays and lignites then deposited. Most of the Claiborne formation is marine sand and nonmarine clay and includes strata formerly assigned to the Late Wilcox. The clays of the Wilcox and Claiborne are used for the manufacture of bricks and pottery and are repositories of the relics of a large land flora that grew along the bayous and lagoons of the ancient coast and whose affinities, because of the palms, mangroves, cinnamons and many other species, are clearly with

living subtropical floras. A few scattered leaves and other parts, however, probably floated into the area of deposition from the temperate foothills and higher elevations of the ancient Appalachians. The Jackson formation, and its correlatives, overlying the Claiborne and completing the Eocene record in the Gulf embayment, is composed of sands, marls and limestones. The scenic Silver springs, one of the largest springs in the world, emerges from the Ocala limestone of Jackson age, 6 mi. E. of Ocala, Fla. The Jackson marine fauna includes the whale *Zeuglodon*, and the land fauna, a herbivorous odd-toed ungulate, the titanothere *Notiotitanops*.

The western interior basins, in part residual from Paleocene time, continued through the Eocene to receive thousands of feet of sediments which entombed the faunas that are the basis for the dating of those strata: Wasatchian, Bridgerian, Uintan and Duchesnean.

Lignites, sandstones, clays and shales succeeded one another repeatedly in the marshes and on the flood plains of the still-sinking trough east of the Rocky mountains. For a short distance along the east front of the Big Horn mountains in Wyoming the earliest Eocene sediments, resting somewhat unconformably on the underlying Paleocene Fort Union, are conglomeratic and contain remains of early horses. Farther east in the Powder river basin the unconformable relations are no longer apparent, making it difficult, in the absence of distinctive fossils, to place the Paleocene-Eocene boundary. Post-Eocene uplift of the mountains tilted the Eocene beds slightly along the mountain flanks, and erosion has removed a large portion, leaving a dissected region of barren, colourful badlands and grassy sagebrush plains.

A small basin on the east side of the Rocky mountains, 20 mi. N.W. of Walsenburg, Colo., includes the brownish to varicoloured but mostly reddish beds referred to the Huerfano formation. The basal strata contain the large ungulate *Coryphodon* and the small titanothere *Lambdotherium*, and later strata contain the tapir-sized titanothere *Palaeosyops* and the large insectivore *Tillotherium*, indicating early and middle Eocene age, respectively.

Between the Big Horn mountains and the Wind River range in Wyoming lie the Big Horn and Wind river basins. The Green river basin, composed of the subsidiary Wasatch, Bridger and Washakie basins, lies between the Wind River range and the Uinta mountains of Utah. Just south of the Uintas is the Uinta basin. The earliest Eocene sediments in some of these basins are the sombre and varicoloured sands, clays and carbonaceous shales usually called Wasatch, similar to those described from the Powder river basin east of the Big Horn mountains. The brilliantly coloured minarets, columns and badlands that comprise the attractiveness of Bryce Canyon National park in Utah are erosional features in the Wasatch formation.

Studies of the ill-defined Wasatch resulted in such subdivisions as Indian Meadows, Gray Bull, Lysite, Lost Cabin, Willwood and Wind River for different localities and basins.

By mid-Eocene the Green river and Uinta basins had developed broad, shallow and somewhat evanescent lakes in which, during appropriate times when precipitation did not muddy the water and temperatures were fairly high, a microscopic fauna and flora prospered, resulting in dark or gray to brownish, fine-grained, many-layered, limy and carboniferous deposits, now the Green river oil shales, with billions of barrels of oil in the form of kerogen (*see* FUELS: *Liquid Fuels: Carbonization*).

The lakes supported many kinds of fishes, whose skeletal remains are quarried extensively at Fossil, Wyo. Most of these fishes were fresh-water species, but a few, including some herrings and a sting ray, were marine or brackish-water forms that presumably came from the sea to spawn. At some stages of the Green river lakes algae caused the formation of extensive reefs. Remains of well-preserved insects, as well as leaves, fruits and seeds of numerous kinds of land plants, are common in the later beds of the formation. Among the plants are palms and other warm temperate species that grew on the plains surrounding the lakes and at not many hundreds of feet above sea level. The distant, higher hills supported a temperate flora that included conifers and broad-leaved species.

The Green river lakes eventually filled up with sediments having a large volcanic ash content and that became the greenish to variegated strata of the Bridger, Uinta and Duchesne river formations. These strata, particularly the Bridger in southwestern Wyoming, have yielded a large assortment of fossil animals, including turtles, crocodiles, birds, rodents, rhinoceros and giant titanotheres.

Eocene Mammals.—The mammals showed distinct advancement from their Paleocene ancestors, developing chiefly many kinds of carnivores and herbivores. The carnivores were represented by creodonts of great diversity and included the family Miacidae, the probable source of all modern forms. The true ungulates, divided into even-toed (artiodactyl) and odd-toed (perissodactyl) lines, made their first appearance in this epoch. The artiodactyls included primitive pigs, oreodonts and camels, and the perissodactyls included *Hyracotherium* (*Eohippus*), the first recognizable member of the horse tribe, the large titanotheres and primitive tapirs and rhinoceros. There were also marsupials, rodents, tillodonts, taeniodonts, insectivores and primates. The largest and most fantastic were the bizarre uintatheres, with sabre tusks and skulls with strange, bony protuberances. Although some of these Eocene animals were bulky, the general impression given by the fauna is one of moderate speed as befits animals adapted to plains and open, wooded areas.

The early Eocene land faunas and floras preserved in the Wasatch strata are notable for their variety. Characteristic mammals of Wasatchian age are the ungulates *Coryphodon*, *Meniscotherium* and *Diacodexis*, the tapirlike *Heptodon* and *Homogalax*, the dawn horse *Hyracotherium* or *Eohippus*, the catlike carnivore *Palaeonictis* and the primate *Pelycodus*.

The characteristic mammalian fauna of Bridgerian Age includes the primates *Apatemys* and *Notharctus;* the ungulates *Helaletes*, *Homacodon*, *Palaeosyops*, the running rhinoceros *Hyrachyus* and the four-toed horse *Orohippus;* the carnivores *Harpagolestes*, *Patriofelis* and *Thinocyon;* and the insectivore *Tillotherium*.

In the Uintan Age the following mammals were characteristic: the primitive swine *Achaenodon;* the running rhinoceros *Amynodon;* the titanotheres *Diplacodon*, *Dolichorhinus* and *Protitanotherium;* the early horse *Epihippus;* the hare- or rabbitlike *Mytonolagus;* the carnivore *Oxyaenodon;* the primitive ruminant *Protoreodon;* and the ancestral camel *Protylopus*.

The Duchesnean Age was characterized by species of the following mammals: the running rhinoceros *Epitriplopus* and *Hyracodon* and the hippopotamuslike *Mesamynodon;* the insectivore *Protictops;* the carnivores *Hessolestes* and the large *Hyaenodon;* the ruminants *Mesagriochoerus* and the camel *Poabromylus;* and the titanothere *Teleodus*.

Other Occurrences and Characteristics.—On the Pacific coast the Eocene is represented by a variety of marine and non-marine strata referred to such formations as Cowlitz, Puget, Roslyn, Manastash, Clarno, Kenai and others in western Oregon, Washington and Alaska; and Meganos, Tejon and Ione in central and southern California. The Roslyn and Kenai formations contain considerable coal, and the Ione correlatives of the Sierra Nevada are part of the famous auriferous gravels. The latter, as bench and channel gravels on the west slopes of the Sierra Nevada in the region north and east of Colfax, Calif., were extensively exploited for gold in the second half of the 19th century. The hydraulic operations exposed many fossil logs and also lenses of clay which, fortunately, contained no gold and were therefore left undisturbed. These clay strata at many localities, however, contain beautifully preserved leaves and seeds that serve to give a fairly clear picture of the vegetation of middle Eocene time in that area. The Sierra Nevada was apparently a low range at that time, and its western slopes graded into wide flood plains and river bottoms bordering on a sea that occupied the present Sacramento valley. Students of the fossil flora identify in it 77 forms including species of climbing fern, cycad, palm, hickory, oak, magnolia, laurel, bean, sycamore, breadfruit, sweetgum, hydrangea, maple, ash, viburnum, grape and dogwood, indicating a warm temperate or subtropical flora, related by numerous genera to the flora of the Wilcox group of the southeastern states and to the middle and late

Eocene floras of Washington, Oregon, Wyoming and Colorado. The flora is interesting also in containing species of *Ailanthus* (tree of heaven), *Cercidiphyllum* (katsura) and *Koelreuteria* (luanshee), whose living counterparts are restricted to eastern Asia. These three, as well as many others now native to the eastern hemisphere, were widely distributed in the northwestern states during the Tertiary but became extinct there in the Late Miocene or Early Pliocene. Floras somewhat younger than that in the auriferous gravels near Colfax have been described from LaPorte, Calif., Comstock, Ore., and Renton, Wash.

The development of a modern-looking vegetation, especially the grasses and other herbs, many of which are peculiarly adapted in flower and fruit to animal users, was most likely both a cause and a result of the progress in animal evolution. As in the Paleocene, palms and other warm temperate to subtropical plants grew in high latitudes in some areas of the north temperate zone, indicating that, for such areas at least, the climate was warmer than it is today. One cooler interval was formerly assumed because supposed glacial till was found in southwest Colorado, but this deposit may be a mudflow or rockslide instead.

Eocene strata have been reported from Australia, New Zealand, Tasmania, Africa, Mexico, Central America, the West Indies and South America. Those from the Caribbean region seem to be best defined. There the early Eocene sea overlapped Trinidad and probably a portion of Venezuela; and the embayment from the Gulf of Mexico that inundated the Gulf coast states spread into northeastern Mexico and the northern part of Yucatán. By middle Eocene time all these areas were still further submerged, the maximum being reached in the late Eocene when Cuba, Hispaniola, Jamaica, the Lesser Antilles and portions of Venezuela and Colombia were covered. An embayment from the Pacific ocean entered a part of Peru. The result of these marine invasions was the deposition of several thousands of feet of fossiliferous sediments, principally limestone.

The rich marine life of the Eocene has already been referred to in the discussion of the several provinces. In fresh waters, turtles, crocodiles and fishes of many kinds abounded. Remains of true birds, such as impressions of feathers and petrified bones, indicate ancestral buzzards, eagles and others, including the large flightless form *Diatryma*.

During the Eocene epoch a series of strata amounting to perhaps 15,000 ft. accumulated. Considerable volcanic activity is reflected by the large percentage of ash disseminated through the sediments in the western United States, as, for example, in many of the leaf- and wood-bearing strata of Yellowstone National park and surrounding areas. Volcanic eruptions also occurred in Central America and the Antilles. Similar evidence of ash falls in the Eocene rocks of Denmark and northwestern Germany point toward a correlation with the activity that produced the post-Cretaceous and pre-Quaternary lavas of western Scotland and Ireland. The Deccan traps of India probably date from early Eocene time.

Except for the movements connected with the incursions and regressions of the sea along some seacoasts and the subsidence of some interior basins, there were no great tectonic disturbances of the earth's crust during this epoch. The uplift of the Rocky mountains or Laramide revolution, which began in the Cretaceous, continued intermittently through the Eocene. The Isthmus of Panamá, it may be noted, submerged and emerged at least once.

The principal economic products derived from Eocene strata are clay, coal, bauxite, building stone, gold, gypsum, petroleum and phosphate.

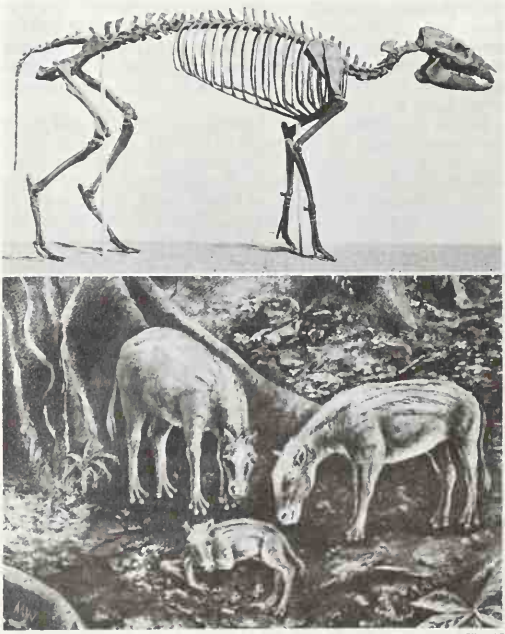
See also references under "Eocene and Paleocene" in the Index volume.

BIBLIOGRAPHY.—*General:* S. von Bubnoff, *Geologie von Europa* (1935); M. Gignoux, *Stratigraphic Geology* (1955); E. Kayser, *Lehrbuch der Geologie* (1909); E. Neaverson, *Stratigraphical Palaeontology* (1955); W. B. Scott, *A History of Land Mammals in the Western Hemisphere* (1937); L. D. Stamp, *Stratigraphy* (1923). *Special:* E. W. Berry, *The Lower Eocene Floras of Southeastern North America* (1916); R. W. Brown, *Cretaceous-Tertiary Boundary in the Denver Basin, Colorado* (1943), *Tertiary Strata in Eastern Montana and Western North and South Dakota* (1952); A. ten Dam, *Die stratigraphische Gliederung des niederlandischen Paläozäns und Eozäns nach Foraminiferen* (1944); Julia Gardner, *The Midway Group of Texas* (1935); C. L. Gazin, *The Lower Eocene Knight Formation of Western Wyoming and Its Mammalian Faunas* (1952), *Paleocene Mammalian Faunas of the Bison Basin in South Central Wyoming* (1956); C. L. Gazin and Harry Tourtelot, *The Geology and Vertebrate Paleontology of Upper Eocene Strata in the Northeastern Part of the Wind River Basin, Wyoming* (1956, 1957); G. L. Jepsen, *Stratigraphy and Paleontology of the Paleocene of Northeastern Park County, Wyoming* (1930); V. E. Khain and L. N. Leontev, *Contribution to Palaeogeology of the Caucasus in Lower Palaeogene* (1947); H. D. MacGinitie, *Middle Eocene Flora From the Central Sierra Nevada* (1941); W. D. Matthew, *Paleocene Faunas of the San Juan Basin, New Mexico* (1937); H. F. Osborn, *The Titanotheres of Ancient Wyoming, Dakota, and Nebraska* (1929); Bryan Patterson, *New Pantodonta and Dionocerata From the Upper Paleocene of Western Colorado* (1939); Eleanor M. Reid and Marjorie E. J. Chandler, *The London Clay Flora* (1933); G. G. Simpson, *The Fort Union of the Crazy Mountain Field, Montana, and Its Mammalian Fauna* (1937); *The Beginning of the Age of Mammals* (1937); F. B. Van Houten, *Early Cenozoic Facies in the Rocky Mountain Region* (1945). (R. W. Br.)

EOHIPPUS, the "four-toed horse," the earliest known stage in the evolution of the horse, fossils of which are found in Lower Eocene formations, about 50,000,000 years old, in the western United States and western Europe.

In Europe the animal is called by its scientific name, *Hyracotherium*, but the more picturesque synonym *Eohippus*, meaning "dawn horse" continues in popular use.

The smallest species were no bigger than a fox, but others were nearly twice as large. Although linked to living horses (genus *Equus*) by an almost complete series of intermediate forms, eohippus is so primitive that the relationship could not be established without these links. The eyes were large and were in the middle of the skull, not far back as in recent horses. The teeth were small, low crowned and simple in pattern. The forefoot had four well-developed toes and the hind foot three. The weight rested largely on a pad under the toes, as it does in dogs and cats, and not wholly on the central hoof as in modern horses. The brain was smaller and simpler in structure than in later horses. *See* EQUIDAE. (G. G. Si.)



BY COURTESY OF (TOP) CHICAGO NATURAL HISTORY MUSEUM; (BOTTOM) AMERICAN MUSEUM OF NATURAL HISTORY

EOHIPPUS (HYRACOTHERIUM): (TOP) RESTORED SKELETON; (BOTTOM) ARTIST'S RENDERING OF THE EOHIPPUS IN ITS SUPPOSED NATURAL HABITAT

EOLITH (literally, "dawn stone"), stone, either in its natural state or crudely broken, that was used as a tool in ancient times. A modern parallel would be a paperweight consisting of a large pebble. Presumably the earliest stone tools used by primitive man or manlike primates were eoliths. But the simplicity of their structure makes it difficult to differentiate between such crude tools and ordinary stones. *See* also ARCHAEOLOGY: *Culture History of the Pleistocene;* FLINTS.

ÉON, THE CHEVALIER D' (1728–1810), French secret agent who spent most of his life in female dress, was born at Tonnerre on Oct. 5, 1728. His first mission was to the Russian empress Elizabeth in 1755, on which he seems to have disguised himself as a woman. After good service as a dragoon captain, he went to London in 1762 with the duc de Nivernais (Louis Jules Mancini). On returning to Versailles with the treaty of Paris ratified (1763), he received the cross of St. Louis. Sent back to London with secret instructions from the king for espionage, he became involved in a quarrel with the French ambassador, but refused to return to France—afterward claiming that Louis XV had instructed him to redisguise himself as a woman and to hide in the city.

Speculation began in London as to the chevalier's sex. Eventually, in need of money, he pretended to Beaumarchais in 1775 that he was really a woman. Believing this, Beaumarchais managed to get him a pension, but Éon was ordered to return to France wearing women's dress. In 1777 he received the command: "By order of the king: Charles Geneviève Louise Auguste Andrée Timothée d'Éon de Beaumont is commanded to leave off the dragoon's uniform which she is wearing, and to dress according to her sex . . ." Thenceforth he always wore women's dress. Having returned to London in 1785, he died there on May 21, 1810. An autopsy performed two days later certified him male.

(AN. CA.)

EÖTVÖS, JÓZSEF, BARON (1813–1871), Hungarian writer and statesman who devoted his energies to the creation of modern Hungarian literature and the establishment of a modern democratic Hungary, especially through educational reform. Born at Buda, Sept. 13, 1813, he studied at Budapest (1826–31) and became inspired with liberalism and the desire to reform Hungarian society. During 1836–37 he studied social conditions in England and France, returning deeply impressed by liberal philanthropy, romanticism and utopian socialism.

Eötvös proclaimed the social mission of literature and in all his writings fought for alleviation of poverty. His first novel, *Karthausi* (1839–41), expresses the romanticism of the French restoration period. Its theme is the disappointment caused by the failure of the July Revolution in France, but it was intended as a criticism of conditions in Hungary. Eötvös showed the contradictions underlying bourgeois civilization because he wanted to replace Hungarian feudalism by an "improved" capitalism. In his essays he advocated modernization of the penal code and prison reform and in a study of poverty in Ireland he pointed out the social problems of western Europe. The novel *A falu jegyzöje* (1845; Eng. trans. *The Village Notary,* 1850), a bitter satire of old Hungary, was influenced by Disraeli's *Coningsby.* Through its picture of the 16th-century Hungarian peasant rebellion, his great historical novel, *Magyarország 1514-ben* (1847), mobilized opinion to abolish serfdom.

Eötvös became minister of education in the revolutionary government of 1848 and prepared a plan to raise educational standards. In Sept. 1848, however, disagreement with Kossuth caused him to resign, and until 1851 he lived in Munich, where he began his great work on the influence on the state of the principal ideas of the 19th century (1851–54; also in German). This attempted to work out the principles of the French Revolution and depicted an ideal liberal state, based on English constitutional ideas and practice. In his political writings, Eötvös attacked both the regime of Napoleon III and Austrian absolutism. He wished to base relationship between Austria and Hungary on the principles of 1848 and the compromise of 1867 was partly his work.

His later years were devoted to political and philosophical activity. His collected reflections (published 1864) show a growing stoicism of a type peculiar to Hungarian literature in the post-revolutionary period. He played a distinguished part in the reorganization of the Hungarian academy, he maintained close relations with western scholars (*e.g.,* J. K. Bluntschli), and was one of the first to apply Macaulay's theories to Hungary and to accept Darwin's ideas. His theory of scientific method was based on Francis Bacon's inductivism, which he expounded in his essays. Like Montalembert, with whom he corresponded, he tried to find a liberal solution to the problem of the relationship between the state and the Roman Catholic Church. He became minister of education again after 1867 and was concerned with the modernization of the educational system and with the organization of a new university system. He died in Budapest, Feb. 2, 1871.

After the revolution, Eötvös wrote only one novel, *Növerèk* ("The Sisters," 1857), explaining his ideas on education. He also gave up writing poetry, which had been sentimental and romantic in tone. Yet his literary work is of great importance. His short stories mark the beginning of a new portrayal of the peasant in Hungarian literature, and at a time when the romantic novel was in fashion, he was a pioneer of realism, initiating trends which were only to be developed in the 20th century.

BIBLIOGRAPHY.—Zoltán Ferenczi, *Báró Eötvös József* (1904); Géza Voinovich, *Báró Eötvös József* (1903); István Soter, *Eötvös József* (1952).

(I. SR.)

EÖTVÖS, ROLAND, BARON VON (1848–1919), Hungarian physicist and experimenter in terrestrial magnetism and gravitation. Son of Baron Josef Eötvös, he was born in Budapest July 27, 1848. He studied at Budapest and Heidelberg, and in 1872 he was appointed professor of physics in the University of Budapest. In the following year he was elected member, and in 1889 president, of the Hungarian Academy of Sciences. From 1894–95 he was minister of public instruction. His first scientific studies were connected with capillarity; in 1885 he published the result of these investigations in which he deduced the law connecting the variations of surface tension with temperature. From 1890 onward he examined the phenomena connected with the problems of gravitation and of terrestrial magnetism. He constructed a number of magnetic instruments. Later he constructed the double-armed torsion balance. This so-called Eötvös balance is an extremely delicate instrument for determining the variations of gravity quickly and accurately; it became an essential instrument for the location of subterranean materials. In a treatise which won him the Benecke prize from Göttingen university in 1909, Eötvös demonstrated with great exactitude that the attraction of bodies is independent of the quality of the substance. Toward the end of his life he studied the variations of the gravitation of bodies moving on the earth. He demonstrated by experiments that if a body moves toward the east it loses in weight. Eötvös showed the rotation of the earth by means of a slowly rotating gravity balance.

Eötvös died in Budapest on April 8, 1919.

EPACRIDACEAE, a family of rather showy, mostly small shrubs, comprising over 20 genera and over 300 species, chiefly Australasian, but a few are found in India and South America, and *Styphelia* is found in Hawaii. They have small, heath-like leaves, and solitary, bell-shaped, often very fragrant flowers in the leaf axils. Some have a fleshy, edible fruit. A number are cultivated for ornament in British greenhouses, especially *Epacris longiflora,* with rosy-crimson flowers, and *E. obtusifolia,* with waxy, yellowish-white flowers, both fragrant. They are little cultivated in the United States.

(N. TR.)

EPAMINONDAS (*c.* 418–362), Theban general whose tactical innovations, employed at the battle of Leuctra, helped to establish the predominance of Thebes among the land powers of Greece, was born about 418 B.C. He was educated by Lysis of Tarentum, a Pythagorean philosopher in exile. In 385 he was fighting on the Spartan side in the attack on Mantinea, where he saved the life of Pelopidas. In 371 he represented Thebes at the congress at Sparta, and by his insistence on signing for the Boeotian cities under Theban control prevented the conclusion of a general peace. In the ensuing campaign he commanded the Boeotian army that met the Peloponnesian levy at Leuctra (*see* LEUCTRA,

BATTLE OF, 371 B.C.) and there achieved a brilliant victory, due mainly to his innovations.

These innovations were the use of a 50-deep phalanx on the left as a striking head, with the centre and right *en échelon* behind it, in close co-operation with cavalry on the flank. In winter 370–369, in support of the new Arcadian league, he took a large army into Laconia; there he dealt a lasting blow at Sparta's predominance in the Peloponnese by liberating the Messenians and rebuilding Messene. In 369 or 368 he forced the isthmus lines and secured Sicyon for Thebes; it was probably in this campaign that he founded Megalopolis, the centre of the Arcadian league. Somewhere about this time he seems to have been put on trial on some charge connected with his command; the evidence is conflicting. Plutarch says it was after the first campaign, that the charge was of exceeding his term and that he was acquitted. Diodorus' account is that it was after the second campaign, that the charge was laxity and that he was removed from the command.

In the following year he served as a common soldier in Thessaly, and when Pelopidas (*q.v.*) was seized by Alexander, tyrant of Pherae, Epaminondas was reinstated in command and contrived the safe retreat of the Theban army from a difficult position. Returning to Thessaly in 367 at the head of an army he procured the release of Pelopidas without striking a blow. In his third expedition (366) to the Peloponnese, Epaminondas again eluded the isthmus garrison and won over the Achaeans to the Theban alliance. When subsequent complications threatened the position of Thebes in the Peloponnese he again mustered a large army in order to crush the newly formed Spartan league (362). After some masterly operations between Sparta and Mantinea, by which he nearly captured both these towns, he engaged in a battle on the latter site. His tactics (*see* MANTINEA) were similar to those of Leuctra. His death on the field converted what was practically a complete victory into an unprofitable draw.

His title to fame rests mainly on his brilliant qualities both as a strategist and as a tactician; he had great influence on military art in Greece. For the purity and uprightness of his character he likewise stood in high repute; his culture and eloquence were outstanding even by Athenian standards. In politics his chief achievement was the final overthrow of Sparta's predominance in the Peloponnese; as a constructive statesman he displayed no special talent, and the lofty Panhellenic ambitions which are imputed to him at any rate never found a practical expression.

BIBLIOGRAPHY.—Nepos, *Epaminondas;* Plutarch, *Pelopidas;* Diodorus, xv, 52–88; Xenophon, *Hellenica,* vii; M. Cary, in *Cambridge Ancient History,* vol. vi, ch. 4 (1927); N. G. L. Hammond, *A History of Greece,* book v, ch. 2 and 3 (1959).

EPAULET (EPAULETTE), a shoulder insignia formerly widely worn in military and naval service. The epaulet derives from a knot of ribbons used to fasten a baldric to the shoulder (French *épaule*). In the 18th century elite French troops were distinguished by clusters of red ribbons on their shoulders. The device was adopted by the duc de Choiseul in 1758, using metal lace as a precise method of differentiating commissioned rank. The epaulet was introduced into the British army about 1768 and thereafter its use became widespread although not universal. From the start it was employed to indicate the rank and function of both officers and men.

The parts of an epaulet are the strap lying along the shoulder; the crescent at the outside end; and the fringe of bullion hanging from the crescent. Rank was indicated by the size of the bullion; by devices attached to the strap; or by the number and way worn—some officers wore only one epaulet, in which case its being on the right or left shoulder was significant. Officers' epaulets were commonly of metal lace, men's of coloured worsted or silk.

By 1850 the cost of a pair of officer's epaulets had so increased that they were largely confined to full-dress uniforms. Thereafter their use declined and they were replaced by less expensive devices such as shoulder knots of twisted cord, metal scales (ornaments similar to an epaulet but without a fringe) or cloth shoulder boards. In the United States officers adapted the lace strap used to hold the epaulet on the shoulder as a device to bear insignia of rank.

The epaulet was given up by most U.S. army officers in 1872; naval officers wore it until 1951. British army officers ceased wearing it in 1855 but naval officers still use it on rare occasions. Worsted epaulets in various colours also form part of the full dress of several celebrated military corps such as the French Foreign Legion. (F. P. T.)

ÉPÉE, CHARLES MICHEL, ABBÉ DE L' (1712–1789), French educator noted for his work on behalf of the deaf-mute, was born at Versailles on Nov. 25, 1712, being the son of the king's architect. He studied for the church, but as a Jansenist had considerable difficulty in being ordained.

He was eventually granted ordination by the bishop of Troyes in 1736, but the bishop died soon afterward and Épée was deprived of his ecclesiatical functions.

About 1760 Épée took the place of a recently deceased priest who had been instructing two deaf-mutes by means of pictures. At the time Giacobbo Pereire (*q.v.*) was in Paris, and from him Épée obtained his manual alphabet. Since the abbé believed that the language of signs was the vernacular of the deaf, he considered it hardly possible for them to express their ideas directly in writing. Speech for the deaf he regarded as a purely mechanical operation having no sort of relation to their intellectual activities, and lipreading, as far as it went, as of precisely the same value as the manual alphabet.

He attained far greater success than Pereire or any of his predecessors, and the whole system of instruction of deaf-mutes that followed virtually owed its origin to his intelligence and devotion. In 1755 he founded at his own expense a school for deaf-mutes, which afterward was succeeded by the Institution Nationale des Sourds Muets, founded by the national assembly at Paris in 1791. He died on Dec. 23, 1789.

Épée published a number of books on his method of instruction, including *Instruction des sourds et muets par la voie des signes méthodiques* (1776) and *La Veritable Manière d'instruire les sourds et muets, confirmée par une longue expérience* (1784). His *Dictionnaire général des signes* was completed by his successor, the abbé Sicard. *See* also DEAF AND HARD OF HEARING, TRAINING AND WELFARE OF. (M. E. F.)

ÉPÉE FENCING: *see* FENCING.

ÉPERNAY, a town of northeastern France, *département* of Marne, lies on the left bank of the Marne, surrounded by hills covered with vineyards and forests, 27 km. (17 mi.) S.S.W. of Reims by road. Pop. (1954) 19,527. Épernay, destroyed or burned 22 times, has few old buildings left. The large modern church of Notre Dame is in late 12th-century style and has interesting 16th-century windows. In the Château Perrier are housed the library (with 9th-century manuscripts) and the museums of champagne wines (showing the history of the cultivation of the vines and the making of the wine), fine arts and local prehistory and archaeology. Épernay is the chief entrepôt of the Champagne wines, which are bottled and kept in vast vaults cut in the chalk rock on which the town is built. Apparatus and material used in the champagne industry are also manufactured and there are railway works.

Known to the Romans as Sparnacum, the town belonged to the archbishops of Reims from the 5th to the 10th century, when it passed to the counts of Champagne. It suffered severely during the Hundred Years' War and was burned by Francis I in 1544. It was captured by Henry of Navarre in 1592, and in 1642 along with Château-Thierry, it was made a duchy and assigned to the duke of Bouillon. In the Franco-German War, Épernay was in German occupation, in World War I from Sept. to Oct. 1914 and in World War II from June 1940 to Aug. 1944. (K. v. G.)

ÉPERNON, JEAN LOUIS DE NOGARET DE LA VALETTE, DUC D' (1554–1642), one of the most powerful new magnates in French politics in the last years of the Valois dynasty and in the early years of the Bourbon, was born at Caumont, in Gascony, in May 1554. He rose to prominence as one of the Mignons or favourites of Henry III, under whom he and Anne de Joyeuse acted more or less as prime ministers in the 1580s. Colonel-general of the infantry and duc d'Épernon from 1581 and admiral of France from the end of 1587, he was also made governor of several provinces. His rapacity was notorious, but he also

showed great political ability, and his energetic action did much to defend the crown in the civil wars. By seizing Boulogne, he prevented the Spanish Armada's taking refuge there. Temporarily out of favour with the king, he returned to his side after the murder of Henry de Guise (1588) and advised reconciliation with Henry of Navarre, the future Henry IV of France. When Henry III had been murdered (1589), the duc at first refused to serve Henry IV (then still a Protestant) and even made a secret treaty with Spain, but when Henry had established himself he reappeared at court and posed as a loyal subject of the king, who treated him well. Épernon, however, knowing that Henry was only waiting for an opportunity to break his power, joined every conspiracy of the reign—without ever being caught out. There are some grounds for believing that, together with Spain, Concino Concini and the marquise de Verneuil, he arranged the murder of the king by François Ravaillac, who is known to have received money from Épernon's mistress, Charlotte du Tillet, and who was kept in custody by the duc for a whole day after the murder (May 14, 1610). Épernon then managed a sort of *coup d'état* whereby Marie de Médicis became regent for Louis XIII, but she did not take him into the government. After Louis XIII had taken power (1617), Épernon fell into disgrace (1618). He avenged himself by organizing the escape of Marie de Médicis from Blois (1619) and by supporting her in the civil war that ensued. On her return to court, she secured his appointment as governor of Guienne (1622). There he lived in wanton luxury punctuated by violent quarrels—for instance with Henri de Sourdis, archbishop of Bordeaux, whom he actually struck with his cane. Finally the Cardinal de Richelieu deprived him of his governorship (1638) and exiled him to Loches (1641), where he died on Jan. 13, 1642.

Épernon's second son, BERNARD (1592–1661), duc de La Valette till he succeeded as duc d'Épernon, a soldier of fierce and ungovernable temper, was married first to Henry IV's natural daughter Gabrielle de Verneuil (whom he is said to have poisoned in 1627), then to Richelieu's niece Marie du Cambout. Even so, after the loss of the battle of Fuenterrabia (1638) he was accused of high treason and fled to England. Returning to France on Louis XIII's death, he was given his father's governorship of Guienne. His oppressive conduct there did much to provoke Bordeaux into rebellion during the Fronde, and he was universally detested.

Bernard's younger brother LOUIS (1593–1639), Cardinal de La Valette from 1621, resigned the archbishopric of Toulouse in 1627 to take up a military career. Constantly loyal to Richelieu, he gave him decisive support on the Day of Dupes (1630).

See L. Mouton, *Un demi-roi: le duc d'Épernon* (1922) and *Le Duc et le roi* (1924); P. Erlanger, *L'Étrange Mort de Henri IV* (1957).
(P. ER.)

EPHEBI, a name given, in ancient Athens and other Greek towns, to a class of young men from 18 to 20 years of age, who formed a sort of college under state control. At the age of 18 the Athenian youth attained his civil majority, and, provided he passed the scrutiny (*dokimasia*) as to age, civic descent and physical capability, was enrolled on the register of his deme or township. He thereby at once became liable to the military training and duties which, at least in the earliest times, were the main object of the ephebia. In the time of Aristotle, the names of the enrolled ephebi were engraved on a bronze pillar (formerly on wooden tablets) in front of the council chamber.

In Aristotle's day the ephebi were under the general supervision of ten *sophronistai* who, as their name ("chasteners") implies, took special interest in their morals. These ten, one from each tribe, were elected from a list of 30 (all over 40 years old) chosen by the fathers of the ephebi. At their head was a *kosmetes* elected from the whole people. After a ceremonial circuit of the public shrines the ephebi spent their first year on garrison duty at the Piraeus on the walls of Munychia or Acte, living in tribal messes. Their military training was in the hands of experts, of whom the chief were the *hoplomachos,* the *akontistes,* the *toxotes* and the *aphetes* (instructors respectively in the use of arms, javelin throwing, archery and the use of artillery engines). At the end of this year a special meeting of the assembly was held in the theatre, at which those whose service was satisfactory were presented by the state with shield and spear, which, with the *chlamys* ("cloak") and *petasos* ("broad-brimmed hat"), made up their equipment.

The ephebi took part in some of the most important Athenian festivals. During the Eleusinia they fetched the sacred objects from Eleusis and escorted the image of Iacchus on the sacred way (*see* MYSTERY). They also performed police duty at the meetings of the assembly.

On their enrollment or, more probably, at the end of their first year's probation, all ephebi took a traditional oath not to discredit their arms and to protect the sanctuaries and their country. This oath was preserved with minor variation in the rhetorical lexicon of Pollux (2nd century A.D.) and the anthology of Stobaeus (5th century A.D.), and a more complete copy has been found in an inscription of the late 4th century B.C. During their two years' service the ephebi were exempt from taxation and other obligations to the state, and they could not appear in the courts as plaintiff or defendant, except in cases of inheritance.

There is no firm evidence of a formal institution of ephebi before the middle of the 4th century B.C., but there are archaic features in the oath, which in the late 4th century is called "ancestral," and though the organization may have been less elaborate it is probable that these two age groups were from early times regarded as special.

After the end of the 4th century B.C. the institution underwent a radical change. Enrollment ceased to be obligatory, lasted only for a year, and the limit of age was discontinued. Inscriptions attest a continually decreasing number of ephebi, and with the admission of foreigners the college lost its representative national character. In 276 there were only 29 members. Later there was some recovery but in the early empire the number did not exceed three hundred, and often less than half were Athenian.

The military element was no longer all-important, and the ephebia became a sort of university for well-to-do young men of good family, whose social position has been compared with that of the Athenian "knights" of earlier times. The military instructors were replaced by philosophers, rhetoricians, grammarians and artists. In Roman imperial times several new officials were introduced, one of special importance being the director of the *Diogeneion,* where youths under age were trained for the ephebia. At that period the college of ephebi was a miniature city; its members called themselves "citizens," and it possessed an archon, strategus, herald and other officials, after the model of Athens itself. The institution lasted to the end of the 3rd century A.D.

Ephebi were found in most Greek cities, though the arrangements for their training and supervision varied. The Roman emperor Claudius (A.D. 41–54), in order to settle the disputes there, declared admission as an ephebus to be the prerequisite for the citizenship of Alexandria.

BIBLIOGRAPHY.—Aristotle, *Constitution of Athens,* ch. 42; for the oath, M. N. Tod, *Greek Historical Inscriptions,* vol. ii, p. 303 (1950); C. A. Forbes, *Greek Physical Education,* ch. 6 (1929). (R. ME.)

EPHEDRA, an evolutionally early genus of mostly low, straggling or climbing desert shrubs that resemble the horsetails (*see* HORSETAIL). *Ephedra* is the sole genus of the family Ephedraceae, order Gnetales, and is of academic interest because of its alleged position between the gymnosperms and angiosperms. A few species, popularly called joint fir, joint pine or Mexican tea, are cultivated as ground covers in arid regions, and by Mexican Indians as a source of a medicinal tea. A Chinese species, *E. sinica,* yields the important drug ephedrine (*see* DECONGESTANT DRUGS). Certain species were the source of a fermented juice drunk ceremonially by Vedic and Zoroastrian priests (*see* SOMA).

See also GYMNOSPERMS: *Gnetales.*

EPHEDRINE: *see* DECONGESTANT DRUGS.

EPHEMERIS. An ephemeris is a table giving, for stated points of time, the position or other numerical particulars of one or more of the heavenly bodies. Such tables are essential to the navigator, whose observations of the sun, moon and stars would be valueless without a knowledge of their exact positions; to the astronomer who wishes to direct his telescope to a particular celestial object, or to check the timekeeping of his clock; and to the mathematician who predicts eclipses or the return of comets.

The word ephemeris is frequently applied to a collection of such tables embodied in the form of an almanac, and it is used in that sense in the following remarks.

National Ephemerides.—The oldest of the national astronomical ephemerides is the *Connaissance des temps ou des mouvements célestes*, founded by J. Picard in 1679 and published at Paris under the auspices of the bureau des longitudes.

The British *Nautical Almanac and Astronomical Ephemeris for the Meridian of the Royal Observatory, at Greenwich* commenced with the year 1766 through the initiative of Nevil Maskelyne, then astronomer royal. At that time no wholly satisfactory method of determining longitude at sea existed. The method of lunar distances, destined to become the standard method during the 19th century but since discarded, was then impracticable because of the inaccuracy of the existing lunar tables, and because special knowledge and long and tedious computation were necessary in their use. When the tables of Johann Tobias Mayer, of Göttingen, came into the hands of Maskelyne, he quickly realized their value to navigation. To make them practically useful, however, it was necessary to prepare from them an ephemeris of the moon for every noon and midnight. Maskelyne's proposals to the Board of Longitude resulted in the establishment of the *Nautical Almanac* under his superintendence.

After the death of Maskelyne in 1811 the *Nautical Almanac* fell into disrepute and became notorious for its errors. As a result of many attacks made on it, principally by Francis Baily and Sir James South, the admiralty appealed to the Royal Astronomical society in 1830. The society appointed a strong committee, whose many recommendations were accepted in their entirety and embodied in the *Almanac* for 1834. Apart from improvements in the basic data from which the ephemerides are constructed, and a few changes made in 1896 at the suggestion of the Royal Astronomical society, the form of the *Almanac* remained substantially unaltered until the issue of that for 1923, when the volume was completely recast. Changes necessitated by the alteration in 1925 of the beginning of the astronomical day from noon to midnight were made and consideration was given to the changed requirements of modern astronomers arising from the introduction of calculating machines.

The third great national ephemeris emanated from Berlin, under the editorship of J. E. Bode. The first volume appeared in 1774, with the title *Astronomisches Jahrbuch oder Ephemeriden für das Jahr 1776*. In 1830 J. F. Encke changed the title to *Berliner astronomisches Jahrbuch*. This ephemeris, which had been produced by the Astronomisches Rechen-Institut, Heidelberg, ceased publication with the edition for 1959.

The *Almanaque Náutico y Ephemerides Astronomicas*, adapted to the naval observatory of San Fernando, near Cadiz, in Spain, appeared first in 1827 for the year 1830.

Finally the *American Ephemeris and Nautical Almanac* was first published at Washington, D.C., in 1852 for the year 1855, the editor being Adm. C. H. Davis. Later, under the superintendence of Simon Newcomb, this ephemeris stepped into the front rank and became the best of the national ephemerides.

Among the similar ephemerides published by other countries are the *Annuaire Astronomique de l'URSS*, first issued for 1925, and the *Indian Ephemeris and Nautical Almanac*, first issued for 1958.

Although these ephemerides were started to provide for the needs of navigators at sea, the emphasis shifted in time to fulfilling the more extensive requirements of astronomers. This prompted the appearance of the *American Nautical Almanac* and a shortened edition of the British *Nautical Almanac* containing only that material relevant to celestial navigation. As the period of World War I approached, the increasing use of faster naval and maritime vessels, and the added requirement for determining a position at sea more quickly and more precisely, made it apparent that tables designed specifically for the needs of celestial navigation were required. The first British *Nautical Almanac, Abridged for the Use of Seamen* appeared as a separate publication in 1914. In the following decade changes were introduced in other national almanacs to meet the needs of the celestial navigator. In the period between World Wars I and II, the growing importance of the airplane for military and commercial use imposed yet further requirements on the producers of the national ephemerides for specialized tables of the sun, moon, planets and selected stars by means of which the location of an aircraft could quickly and easily be determined. After several attempts to utilize the existing nautical almanacs, the new concept and format of the present *Air Almanac* was introduced. Numerous specialized ephemerides for air navigation are being published.

International Co-operation.—Although a number of national ephemerides are printed they are not all computed independently. Conferences of the directors, commencing with the Paris conference of 1896, have been held at irregular intervals to settle upon international values of the astronomical constants and to apportion the extensive computing load among the various co-operating countries. Opportunities for co-operation have also been provided by the triennial meetings of the International Astronomical union, one of whose "commissions" is devoted to ephemerides. The British Nautical Almanac office is responsible for computing the fundamental ephemerides of the sun and planets, occultations of stars by the moon, and the hourly ephemeris of the moon, based on fundamental data derived from the theory of E. W. Brown. The office of the *Connaissance de temps* produces the apparent places of circumpolar stars and the ephemerides of the four great satellites of Jupiter. The American Nautical Almanac office is charged with the production of the elements of eclipses from the sun and moon data supplied by the British Nautical Almanac office, of sunrise, sunset, moonrise and moonset tables, and of tables relating to the natural satellites and to physical observations of the sun, moon and planets. The Astronomisches Rechen-Institut computes the apparent places of the more than 1,000 fundamental stars. *Ephemerides of Minor Planets*, compiled and published annually by the Institute for Theoretical Astronomy, U.S.S.R., represents further international co-operation in the preparation of astronomical tables. To avoid duplication of compilation and printing costs the British and American Nautical Almanac offices have unified their ephemerides. The *American Ephemeris* and the British *Astronomical Ephemeris* are of identical format and are reproduced from the same copy by photo-lithography separately in each country, commencing with the issue for 1960. The *Nautical Almanac* and the *Air Almanac* are similarly unified, beginning with the volumes for 1958 and 1953 respectively.

Tables of Heavenly Bodies.—Some reference has already been made to the tables used in producing the ephemerides. When a theory of the motion of a heavenly body has been evolved, the available observations are used to supply the numerical constants of the theory. Then tables are formed by means of which the position of the body may be found for any desired date. These tables allow for every known gravitational effect acting on the body; for instance, the effect of the attractions of other members of the solar system, and of changes in the size, shape, plane and orientation of the orbit of the body. In Brown's *Tables of the Moon* no fewer than 1,400 terms are included, and the determination of a single position, which involves the use of 180 tables, is a good day's work. The introduction of high-speed electronic calculators has made feasible, in some cases, the direct numerical evaluation of the theory of motion of a celestial body, thus obviating the necessity of forming tables.

Ephemerides of the apparent positions of stars are based on catalogues produced by combining all available observations of these stars and on a knowledge of the motions of the earth, which give rise to precession, nutation and aberration.

In the making of the observations from which these all-important tables and catalogues are formed, the national observatories have, of course, played a leading part; and in several of the ephemeris offices much work has been done in the analysis of these observations, in preparation of the numerical material on which the predictions are based, and in the determination of improved values of the astronomical constants and new general planetary theories.

Computation of an Ephemeris.—The production of an ephemeris involves very heavy computing and careful checking, though it may perhaps be a matter for surprise that a calculation

is rarely repeated. The principal check employed is known as differencing, and depends on the fact that the quantities are tabulated for equal intervals of time, and should exhibit a certain smoothness of progression. As a simple illustration of checking by differences, a column of numbers and their cubes may be considered. The process consists in subtracting each cube from the one following and setting down the result as the first difference. Similarly the second, third and higher differences may be found.

No.	Cube	Differences First	Second	Third
1	1			
		7		
2	8		12	
		19		6
3	27		18	
		37		6
4	64		24	
		61		6
5	125		30	
		91		6
6	216		36	
		127		6
7	343		42	
		169		6
8	512		48	
		217		
9	729			

No.	Cube	Differences First	Second	Third
1	1			
		7		
2	8		12	
		19		6
3	27		18	
		37		5
4	64		23	
		60		9
5	124		32	
		92		3
6	216		35	
		127		7
7	343		42	
		169		6
8	512		48	
		217		
9	729			

In the table on the left the third differences are uniformly 6. In the other table 5^3 or 125 has been replaced by 124, with the result that the smoothness of the third difference is interrupted. The effect of an error of even a single unit in the last figure of a function is seen to be a considerable distortion of the smoothness of the final order of differences. Not only does the process of differencing provide a most powerful method of detecting accidental errors but it also enables the computer to locate the erroneous value and to determine the sign and magnitude of the error.

Calculating Machines.—Until the early part of the 20th century, tables of logarithms formed the primary assistance to computation. The gradual introduction of mechanical calculators increased both the speed and accuracy with which the calculations can be performed. Of even greater impact on the field of computation was the development of punched-card machines and high-speed electronic calculators. Among the earliest users of punched-card machines for scientific computations were the British and American Nautical Almanac offices. The summation of the numerous terms necessary to determine a single position of the moon from Brown's tables was adapted to punched-card machines at the British Nautical Almanac office as early as 1930, while the first American *Air Almanac*, prepared in 1940 for the year 1941, was computed almost entirely by such equipment. In addition to their use in computation, specialized punched-card machines are also utilized to prepare copy for, and compare proof of, several of the national ephemerides and navigational almanacs. High-speed electronic digital calculators have made possible the solution of problems formerly considered impossible because of the tremendous labour of computing involved. The simultaneous integration of the equations of motion of the five outer planets to compute their positions for every 40th day from the year 1653 to 2060 is typical of the astronomical calculations which can be performed with these machines. Such computations, which form the basis of the national ephemerides, are produced with incredible speed, the basic arithmetic operations of addition, subtraction, multiplication and division being performed in milliseconds (*see* COMPUTING MACHINES, ELECTRONIC). (L. J. C.; R. L. DE.)

EPHEMEROPTERA, the May flies (*q.v.*) an order of winged insects (*q.v.*) which is very distinctive and isolated, having no close living allies; the extinct order Paleodictyoptera was most nearly related. The ordinal name refers to the fleeting existence of the adult (a few hours to a few days). (G. F. Es.)

EPHESIANS, EPISTLE TO THE, a book of the New Testament traditionally included among the Epistles of Paul and belonging (with Colossians, Philemon and Philippians; *qq.v.*) to the so-called "captivity epistles," claiming to be written by Paul while a prisoner. But its authorship and origin are much debated. The value of its contents, however, is to a considerable extent independent of such questions.

Contents and Themes.—After the greeting (i, 1–2) the book launches into a glowing thanksgiving for the mysterious design of God's purpose as summed up in Christ and, in particular, as including and binding together, through the redemptive death of Christ, the two hostile camps of Jew and Gentile in a single body (i, 3–iii, 21)—a passage which moves from praise to prayer and back to praise, in words which may well echo the language then current in liturgy. Then follows (iv, 1 ff.) exhortation to moral response to this gospel (again containing much that may have been current in the Christian church of the time, as instruction for inquirers and in public worship and preaching), which merges (at vi, 19) into a personal appeal for support in prayer, and (vi, 21–24) personal messages and farewell.

In the course of these approximately 150 verses, doctrinal and homiletic themes of great weight are included, such as: the "summing up" of God's agelong purpose in Christ (i, 10); Christ as the head of the church which is his body (i, 22, 23; iv, 12 ff.; *cf.* Col. i, 18, ii, 19, and contrast Rom. xii, 4–5, I Cor. xii, 12 ff.); the church as a witness to and a revealer of God's design (iii, 10), as the recipient of the gifts of the risen Christ which are intended to be manifested in all their diversity within a unified harmony (iv, 1 ff.) and as the bride of Christ (v, 25 ff.); the qualities of the Christian community (iv, 17 ff.), household (v, 22 ff.) and individual (vi, 10 ff.—the soldier in Christ's army equipped with divine armour and weapons for combat against supernatural forces of evil; *cf.* Isa. xi, 5, lix, 17 and Wisd. Sol. v, 17 ff.).

In sum, this splendid letter is aglow with the thought of the harmony actually achieved by God in Christ and waiting to be realized, implemented and entered into. It is filled with an optimism springing from the gospel and capable of facing titanic evil undismayed. It is often couched in the language of worship (v, 14 may even be a snatch from an early hymn).

Authorship and Origin.—Various problems attend the question of the book's origins.

Destination.—In i, 1 the destination "at Ephesus" occurs in the majority of manuscripts. But there is a strong case against the originality of this reading, which is omitted by important manuscripts, including one of perhaps the 3rd century, as well as by Origen's text and early texts known to Basil and to Jerome. Moreover, both Marcion (*c.* A.D. 150) and the so-called Marcionite prologue to the epistle described the epistle as "to the Laodiceans." Yet it is unlikely that the original opening sentence intended no indication of destination; Origen's attempt to construe it without one, yielding the sense "To the saints who are also faithful in Christ Jesus," as in the Revised Standard Version, is a tour de force.

Other explanations have therefore been sought. Archbishop James Ussher suggested in 1654 that there was a blank where the destination should have stood, and that the letter was intended for several destinations, beginning with Ephesus (to which it was carried doubtless by Tychicus; vi, 21), whence it was to be sent on to the others, each name being inserted, perhaps orally, as required. A modification of this theory would be to postulate a master copy with a blank space, from which copies were made, each with the relevant place name filled in. It is not impossible that Paul had facilities for such reproduction of copies, and it would help to account for the three variations—no name, "Ephesians," and "Laodiceans." In either case, if the letter were intended for a plurality of churches, this would help to explain why it is much less specific than, for example, Colossians, and contains no evidence of personal acquaintance between writer and readers. The letter to be passed on from Laodicea, referred to in Col. iv, 16, might be this encyclical on its way round.

Alternatively, it has been suggested that it was originally intended for some single destination other than Ephesus (had it been for Ephesus alone, it could scarcely have failed to show personal acquaintance with the church there), but that the address was for some unknown reason in some instances erased or changed to Ephesus. But how explain this change? The degeneracy of

Laodicea (Rev. iii, 14 ff.) is scarcely reason: if it comes to that, Ephesus was not blameless either (Rev. ii, 4–5). Besides, Marcion's description of it as "to the Laodiceans" may only have been a deduction from Col. iv, 16. Of the two explanations some form of Ussher's theory seems preferable.

A third possibility, that the writing is not a true Pauline letter at all, would place the whole question of the address in a different light (*see* below).

Possible Reference to Another Epistle.—A problem is presented by Eph. iii, 2–4, "assuming that you have heard of the stewardship of God's grace that was given to me for you, how the mystery was made known to me by revelation, as I have written briefly. When you read this you can perceive my insight into the mystery of Christ . . ." (Revised Standard Version). For "I have written," a more exact translation would be "I have already [or previously] written." To what, then, does this refer? Scarcely to the preceding chapter, for although the union of Jew and Gentile (the "mystery" in question—*see* iii, 6) is there mentioned, there is no reference to its being "revealed" to the writer; besides, why, in that case, "assuming that you have heard . . . ," and "when you read this"? It is more natural to refer it to a passage in some other epistle; *e.g.*, Col. i, 24–29 or Gal. i, 12. But if so, that seems to presuppose that the readers had access to the collected epistles, which would almost certainly place Ephesians after Paul's death. The alternative, of postulating a lost letter, is pure speculation. The problem thus remains for those who hold Pauline authorship.

Style, Vocabulary and Thought.—Specially close parallels between Ephesians and Colossians, combined with subtle differences, constitute a curious phenomenon. Eph. vi, 21–22 (reference to Tychicus) is almost identical with Col. iv, 7–8. But elsewhere what appear at first to be exact parallels turn out to be subtly distinct. For instance, in Eph. ii, 14 ff. and Col. ii, 14 the figure and construction are different; in Eph. ii, 16 and Col. i, 20–22 the uses of "body" are different. Is pleroma ("fullness") quite the same in both Eph. i, 23 and iv, 13 and Col. i, 19 and ii, 9, and is *mysterion* ("mystery") used identically in Eph. iii, 3, 9 and Col. i, 26 and ii, 2?

Then, too, the style of Ephesians, in many respects like that of Colossians, is notoriously more turgid and tautologous: there are a number of loosely attached genitives, sometimes nearly synonymous with the words they qualify (*e.g.*, i, 5, "the purpose of his will"; i, 11 "the counsel of his will," etc.); there are vast sentences (*e.g.*, i, 15–23), the great parenthesis at iii, 2–13, interrupting the sentence begun at iii, 1 and not resumed till iii, 14, and other peculiarities of style. There is nothing quite like it in the other Pauline Epistles. Liturgical echoes, as has been said, may partly be the cause.

Further, the actual vocabulary is distinctive. There are some 40 words not found elsewhere in the New Testament; some 40 more which, though used elsewhere in the New Testament, are not found in the acknowledged Paulines; and some 25 which Ephesians shares only with Colossians in the Paulines. There are distinct affinities with I Peter, and some with Luke-Acts and the Johannine literature, with the Pastoral Epistles and Hebrews. Specially striking is the use twice of the word *diabolos* ("devil"), whereas "Satan" (with other terms) is used in Romans, I and II Corinthians and I and II Thessalonians (in the Pastorals, both words occur); the use five times of the phrase "in the heavenly places" (nowhere else in the New Testament); and *halysis* for "chain" (Eph. vi, 20), whereas *desmos* is used in Philippians, Colossians and Philemon. On the other hand, it must not be forgotten that there are in Ephesians many words (such as "adoption") and phrases found elsewhere only in the Paulines and not elsewhere even in early Christian literature of the succeeding period.

More important even than language are differences of thought. The possibility that pleroma and *mysterion* are used differently has been mentioned already. The prayer that the recipients may be given "a spirit of wisdom and of revelation" (i, 17) and the exhortation to "be filled with the Spirit" (v, 18) are not easy to combine with the usual Pauline assumption that the Spirit is given with baptism (Rom. viii, 9, I Cor. xii, 13, etc., *cf.* Acts xix, 1–6). The conception of Christians as already saved (Eph. ii, 5, 8) is not exactly paralleled in the other Paulines. This is the only New Testament Epistle except Hebrews where Christ is said to have offered himself as a sacrifice to God (Eph. v, 2). Only here is it the church's function to make the manifold wisdom of God known to the supernatural powers (iii, 10). The church is built upon the foundation of the apostles and prophets (ii, 20; *cf.* Rev. xxi, 14; contrast I Cor. iii, 11, where Jesus Christ is the only foundation). The word ecclesia is used always for the entire church, never for a local congregation. Nowhere else is there so elaborate a terminology for angelic powers (ii, 2; vi, 10 ff.). The eschatology is expressed in a unique way—in terms of the growing up of the church to the full stature of Christ's totality (iv, 12–16). There is nothing apocalyptic.

Conclusions.—Advocates of Pauline authorship explain these phenomena by postulating that Ephesians was written at about the same time as Colossians but as a general, perhaps an encyclical, letter instead of (as Colossians) to meet the specific needs of a small group of churches. It is a great apologia for the church as spanning and transcending time and space. And it is urged that Paul need not have used the same terms twice over with mathematical precision and without any flexibility or variation in meaning; and that, in any case, the supposed differences have been exaggerated.

Those who cannot believe this have to resort to a theory of pseudonymity. One of the most specific of such theories is that of E. J. Goodspeed, attributing Ephesians to an admirer of Paul (Onesimus?) who had been led to collect the epistles and to write this glorious summary and recapitulation and development of the Pauline Gospel in the name of his master. In view of such uncertainty it appears that Ephesians must either be a Pauline encyclical or the writing of an unknown admirer, using his master's name with sincere intention to commend his doctrine.

Date.—The "captivity epistles," generally assumed until the modern critical period to have been written during the apostle's imprisonment in Rome, have been placed alternatively at Ephesus (in which case they could not include an epistle to the Ephesians) or Caesarea. For a brief discussion of the relative merits of these theories *see* COLOSSIANS, EPISTLE TO THE. The date, according to the imprisonment in question, will be roughly A.D. 52 or 62. If Paul was not the author, the date is doubtful but certainly later.

See also BIBLE; PAUL, SAINT.

BIBLIOGRAPHY.—Greek text ed. by S. D. F. Salmond with introduction and commentary (1903) in *Expositor's Greek Testament;* Greek text ed. by J. A. Robinson, with Eng. trans., introduction and commentary, 2nd ed. (1904; reprinted 1961); Greek text ed. by B. F. Westcott with Vulgate, Eng. trans. by Wycliffe and by W. Tyndale and introduction and commentary (1906); introduction and commentary on the Greek text by T. K. Abbott (1897) in *International Critical Commentary;* Eng. trans. with introduction and commentary by E. F. Scott (1930) in *Moffatt New Testament Commentary,* by W. Lock (1919) in *Westminster Commentaries* and by F. W. Beare and T. O. Wedel (1953) in *Interpreter's Bible;* French trans. with introduction and commentary by C. Masson (1953) in *Commentaire du Nouveau Testament;* German trans. with introduction and commentary by E. Haupt (1902) in *Meyers Kommentar,* by M. Meinertz, 4th ed. (1931), in *Die heilige Schrift des Neuen Testaments,* by M. Dibelius, 3rd ed. (1953) in H. Lietzmann (ed.), *Handbuch zum Neuen Testament,* and by H. Schlier, 2nd ed. (1958).

Studies: E. J. Goodspeed, *New Solutions to New Testament Problems* (1927), *The Meaning of Ephesians* (1933), *The Key to Ephesians* (1956); W. L. Knox, *St. Paul and the Church of the Gentiles,* ch. ix (1939); J. Knox, *Marcion and the New Testament* (1942), *Philemon Among the Letters of Paul,* 2nd ed. (1959); S. Hanson, *The Unity of the Church in the New Testament, Colossians and Ephesians* (1946); C. L. Mitton, *The Epistle to the Ephesians* (1951); F. L. Cross (ed.), *Studies in Ephesians* (1956); H. Chadwick, "Die Absicht des Epheserbriefes," *Zeitschrift für die neutestamentliche Wissenschaft* 51:145–153 (1960); H. Schlier, *Christus und die Kirche im Epheserbrief* (1930). *See* also works by G. S. Duncan, E. Percy, J. Coutts and H. J. Cadbury in bibliography to COLOSSIANS, EPISTLE TO THE. (C. F. D. M.)

EPHESUS, an ancient Ionian city the ruins of which lie near the modern village of Seljuk in the Izmir *il* (province) of Turkey (*see* also IONIA). In Roman times it was situated on the northern slopes of the hills Coressus and Pion and south of the Cayster (Küchük Menderes) river, the silt from which has formed a fertile plain but has driven the coast line ever farther west. The

temple of Artemis, or Diana, to which Ephesus owed much of its fame and which seems to mark the site of the classical Greek city, was probably on the seaboard when founded, about 600 B.C., 1 mi. E.N.E. of Pion (modern Panajir Dagh). But in Roman times a sea channel was kept up with difficulty to a harbour well west of Pion. In late Byzantine times this in turn became useless, and the coast by the middle of the 20th century was 3 mi. farther west. Ephesus commanded the west end of one great trade route into Asia, that along the Cayster valley, and had easy access to the other two, along the Hermus (Gediz) and the Maeander (Menderes) rivers.

History.—The first inhabitants mentioned by Greek writers are the "Amazons," with the Leleges, Carians and Pelasgians. But Greek Ephesus dates from the colony of Androclus, the son, according to legend, of the Athenian king Codrus. He is said to have arrived in 1087 B.C., and his descendants formed the privileged family of the Basilidae. But his followers were mixed, neither celebrating the chief Ionian festivals nor divided into the usual four Ionian tribes. The deity of the city was Artemis, not the normal Greek huntress but closely related to the primitive goddess of nature and worshiped as the productive power of the earth, the virgin mother of all life. Her cult statue, known from a copy and from coins of about 40 B.C. onward, showed a figure with many breasts, swathed below the waist in graveclothes. It perhaps interpreted the earliest cult object, a meteorite ("the sacred stone that fell from the sky"; Acts xix, 35). Tapering sharply toward the foot, it resembled a bee. The chief priest was a eunuch and was always called Megabyzus. Ortygia, near Ephesus, claimed, as against Delos, to be the birthplace of Artemis and Apollo.

Ephesus enters history with the Cimmerians in the mid-7th century B.C. Unlike its neighbour Magnesia, it survived their attack. Statements that they burned the temple of Artemis are very late. For part of the early 6th century it was under tyrants. Though allied by marriage to the kings of Lydia, they could not hold back the Lydian Croesus, who asserted a general suzerainty over the city. He did, however, present many columns and some golden cows for a new and splendid rebuilding of the temple of Artemis. At this time, according to Strabo, the Ephesians began to live in the plain; and to this period, too, should be set the redrafting of the laws, said to be the work of an Athenian, Aristarchus. Ephesus soon submitted to Cyrus of Persia, and its role in the Ionian revolt (500–493 B.C.) was fitful. Although it served as a base for the Ionian attack on Sardis, it is not mentioned again until Ephesians massacre the Chiot survivors of the battle of Lade; the massacre may have occurred because Ephesus was a commercial rival of the chief rebels, Chios and Miletus. Even in 478 Xerxes, returning from his failure in Greece, honoured Artemis of Ephesus, although he sacked other Ionian shrines, and left his children for safety in Ephesus. Themistocles landed there in the 460s on his flight to Persia. But Ephesus appears as a regular tributary of Athens after 454. Great Ephesians up to this time had been Callinus, the earliest Greek elegist (mid-7th century B.C.), the satirist Hipponax and the famous philospher Heraclitus (*q.v.*), one of the Basilidae.

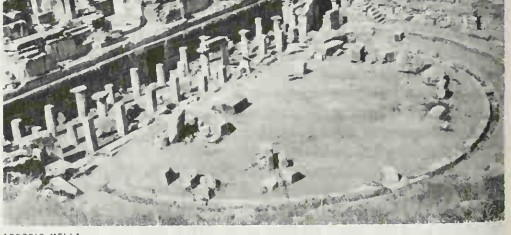
Ephesus shared in the general revolt of 412 B.C. against Athens and remained an effective ally of Sparta down to the end of the Ionian War. Threatened by Persia after 403, it served in 396 as the headquarters of Agesilaus. In 394 the Ephesians deserted to Conon's anti-Spartan maritime league, but by 387 the city was again in Spartan hands and was handed by Antalcidas to Persia. There followed the pro-Persian tyranny of Syrphax and his family, who were stoned to death in 333 on Alexander the Great's taking the city. After 50 years of fluctuating fortune, Ephesus was conquered by Lysimachus (*q.v.*) and resettled around Coressus and Pion (286–281 B.C.). Lysimachus introduced colonists from Lebedus and Colophon and renamed the city after his wife, Arsinoë—a name soon dropped. This was the beginning of Ephesus' Hellenistic prosperity. It became conspicuous for the abundance of its coinage. After the defeat of Antiochus the Great, king of Syria, by the Romans in 189 B.C., Ephesus was handed over by the conquerors to the king of Pergamum. Attalus III of Pergamum bequeathed Ephesus with the rest of his possessions to the Roman people (133 B.C.).

Henceforth Ephesus remained subject to Rome except for some years after 88 B.C., when, at the instigation of Mithradates the Great of Pontus, the cities of Asia Minor revolted and killed their Roman residents. The Ephesians even killed those Romans who had fled for refuge to the Artemision; notwithstanding which they returned in 86 B.C. to their former masters. Their claim, preserved on an extant inscription, that in admitting Mithradates they had merely yielded to superior force was rudely brushed aside by Sulla, who inflicted a very heavy fine. Although it twice chose the losing side in the Roman civil wars, and although it was stoutly opposed by Pergamum and Smyrna, Ephesus became under Augustus the first city of the Roman province of Asia. Strabo called it the biggest emporium this side of Taurus; years later Aristides referred to it as the bank of Asia. The triumphal arch of 3 B.C. and the aqueduct of A.D. 4–14 initiated that long series of public buildings, ornamental and useful, which make Ephesus the most impressive example in Greek lands of a city of imperial times.

Meanwhile the Christian Church began to win converts. The famous protest in the theatre against St. Paul, described in Acts xix, is dated about A.D. 57. According to local belief Ephesus was the last home of the Virgin, who was lodged near the city by St. John and died there. The tradition that St. Luke also died there seems to be less strongly supported. Ephesus was one of the seven churches of Asia to which the Book of Revelation was addressed. (For St. Paul's Epistle to the Ephesians, *see* EPHESIANS, EPISTLE TO THE.)

The Goths destroyed both city and temple in A.D. 262, and neither ever recovered its former splendour. Constantine, however, erected a new public bath and Arcadius rebuilt at a higher level the street from the theatre to the harbour, named after him, the Arkadiane. A general council of the church, held at Ephesus in 431 in the great double church of St. Mary, condemned Nestorius and justified the cult of the Virgin as *theotokos*, "mother of God" (*see* COUNCIL). A few years later the Seven Sleepers are said to have awakened and expired (*see* SEVEN SLEEPERS OF EPHESUS). They too became the object of a famous cultus. Presented by Justinian with a magnificent basilica of St. John, the city partook of the subsequent Byzantine decay. The sea retired westward, life withdrew eastward, and late Byzantine Ephesus, conquered by the Seljuks in 1090, was merely a small town. After brief splendour in the 14th century even this was deserted, and the true site of the Artemision remained unsuspected until 1869.

Excavations and Extant Remains.—J. T. Wood, working at Ephesus for the British museum between 1863 and 1874, excavated the odeum and theatre. In May 1869 he struck a corner of the Artemision. His excavation exposed to view not only the scanty remains of the latest edifice, built after 350 B.C., but the platform below it of an earlier temple of identical size and plan, subsequently found to be that of the 6th century, to which Croesus contributed. The sculptured fragments of both temples were sent to the British museum. In 1904 D. G. Hogarth, heading another mission from the museum, examined the earlier platform and found beneath its centre the remains of three yet older structures.



ARBORIO MELLA

THE THEATRE AT EPHESUS

In its earliest-known phase the temple was apparently a small platform of green schist, containing a sealed deposit of primitive coins and other objects. Hogarth's deep digging left the site liable to floods, but Oxford scholars working over his publication have concluded that platform and deposit cannot be much earlier than 600 B.C. It is impossible to assign the various architects named by ancient authors to the respective phases of the temple. At best Chersiphron and Metagenes can be tentatively assigned to the temple of Croesus, "Chirocrates" or "Dinocrates" to that of the 4th century. There had perhaps been some repairs toward 400 B.C., associated with the architects Paeonius and Demetrius and with the prize-winning dedicatory hymn of the famous musician Timotheus. The Artemision passed rapidly through three phases before *c.* 550 B.C. The temple of Croesus (the fourth phase) was remarkable for its great size (it was more than 300 ft. long and 150 ft. wide), for the carved figures round the lower drums of its columns (Pliny's *columnae caelatae*) and for the smaller but elaborate figured friezes along its roof gutter (*sima*). Croesus' temple seems to have been burned down in 356 B.C. The new temple built shortly afterward copied the old in its *columnae caelatae*, one of which was by Scopas, but the new *sima*, instead of small, crowded figures, had a more conventional, if vigorous, *rinceau* ornament. The cella contained, among other great works, the Amazons of Polyclitus, Phidias and Cresilas.

"Lysimachean" Ephesus has been continuously excavated since 1894 by the Austrian Archaeological institute, but so solid and extensive is the Roman town that by the early 1960s the Austrians had rarely penetrated to Hellenistic levels.

On the hill of Ayasoluk (Hagios Theologos) is Justinian's church of St. John the Theologian (the apostle John's title in the Orthodox Church), built around a shrine variously associated in the early middle ages with the death or bodily assumption of St. John. The church, uncovered since 1922, is a noble structure but badly restored. On the hill there is also a beautiful Seljuk mosque dedicated in 1375.

The public buildings of the city are arranged in a rectangular street pattern going back to Hellenistic days. They include the theatre, capable of seating nearly 25,000 spectators and completed in its present form under Trajan; the agora surrounded by stoas dating from the time of Severus; the library of Celsus, also Trajanic and well known because of its façade; and an immense array of baths and gymnasia.

All these buildings are to the west of Pion. On its north side is the stadium and north of this the gymnasium of P. Vedius Antoninus, relatively small but very complete and with a notable chapel for the cult of Antoninus Pius. South of Pion were the odeum, another gift of Vedius, a roofed semicircular theatre to hold 1,400 persons; also a series of fountains and aqueducts, notably the aqueduct of C. Sextilius Pollio which crossed the valley from Coressus. The unmortared city wall along the crest of Coressus appears to be that of Lysimachus.

Of the early Byzantine city, besides the stretch of curtain wall on Panajir Dagh, there remain the ruined church of the Seven Sleepers to its east and the long double basilica of the Virgin, the scene of the council, to its west. This was rebuilt several times. It was around this building, between the great gymnasia and the stadium of the classical city, that the early Byzantine Ephesians largely gathered. *See* also references under "Ephesus" in the Index.

BIBLIOGRAPHY.—J. T. Wood, *Discoveries at Ephesus* (1877); E. L. Hicks, *The Collection of Ancient Greek Inscriptions in the British Museum*, part iii, 2 (1890); Sir W. M. Ramsay, *The Letters to the Seven Churches* (1904); D. G. Hogarth, *Excavations at Ephesus: the Archaic Artemisia*, (1908): Oesterreichisches Archäologisches Institut, *Forschungen in Ephesos* (1906–); W. R. Lethaby "The Earlier Temple of Artemis at Ephesus," *Journal of Hellenic Studies* (1917); P. Jacobsthal, "The Date of the Ephesian Foundation—Deposit" and E. S. G. Robinson, "The Coins From the Ephesian Artemision Reconsidered," *Journal of Hellenic Studies* (1951); B. L. Trell, *The Temple of Artemis at Ephesus* (1945); J. Keil, *Ephesos: ein Fuehrer durch die Ruinenstatte*, 4th ed. (1957). (WI. H. P.)

EPHIALTES (d. 462/461 B.C.) was leader of the radical democrats at Athens in the 60s of the 5th century B.C. who by his reforms paved the way for the final stages of Athens' democratic development. Cimon (*q.v.*), supported by a powerful group of nobles had dominated Athenian politics since the repulse of the Persians and had developed the Delian League (*q.v.*) while maintaining in a conservative democracy friendly relations with Sparta. Ephialtes like Themistocles (*q.v.*) believed that friendship with Sparta was incompatible with Athenian growth and that the energies of the common people should have wider scope. He was elected general soon after 465 and in 462, when Sparta appealed for help against the Helots, he was opposed to sending help but was outvoted and an expedition under Cimon was sent to besiege the Helots at Ithome. When this Athenian contingent was dismissed from Ithome opinion swung in his favour. In the same year 462/461, during Cimon's absence with the hoplites, he carried reforms which gave much fuller power to the common people. The senior council of the Areopagus was stripped of its political power and the democracy was made to rest more squarely on the popular assembly or *ecclesia* (*q.v.*), the *boule* (*q.v.*) and the law courts, whose juries were now reformed and paid. These reforms were bitterly opposed by the more extreme oligarchs. Ephialtes was assassinated, but his political revolution was consolidated.

(R. ME.)

EPHOD, part of the ceremonial dress of the high priest of Israel in the postexilic period, described in Ex. xxviii, 6–8; xxxix, 2–5. It was worn outside the robe and kept in place by a girdle and by shoulderpieces (?) from which hung the "breastplate" (or rather pouch) containing the sacred lots, Urim and Thummim. It is uncertain whether it covered the back, encircling the body like a kind of waistcoat, or only the front; at all events it was not a garment in the ordinary sense, and its association with the sacred lots indicates that the ephod was used for divination (*cf.* Num. xxvii, 21). A similar vestment, made of linen, was worn by persons other than the high priest; *e.g.*, by Samuel in serving before the tabernacle at Shiloh (I Sam. ii, 18) and by David as he danced before the ark at its entry into Jerusalem (II Sam. vi, 14). *See* also VESTMENTS, ECCLESIASTICAL: *Jewish Vestments.*

See W. R. Arnold, *Ephod and Ark,* Harvard Theological Studies, vol. iii (1917).

EPHOR (Gr. *ephoros*) in ancient Sparta was the title of the highest magistrates apart from the kings. The senior ephor gave his name to the year, like the Athenian archon (*q.v.*) and in antiquity there existed a list of these eponymous ephors reaching back to 754 B.C., the authenticity of which cannot be checked. Their institution was sometimes ascribed to Lycurgus (*q.v.*); others said they were appointed during the Messenian Wars (*see* SPARTA) to take over some duties from the absent kings: there seems to have been no firm tradition in the matter. It was normal for Greek states in their aristocratic phase to set up executive magistrates as well as the king, or replacing him, and this is probably the origin of the ephorate, though the kings retained more effective power at Sparta than elsewhere. In classical times an oath was sworn monthly, by the kings that they would observe the laws, by the ephors on behalf of the city that on this condition they would maintain the kings' authority. This exemplifies the relation between the two offices, and may throw light on the origin of the ephorate.

They were five in number, elected annually, every full citizen being eligible. Their term began with the Spartan calendar year, near the autumn equinox. With the kings, and to some extent in rivalry with them, they formed the main executive of the state. They presided over meetings of the council (*gerousia*) and assembly, and were responsible for the execution of their decrees. They received foreign ambassadors in the first instance. Their police powers were extensive, especially over the helots, on whom they declared war at the opening of their year as advance justification for the measures they might take against them. They enforced Sparta's occasional expulsions of foreigners (*xenelasia*). They might in emergency arrest the king and imprison him pending trial; and they were members, with the *gerousia* and the other king, of the court which conducted such trials. The trial of civil suits between Spartans was distributed among them. When a military expedition was decreed, they determined what age-classes should be called up and when the expedition should set out; two ephors

accompanied a king on campaign, but might not interfere with his orders while the campaign lasted. They had also certain financial and administrative duties.

These were extensive powers for Greek magistrates. Ancient observers sometimes exaggerated them, and they have played too sensational a part in some modern histories of Sparta. Annual election ensured that they reflected to some extent the mood of the assembly, and Aristotle's judgment, that they formed a democratic element in the constitution, seems just. The most famous of them was Chilon in the middle of the 6th century B.C., who was reckoned as one of the Seven Wise Men of Greece and was associated with the expulsion of tyrants from certain other cities.

BIBLIOGRAPHY.—References to the ancient authorities, too numerous to be given here, are collected in G. Busolt and H. Swoboda, *Griechisches Staatsrecht*, vol. ii, pp. 683–690 (1926), with a fairly sober account of the office. (A. As.)

EPHORUS (*c.* 405–330 B.C.), Greek historian, was the author of the first universal history. He was born at Cyme in Aeolis, Asia Minor. The tradition that, together with Theopompus, he was the pupil of Isocrates, has been rejected by some critics, but without sufficient reason. His *Historiai* were divided into 29 books beginning with the return of the Heracleidae to Peloponnesus, and ending with the siege of Perinthus (340) by Philip of Macedon. A 30th book, having as its centrepiece the Sacred War of 356–346, was added by his son Demophilus, who edited the whole work. Ephorus was the first historian to divide his work into books, to each of which he wrote a preface. He treated his material under subject headings, and not according to years. He was used as a source by Diodorus Siculus, whose chronological blunders arise partly in trying to reproduce him in annalistic form. Polybius, while crediting him with a knowledge of naval warfare conditions, ridicules his description of the battles of Leuctra and Mantineia as showing ignorance of the nature of land operations. He usually (though not always) distinguished clearly between the mythical and historical; he even recognized that detail, though it corroborates accounts of recent events, is ground for suspicion in reports of far-distant history. His style was artificial, and he frequently sacrificed truth to rhetorical effect, or even to his own bias. Other works attributed to him were *A Treatise on Discoveries; Respecting Good and Evil Things; On Remarkable Things in Various Countries* (it is doubtful whether these were separate works, or merely extracts from the *Histories*); *A Treatise on My Country*, on the history and antiquities of Cyme; and an essay *On Style*. The view once maintained (*e.g.*, by E. M. Walker) that he was the author of the famous historical fragment discovered at Oxyrhynchus has been rejected. (*See* H. Bloch, in *Harvard Studies in Classical Philology*, special volume, pp. 303 ff. [1940].)

BIBLIOGRAPHY.—Fragments in F. Jacoby, *Die Fragmente der griechischen Historiker* IIA and IIC, no. 70 (1926), with lengthy commentary and criticism partly superseding the article of E. Schwartz, in Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft* VI, i (1907). *See* also G. L. Barber, *The Historian Ephorus* (1935). (G. T. Gr.)

EPHRAEM (Syriac, APHREM) **SYRUS, SAINT** (EPHRAIM THE SYRIAN) (*c.* A.D. 306–373), is the most celebrated writer among the Syrian fathers. He was created a doctor of the church by Pope Benedict XV in 1920. He was born at Nisibis (southeast Turkey); his father according to Syrian sources was a pagan priest though Ephraem's writings suggest he was of Christian parentage. He was ordained deacon but never became a priest. The pupil of St. James, bishop of Nisibis, he may have accompanied his master to the Council of Nicaea (325). When the emperor Jovian ceded Nisibis to the Persians in 363, Ephraem settled at Edessa and devoted himself to teaching and writing. He died there on June 9, 373. His feast day, formerly Feb. 1, is June 18 in the west, Jan. 28 in the Eastern Church.

His biography, written soon after his death, has survived only in two later recensions full of miraculous anecdotes and unlikely details. Among these is the account of an eight years' visit to Egypt refuting Arianism and a visit to Caesarea of Cappadocia, whose bishop, St. Basil, he is reputed to have met and to have eulogized on his death. Syriac writers give credit to Ephraem for the deliverance of Nisibis which was besieged by Shapur II in 338.

A prolific writer, popular with his contemporaries, though to modern taste his verse is tedious, repetitious and overfull of metaphor and allegory, Ephraem produced Syriac hymns, homilies and biblical commentaries, of which early versions in Greek, Armenian, Coptic, Arabic and Ethiopic are often extant. His hymns, many in his favourite seven-syllable metre, deal with such themes as the Nativity, Epiphany, Crucifixion and Passover or are directed against heretics and sceptics. Important for the historian are the *Carmina Nisibena*, some 70 hymns, 21 of which were composed at Nisibis between 350 and 363 and deal with the bishops of that city and its history during the Persian invasion, while the remainder were composed at Edessa between 363 and 373. Notable among the homilies are the refutations of Mani, Marcion and Bardaisan. Of his commentary on the Old Testament only the sections on Genesis and a large part of Exodus are extant. The rest is accessible only as cited by later writers, notably in the *Catena Patrum* of Severus (861). Of his commentary on the *Diatessaron* (*i.e.*, the four gospels in a continuous narrative) only the Armenian version was known until in the late 1950s it was reported that a considerable portion of this work in Syriac was extant in the library of Sir Chester Beatty. It was being prepared for publication by L. Leloir, who had already edited a collection of the biblical quotations in Ephraem's works.

Though Ephraem had a knowledge of Greek and Hebrew and knew the Jewish haggada, he was less theologian than preacher and his fame rests more on his sanctity than on his originality.

BIBLIOGRAPHY.—A. Baumstark, *Geschichte der syrischen Literatur* (1922); C. W. Mitchell, *St. Ephraim's Prose Refutations of Mani, Marcion and Bardaisan* (1912–21). In *Corpus scriptorum Christianorum orientalium* see L. Leloir, *Commentaire de l'évangile concordant, version arménienne éditée et traduite* (1953–54) and *L'évangile d'Ephrem d'après les oeuvres éditées: recueil des textes* (1958); R. M. Tonneau, *Sancti Ephraem Syri in Genesim et in Exodum commentarii* (1955); E. Beck, *Des Heiligen Ephraem des Syrers Hymnen . . . de fide* (1955) . . . *contra haereses* (1957) . . . *de paradiso und contra Julianum* (1957). (W. D. McH.)

EPHRAIM, one of the 12 tribes of Israel, the stronger of the two (the other being Manasseh) into which the house of Joseph was divided. Ephraim included the famous sites of Shechem, Shiloh, Timnath-serah and Samaria. *See* TWELVE TRIBES OF ISRAEL.

EPICHARMUS (*c.* 530–*c.* 440 B.C.), Greek poet, was, according to the Suda lexicon, the originator of Sicilian (or Dorian) comedy. He was born in either Syracuse or (Sicilian) Megara or Cos about 530 B.C. and died about 440. He has been credited with more than 50 plays, the titles of 35 of which survive; but remains of his works are scanty. Many of his comedies were obviously mythological burlesques, as can be seen from such titles as *Odysseus Automolos* ("Odysseus the Shirker"), *Cyclops*, *Sirens*, etc. The gods are not spared: in *Hebes Gamos* ("The Marriage of Hebe") Poseidon appears as a fishmonger, and in *Busiris* the greed of Hercules is satirized. Many of Epicharmus' comedies must have contained, or even consisted in, a set debate, as *Logos kai Logina* ("Male and Female Reason"), *Elpis e Ploutos* ("Hope or Wealth") and *Ga kai Thalassa* ("Earth and Sea"); some have titles which suggest the stock characters of Middle or New Comedy, such as Agrostinos ("The Countryman"); others again, such as *Persai* ("The Persians") or *Philoctetes*, suggest parodies of tragedy. Though they seem to have had some musical accompaniment the plays had no chorus; such plural titles as *Thearoi* ("The Envoys") refer to actors. Epicharmus' style was lively, and his comedies seem more akin to mime and to New Comedy than to Old Comedy. They were apparently short and in the main farcical, but contained an odd admixture of philosophical moralizing. Epicharmus himself was often regarded as a philosopher. Ancient authorities attributed to him works on ethics and medicine, and tradition made him the pupil of the philosopher Pythagoras.

BIBLIOGRAPHY.—A. O. F. Lorenz, *Leben und Schriften des Koers Epicharmos* (1864); G. Kaibel, *Comicorum Graecorum Fragmenta* (1899) and the article in Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft* (1907); A. W. Pickard-Cambridge, *Dithyramb, Tragedy, and Comedy* (1927); G. Norwood, *Greek Comedy* (1931). (M. Pr.)

EPIC POETRY or EPOS, terms rather loosely used to designate a widely extant form of narrative verse conspicuous for its

length and its elevated heroic mood. No more exact definition is possible, because epic varies markedly from age to age and language to language according to native literary conventions, interests and talent. There is, however, an important distinction to be drawn between oral epic—poetry of heroic incident or legend composed without aid of writing, recited and not read, transmitted by memory from poet to poet and often from generation to generation—and the epic of the written manuscript and the printed book, of quite different character because created under other formative influences.

As particularly outstanding examples of this second, or written, class of epic may be cited: in Latin, Virgil's *Aeneid* and Lucan's *Pharsalia;* in medieval Greek, *Digenes Acritas;* in medieval French, the *Chanson de Roland;* in modern Italian, Ariosto's *Orlando Furioso* and Tasso's *Gerusalemme Liberata;* in Spanish, the *Chronicle of the Cid;* in Portuguese, Camoens' *Lusiads;* in English, Spenser's *Faerie Queene* and Milton's *Paradise Lost.* Detailed discussions of the character, reputation and literary merit of these epics whose authors are named will be found in the biographies of the authors. For further information on *Digenes Acritas* see DIGENES AKRITAS, BASILIUS; for the *Chanson de Roland* see ROLAND, CHANSON DE; for the *Chronicle of the Cid* see SPANISH LITERATURE: *Castilian Literature.*

It is less easy to acquire information on epic of the unwritten class, since its authors are often unknown, its preservation has been fitful and uncertain and critical interest in it has been a comparatively recent event. However, an extensive survey of oral epic exists in H. M. and N. K. Chadwick's *The Growth of Literature,* 3 vol. (1932–40), from which the vastness of the topic will become apparent.

Certain traits are very generally distinctive of epic poetry in its oral state and are frequently carried over into the written form. Among these are the exaggeration of minor events to heroic proportions, the projection of unrelated historical characters and happenings upon a common scene and the fusion of remembered incident with imaginary accessories drawn from legend and myth in an inextricable patchwork of fact and fancy. The elevation of minor incident to heroic proportions is well illustrated by the *Chanson de Roland,* originally in the oral tradition before it was reduced to text. Conflation of temporal setting can be proved for the *Nibelungenlied,* which introduces Attila the Hun and Theodoric the Ostrogoth as contemporary figures, though in fact the former died in the year preceding the latter's birth. Fusion of history with folk tale is well illustrated by *Beowulf,* where the Danish setting is veridical but most of the incident is derived from fairy tale. One of the finest of the Icelandic epics (these defy the usual definition of epic by being narrated in prose), the *Saga of Grettir the Strong,* combines precise topography, authentic genealogy and historic event with sheerly impossible fantasy drawn from folk tale.

Such heroic exploits, mingled with adventures from myth and legend, naturally become the favoured theme for epic recitation, whose measured cadences and memorable phrases could keep past glory alive for the edification of the present and the emulation of the future. Presumably it was this lofty and serious subject matter which everywhere caused epic to express itself in a dignified and poetical vocabulary framed in many-syllabled lines or recurrent stanzas. Whatever similarity underlies the world's oral epic is hence ascribable to its identity of spiritual mood and cultural interest rather than to any diffusion of literary traditions.

As knowledge of writing spread through the world, oral epic tended to give place to the more strictly literary form. This conversion is strikingly embodied in the Homeric poems; there is now universal agreement among scholars that they were originally oral compositions which were subsequently reduced to writing. How and when and with what changes this redaction was accomplished remains today's still unsolved residue of the redoubtable Homeric question (*see* HOMER). With the written genus once established, all subsequent epic in Greek or Latin, and much composed in other tongues, was profoundly influenced by the *Iliad* and *Odyssey,* spiritual ancestors of most of the western world's conception of this form of literature.

None of the literary epics has rivaled the Homeric poems in those qualities which belong by inherent right to oral, as opposed to printed, composition. Conversely, it is unreasonable to demand of Homer the peculiar literary virtues which the printed book brings into being—the tense brevity of felicitous phrase or adjective, the complex description of studied narrative, the balanced structure of paragraph and page, the personal intrusion of the author into his work. In all of these, Virgil's *Aeneid* contrasts strikingly with Homer, despite the Latin poet's unmistakable dependence on the *Iliad* and *Odyssey.* The effect is un-Homeric, if only because a close inspection of every word and every verse, with a critical consciousness of each epithet and phrase, has been substituted for the larger and looser movement of oral recitation.

Yet all written epic bears some trace of an underlying oral cast, partly because of Homer's unshakable authority, and partly because the ultimate origin of the genus has everywhere been oral recitation. But by continuous adaptation to varying purposes and ideas, and the transmutations due to individual genius, the range and modality of epic have been expanded to include such extreme variety as that evidenced in, for example, *The Faerie Queene* and *Paradise Lost.*

Finally, epic exists also in a serio-comic or outright burlesque form, of which the pseudo-Homeric *Battle of Frogs and Mice* is the earliest example, with the 15th-century Italian poet Pulci responsible for its more modern vogue.

Further information related to the epic may also be found in the articles CHANSONS DE GESTE; EDDA; HESIOD; HOMER; LATIN LITERATURE and SANSKRIT LITERATURE, and in the separate entries for the various authors, such as (in addition to those previously mentioned) APOLLONIUS; STATIUS, PUBLIUS PAPINIUS; VALERIUS FLACCUS, GAIUS; SILIUS ITALICUS; and BOIARDO, MATTEO MARIA. *See* also references under "Epic Poetry" in the Index.

BIBLIOGRAPHY.—The literature dealing with epic poetry is enormous but confines itself almost exclusively to the written and printed form; extensive bibliographies on this form of epic will be found in the various articles referred to in this article. On oral epic, in addition to the previously mentioned treatise by H. M. and N. K. Chadwick, *The Growth of Literature,* 3 vol. (1932–40), attention may be directed to H. M. Chadwick, *The Heroic Age* (1912); the opening chapter of C. M. Bowra, *From Virgil to Milton* (1945); Rhys Carpenter, *Folk Tale, Fiction and Saga in the Homeric Epics* (1958). On the general theme, W. P. Ker, *Epic and Romance* (1897) remains authoritative. (R. C.)

EPICTETUS (born *c.* A.D. 50), Greek philosopher, the religious tone of whose teaching commended him to many early Christian thinkers. He was probably from Hierapolis in Phrygia. The name Epictetus is the Greek for "acquired"; his original name is not known. As a boy he was a slave, but managed to attend the lectures of the Stoic Musonius Rufus; subsequently he became a freedman. He was lame and weak. In 90 he was expelled from Rome with other philosophers by Domitian, who was irritated by the encouragement which the opposition to his tyranny found among the adherents of Stoicism. The rest of his life he spent at Nicopolis in southern Epirus.

Epictetus wrote nothing; his teaching was transmitted by a pupil, Arrian (the historian of Alexander the Great), in two works: *Discourses,* of which four books are extant; and the *Manual* or *Encheiridion,* a condensed aphoristic version of the main doctrines. In his teaching, Epictetus followed the early rather than the late Stoics and reverted to Socrates and to Diogenes as historical models of the sage.

Primarily interested in ethics, Epictetus described philosophy as learning "how it is possible to employ desire and aversion without hindrance" (*Discourses,* iii, 14, 10). True education lies in recognizing that there is only one thing which is fully our own—that is, our will or purpose. God, acting as a good king and father, has given us a will which cannot be compelled or thwarted by anything external. We are not responsible for the ideas that present themselves to our consciousness, but we are wholly responsible for the way in which we use them. "Two maxims," he says, "we must ever bear in mind—that apart from the will there is nothing good or bad, and that we must not try to anticipate or to direct events, but merely to accept them with intelligence." We must, in short,

believe that there is a God whose thought directs the universe.

Man is a member of a great system, which comprehends God and men. Each human being is primarily a citizen of his own commonwealth; but he is also a member of the great city of gods and men, whereof the city political is only a copy. All men are the sons of God, by virtue of rationality, and kindred in nature with the divinity. Hence man can enter into the method of divine administration, and thus can learn—and it is the acme of his learning—the will of God, which is the will of nature. The natural instinct of animated life, to which man also is subject, is self-preservation and self-interest. But men are so constituted that the individual cannot secure his own interests unless he contributes to the common welfare. The aim of the philosopher, therefore, is to see the world as a whole, to grow into the mind of God and to make the will of nature his own.

There are editions of the works of Epictetus by J. Schweighäuser, 5 vol. (1799–1800) and by H. Schenkl (1894; new ed., 1916); a text with Eng. trans. by W. A. Oldfather, Loeb library (1926); an Eng. trans. by G. Long (1877) and by P. E. Matheson (1916). *See* also STOICS.

BIBLIOGRAPHY.—A Bonhöffer, *Epiktet und die Stoa* (1890), *Die Ethik des Stoikers Epiktet* (1894) and *Epiktet und das Neue Testament* (1911); T. Zahn, *Der Stoiker Epiktet und sein Verhältnis zum Christentum* (1895); T. Colardeau, *Étude sur Epictète* (1903); D. S. Sharp, *Epictetus and the New Testament* (1914); A. Jagu, *Épictète et Platon* (1946); F. Ueberweg, *Grundriss der Geschichte der Philosophie,* Teil 1, *Das Altertum,* 12th ed. by K. Prächter (1926).

(A. R. C. D.)

EPICURUS (341–270 B.C.), Greek philosopher, was notable for his exposition of the atomistic theory of physics; but he was especially celebrated for an ethical teaching that in antiquity gave him the reputation of a saviour of souls, and still entitles him to be respected as a great moral reformer. He was born in 341 B.C. in Samos, where his father, Neocles, an Athenian, had settled about ten years before. Epicurus went to Athens in 323, but shortly afterward joined his father in Colophon. Thereafter he taught in Mytilene and in Lampsacus before returning to Athens in 306. In Athens he then established his permanent home and founded a school of philosophy, teaching his pupils in his garden (near the Dipylon), which he bequeathed to the school under his will. As the society which he gathered round him included women as well as men, some scandalous stories were told about it. The regimen of the school, however, was characterized by an attractive simplicity; and both his country and his friends testified to the philosopher's unsurpassed kindness to all.

Writings and Testimonies.—Diogenes Laërtius devoted book x of his *Lives and Opinions of the Famous Philosophers* to Epicurus. He describes Epicurus as a most prolific writer and gives a list of his chief works. The same Diogenes also preserves the text of three letters of Epicurus and the *Kuriai Doxai* ("Principal Doctrines"). The three letters are (1) *To Herodotus,* dealing with physics; (2) *To Pythocles* (probably a disciple's abridgment of an authentic work of Epicurus, but no doubt reflecting the master's doctrine), on meteorology; and (3) *To Menoeceus* (in some points designed to present a clear contrast to Aristotle's *Protrepticus*), on ethics and theology. The *Kuriai Doxai* consists of a series of 40 short aphoristic statements illustrative of Epicurean doctrine.

Another major source of knowledge on Epicurus and his doctrine is the great collection of papyrus rolls from the library of the so-called "Casa dei Papiri" or "Villa of the Epicurean," discovered at Herculaneum in the years 1752–54. These papyri include not only parts of Epicurus' great work *On Nature* (originally in 37 books), but also numerous fragments of his correspondence with his friends, of great interest for the evaluation of Epicureanism (especially papyrus 176, ed. by A. Vogliano; and papyrus 1,418, ed. by C. Diano, *Lettere;* see *Bibliography,* below).

Two further textual sources are: (1) a 14th-century manuscript in the Vatican library (Codex Vaticanus Graecus No. 1,950), in which C. Wotke in 1888 discovered the so-called *Sententiae Vaticanae,* a series of 80 Epicurean aphorisms (some of them identical with those in the *Kuriai Doxai*); and (2) the fragments of letters and aphorisms which a certain Diogenes of Oenoanda in Lycia caused to be inscribed on the wall of a portico (2nd or 3rd century A.D.) for the instruction of his fellow citizens (discovered in 1884–85 and subsequent years; *Diogenis Oenoandensis fragmenta,* text ed. by J. William, Leipzig, 1907). Further very valuable information is supplied by Lucretius (*q.v.*), whose poem *De rerum natura* is a passionate exposition of Epicureanism; and there are also numerous quotations and references in the philosophical works of other ancient authors such as Cicero and Plutarch.

PHILOSOPHY

The philosophic outlook of Epicurus is fundamentally ethical, and his concern with other studies seems to have been small. Even his interest in physics is purely subsidiary to the ethical end; *i.e.,* to obtaining a theory of life which shall ensure "quietude of mind and a steadfast faith." According to Diogenes Laërtius, his philosophy is divided into three parts: "canonical," "physical" and "ethical," the first (dealing with "approaches to study"; *i.e.,* with criteria of validity) being usually coupled with the second.

Physics.—Diogenes Laërtius states that it was from the writings of Democritus that Epicurus received his first impulse to philosophy. Be that as it may, the doctrine of Democritus regarding the atoms and the void is the basis of the teaching of Epicurus.

Nothing is created out of nothing and nothing passes into nothing. The universe was always such as it is now and always will be such, since there is nothing beyond the universe which could enter it and effect a change. The universe consists of bodies and space (void). The existence of bodies is testified by sense perception, and the existence of space is a necessary inference of reason; for, if the void did not exist, bodies could not move as, in fact, sense assures us that they do move. Besides these two—bodies and space—nothing can even be imagined. Again, bodies are either compound or the elements of which greater bodies are compounded. These elements are indivisible (*atoma,* atoms) and unchangeable; for, when a compound is resolved into its elements, the elements must either pass into nothingness—which is held to be impossible—or remain as ultimate indissoluble entities. Again, the universe is infinite; for the finite has an extremity which can only be observed against something else, which is, by definition, impossible in the case of the universe. Having, then, no extremity, the universe has no limit. Also, the number of atoms and the extent of the void are infinite; for if there were a limited number of atoms in unlimited void, the atoms could not remain anywhere but would drift, scattered through infinite void, not having anything to support or place them by their collisions. If, on the other hand, the void were limited and the atoms unlimited, there would not be room for them.

The atoms are not all of one shape. The number of atoms of any one shape is absolutely infinite; but the number of different shapes, though incomprehensibly large, is not absolutely infinite.

The atoms vary in size but are not of all sizes: otherwise some would be visible. Whatever their size, they are in perpetual motion, moving all with the same velocity, swift as thought. Because of an inherent "swerve" (this, of course, is a necessary postulate) the atoms collide and rebound to a lesser or greater distance, thus forming compound bodies of greater or lesser density.

Psychology.—The soul is a material body of fine parts and is distributed through the whole bodily structure. Compound bodies are of two sorts: those which are capable of holding together of themselves; and those which cannot do so unless protected by some more solid body. The soul belongs to the latter sort. So long as it is protected by the body, it is capable of sensation and of communicating sensation to the body. When it leaves the body, it is dissipated; the body, on the other hand, is no longer capable of sensation.

Sense perception is a purely material process. From the surface of all bodies there are continually being discharged images, viz., hollow films of exceedingly fine texture, which in shape are an exact replica of the bodies. These images reach the soul through the various organs of sense and originate sensation. Sensation is the sole source of knowledge, and all sensuous perceptions are true. Error can arise only when, beyond what is given in sensation, the

mind forms an opinion which is afterward contradicted or unconfirmed.

According to Diogenes Laërtius (x, 31), "In the *Canon* Epicurus says that the criteria of truth are the sensations and concepts and the feelings; and the Epicureans add the imaginative apprehensions of the mind." This passage, with others akin to it, requires some interpretation. By "sensations" is meant that which is immediately given by the senses. By "concepts" is meant the general notion formed by repeated sense-perceptions. Diogenes tells us that by concepts the Epicureans denoted "an apprehension of right opinion or thought or general idea stored within the mind; that is, a recollection of what has often been presented from without" (x, 33); and goes on to emphasize that "the assertions of the many about the gods are not concepts but false suppositions." By the retention of concepts in the mind we are not only able to perceive an individual object as such, but to refer it to the general name. By "feelings" are meant, apparently, the "emotions," or, so to say, passive effects on the mind. Lastly the "imaginative apprehensions of the mind" seem best referred to the power of the mind to form mental images intuitively (for an interpretation of the relevant passages, *see* the appendix on the meaning of *epibole tes dianoias* in the edition by C. Bailey, pp. 259 ff.[Oxford, 1926]). The expression would thus be equivalent to the "immediate knowledge" by which the gods are known (*Letter to Menoeceus*, 123). The best commentary on the whole matter is provided by no. 24 of the *Kuriai Doxai* (Diog. Laërt., x, 147): "If you reject any sensation absolutely, and do not distinguish opinion in regard to that which awaits confirmation from that which is already given by sensation and feelings and every imaginative apprehension of the mind, you will confound all other sensations as well and so reject every criterion."

Ethics.—For Epicurus the criterion of good and evil is sensation; so that "we declare pleasure to be the beginning and end of the blessed life; for we recognize this to be our first and natural good, and from this we start in every choice and avoidance; and this we make our goal, using feeling as the canon by which we judge every good." But while every pleasure is in itself good, not all pleasures are to be chosen, since certain pleasures are produced by means which entail annoyances many times greater than the pleasures. Moreover, a right conception of pleasure itself conduces to right living, since "it is not possible to live pleasantly without living wisely and well and righteously." In this light the declaration that "beauty, virtue and the like are to be valued if they produce pleasure; if they do not, we must bid them farewell" (cited by Athenaeus, *Deipnosophistae*, xii, 546f) is not so hedonistic as it sounds. Pleasure as contemplated by Epicurus is not so much active enjoyment as the absence of pain: "Epicurus differs from the Cyrenaics as regards pleasure; for they do not admit static but only active pleasure . . . But Epicurus in his work on *Choice* says: Quiet of mind and absence of pain are static pleasures, but joy and mirth are seen in activity" (Diog. Laërt., x, 136).

Since happiness lies in quiet of mind, Epicurus is concerned to eliminate every disturbance to the soul. He urges withdrawal from public life and above all wishes to remove the two great disturbing ideas in man, namely the fear of the gods and the fear of death. As for the latter, Epicurus holds that there can be no life to come, since the soul (as has been explained above) is of such a nature as to be dissolved immediately on leaving the body into the primordial atoms of which it was compounded. As for the fear of the gods, it can be avoided through adopting the true teachings of philosophical theology.

Theology.—Epicurus does not deny the existence of the gods; rather he holds that "their existence is known to us by immediate apprehension." They are not, however, such as they are usually supposed to be; that is to say, popular religion errs in crediting them with the government of the world and with feelings of anger or of love toward men. Fashioned of finer stuff than we, they dwell afar in the intermundial spaces, neither troubling human affairs nor troubled by them. The more man is at peace to receive the images emanating from the gods, the more easily will he win participation in the serenity of the heavenly beings. Thus there is in fact an Epicurean counterpart to Plato's doctrine of "assimilation to God": the wise man resembles the gods inasmuch as he shares in their state of pleasure; and his chief way of attaining this is his inner vision of the ideal divine figures in their sublimity. Epicurus' attitude toward religion seems to be twofold: he rejects superstitious fear but lends philosophical support to piety. His devout aspiration to the divine sublimity has an aesthetico-hedonistic foundation, but in no way excludes continued attachment to inherited religion. The theology of Epicurus has in fact been called a "philosophy of the Olympian religion."

Epicureanism as a Message of Salvation.—Insofar as he attained to divine living for himself, Epicurus could be regarded as having acquired also the divine faculty of being a saviour to others; and it is with this latter aspect in mind that his followers again and again worshiped him. Lucretius apostrophizes him in the language of religious awe and veneration (*De rerum natura*, iii, 1–30; v, 1–54); and the feelings of devotion of Colotes for his master are aptly described by Epicurus himself (*see* Plutarch, *Adversus Coloten*, 17; *i.e.*, *Moralia*, 1,117^{b}). Epicurus, then, was regarded as the founder of a religion, and his sayings were taken as "medicaments of salvation" (*Diogenis Oenoandensis fragmenta*, col. 5, line 14). His followers formed communities with which the head of the school kept in constant touch; and it is not too much to say that the letters of Epicurus to these communities and disciples recall, in some way and with all allowance for differences, the epistles of St. Paul. (Nonetheless, although similarities may be discerned between Epicurus and the New Testament, it would be rash to infer influence from mere agreement.) We are entitled to follow Metrodorus, the master's favourite disciple, in considering the essence of Epicurus' thought to have been revelation rather than theory: "Let us change over," says Metrodorus to Timarchus, "from lowly living to the truly divinely revealed mysteries of Epicurus" (quoted by Plutarch, *Adversus Coloten*, 17). It is evident, however, that the doctrine of salvation propounded by Epicurus is a rational one, though one which, at times, shows a preference for language reminiscent of elements of Greek mysticism. It offers to the human soul release from its anxieties, and shows how the simple fact of its being can be a ground for real happiness.

The Epicurean school survived until the 4th century A.D. Among his immediate disciples the most eminent were Metrodorus; Hermarchus, who succeeded Epicurus as head of the school; and Colotes of Lampsacus. Later prominent Epicureans were Apollodorus (in the second half of the 2nd century B.C.), called "Lord of the Garden"; Demetrius Lacon (born *c.* 160 B.C.), outstanding in his grammatical and mathematical research (a feature not to be expected of an Epicurean); Zeno of Sidon (born *c.* 150 B.C.), teacher of Cicero; and Zeno's pupil, Philodemus of Gadara, of whose works some fragments have been recovered from Herculanean rolls. The most eminent Roman follower of Epicurus was, of course, Lucretius. In modern times, the 17th century saw the revival of Epicurean physics by Pierre Gassendi.

See also references under "Epicurus" in the Index volume.

BIBLIOGRAPHY.—For texts and commentary *see* H. Usener (ed.), *Epicurea* (1887), a standard work; H. Diels, "Ein epikureisches Fragment über Götterverehrung," *S. B. preuss. Akad. Wiss.* (1916); E. Bignone (ed.), *Epicuro: Opere, frammenti, testimonianze*, Ital. trans. with notes (1920); P. von der Mühll (ed.), *Epicuri epistulae tres et ratae sententiae a Laertio Diogene servatae* (1922); R. D. Hicks (ed.), Diogenes Laërtius, *Lives of Eminent Philosophers*, 2 vol., with Eng. trans. (1925); C. Bailey (ed.), *Epicurus, the Extant Remains* (1926); A. Vogliano (ed.) *Epicuri et Epicureorum scripta in Herculanensibus papyris servata* (1928), *I Resti dell' XI libro del Περὶ φύσεως di Epicuro* (1940); C. Jensen (ed.), "Ein neuer Brief Epikurs," *Abh. Ges. Wiss. Göttingen: phil.-hist. Kl.*, ser. 3, no. 5 (1933); W. Schmid (ed.), *Ethica Epicurea* (1939); C. Diano (ed.), *Epicuri ethica* (Florence, 1946) and *Lettere di Epicuro e dei suoi* (1946). For fragments of *Peri Physeos* not contained in Vogliano, *op. cit.*, see the list given by W. Schmid, "Geschichte der herculanensischen Studien," *La Parola del passato*, vol. x, p. 498 (1955).

See further W. Crönert, *Kolotes und Menedemos* (1906); C. Bailey, *The Greek Atomists and Epicurus* (1928); R. Philippson, "Neues über Epikur und seine Schule," *Nachr. Ges. Wiss. Göttingen* (1929–30); E. Bignone, *L'Aristotele perduto e la formazione filosofica di Epicuro*, 2 vol. (1936); W. Schmid, *Epikurs Kritik der platonischen Elementenlehre* (1936); O. Gigon, *Epikur: Von der Überwindung der Furcht*

(1949); N. W. DeWitt, *Epicurus and his Philosophy* (1954), *St. Paul and Epicurus* (1954); and A. J. Festugière, *Épicure et ses dieux* (1940; Eng. trans. by C. W. Chilton, *Epicurus and His Gods*, 1956).
(A. W. Ma.; W. Sd.)

EPICYCLE, in ancient astronomy, a small circle whose centre moves in a larger circle. It was used by Ptolemy to represent geometrically the apparent motions of the planets (including the sun and moon). In the Ptolemaic system, each planet was supposed to move around the circumference of a circle called the epicycle, the centre of which in turn moved around the earth on the circumference of a larger circle called the deferent. *See* Astronomy: *History of Astronomy;* Ptolemaic System.

EPIDAURUS, the name of two ancient cities of Greece.

Epidaurus, now Palaia Epidavros in the modern *nomos* (prefecture) of Argolis, a city on the eastern coast of the Argolid (northeastern Peloponnese), on a rocky peninsula with a natural harbour on the northern side and an open bay on the southern. Its narrow but fertile plain is shut in except toward the sea by mountains, of which the most notable are Mts. Arachnaeon (Arakhnaion) and Titthion. The commercial advantages of its position, and the fame of its temple of Asclepius (*q.v.*), the Greek god of healing, made Epidaurus a place of some importance. Portions of the walls can be traced and the locations of several temples are known. Some Mycenaean tombs have been found nearby.

The *hieron* (sacred precinct) of Asclepius, about 8 mi. inland from the town, has been excavated by the Greek archaeological society. In addition to the sacred precinct with its temples and other buildings, the theatre and stadium have been cleared and baths, gymnasiums and a hospital have been found. The sacred road from Epidaurus, flanked by tombs, approaches through a gateway. The chief buildings are grouped together, and include the temples of Asclepius and Artemis, the *tholos,* and the *abaton* (portico) where the patients slept. Architecture and sculpture are of high merit, and inscriptions record cures attributed to the god.

The temple of Asclepius, which contained the gold and ivory statue of the god by Thrasymedes of Paros (early 4th century B.C.), has 6 columns at the ends and 11 at the sides; it was approached by a ramp at the eastern front. An inscription records contracts for building this temple about 380 B.C. The sculptor Timotheus, who later collaborated in the monument to Mausolus (the Mausoleum) at Halicarnassus, undertook to make the *acroteria* (figures on top of the gables) for one end of the temple and also certain reliefs. Two Nereids on horseback found at the western end of the temple, as well as certain other sculptures, may be his work. The great altar lay to the south of the temple, and beyond it a small temple of Artemis. The *tholos,* to the southwest of the temple of Asclepius, must have been one of the most beautiful buildings in Greece; the exquisite carving is only equalled by that of the Erechtheum (*q.v.*) at Athens. It consisted of a circular chamber with a Doric colonnade outside and Corinthian within. The foundations are a series of concentric walls with doors and partitions that make a subterranean labyrinth. The architect, who also designed the theatre, was Polyclitus, probably the younger Argive sculptor of that name (4th century B.C.). In the contracts for its building it is called the *thymele* (altar). Its purpose has been much discussed and many fanciful theories have been propounded. It is perhaps best thought of as the idealized architectural representation of a primitive pit of sacrifice such as may still be seen in the Asclepieion (temple of Asclepius) at Athens. There is no evidence of a well or spring. North of the *tholos* is a long portico on two different levels, the lower of which had two stories. This was probably the building called in the inscriptions the *abaton.* There the invalids used to sleep when consulting the god, and there were found the long inscriptions which record not only the method of consulting the god but the manner of his cures. The patient sleeps in the *abaton,* sees certain visions and comes forth cured. There are some surgical cases, and many examples of cures resembling those at Lourdes or Tenos. It is, however, difficult to make scientific use of the records. The long lists of cases were compiled by the priest from earlier dedications or tradition; there are also some dedications contemporary with the cure. In later times the old faith-healing lost its efficacy and the priest substituted elaborate prescriptions as to diet, baths and regimen which must have made Epidaurus resemble a modern spa. Extensive buildings were provided for the accommodation of invalids, great courtyards surrounded by colonnades and chambers. There was a large gymnasium which was abandoned in Roman times when a small *odeum* (concert hall) was built in the middle. The stadium consists of earth embankments with stone seats only near the centre of each side. The starting lines are well preserved.

The theatre, as Pausanias says, is the most beautiful in Greece, almost every seat being still in place. The complete circle of the orchestra is marked by a sill of white limestone. There are benches with backs not only in the bottom row, but above and below the *diazoma* (horizontal gangway). A speaker in the orchestra can make himself heard throughout the auditorium without raising his voice. The original stage buildings are not preserved much above their foundations. The missing seats were restored and the theatre is used each summer for performances of ancient drama.

On the slopes of Mt. Cynortium, about 15 minutes on foot above the theatre, is the sanctuary of Apollo Maleatas. A massive terrace wall of the 4th century B.C. supports a colonnade, a small temple and other buildings. Pottery finds carry the history of this sanctuary far back into the Bronze Age, and it is clearly the original seat of worship in the area. The cult of Asclepius, Apollo's son, sprang from there, and it was there that suppliants at the Asclepieion regularly made their first sacrifice.

Epidaurus Limera (either "the well-havened" or "the hungry") was a city of the Peloponnese on the east coast of Laconia, founded by the people of Epidaurus in the Argolid. The ruins occupy a rocky hill near the shore about 3 mi. N. of the Byzantine fortress of Monemvasia, the ancient promontory of Minoa. Mycenaean tombs have been found in the neighbourhood.

Bibliography.—*Epidaurus:* Pausanias, *Description of Greece,* ii, 27, and J. G. Frazer's commentary, including addenda (1898). P. Cavvadias, *Fouilles d'Epidaure* (1891), and *To Hieron tou Asklepiou en Epidauro* (1900). E. J. and L. Edelstein, *Asclepius* (1945) contains text, translation and interpretation of the great inscriptions recording cures as well as all other ancient testimonies concerning Asclepius. *Epidaurus Limera:* Pausanias, *Description of Greece,* iii, 23, with J. G. Frazer's commentary; A. J. B. Wace and F. W. Hasluck, *Annual of the British School at Athens,* xiv, pp. 179–182 (1907–08).
(E. Gr.; E. V.)

EPIDEMIC: *see* Epidemiology.

EPIDEMIOLOGY was formerly defined as the science or doctrine of epidemics, or unusually high outbreaks of sickness affecting several persons within a short time. This implies, however, that the phenomena of epidemics are distinct from those of the interepidemic or endemic occurrence of a disease. Such a view is no longer tenable. An epidemic is only a temporary phase in the natural history of a disease. This phase cannot be considered except in relation to its antecedent or subsequent occurrence in the same geographic area and population. Accordingly, usage has broadened the concept of epidemiology to include not only the occurrence of disease in outbreaks but also its existence at any level of prevalence.

Definition and Scope.—For clinical medicine, disease is described as it occurs in individuals—onset, symptoms, physical signs, anatomical involvement, course, termination, etc. For epidemiology, disease is described as it occurs in an aggregation of individuals making up a group or population—the number of cases of a specified illness and their distribution in relation to time, place of residence or employment, race, sex, age, occupation, pertinent environmental conditions or to some other meaningful attribute. These are the "mass phenomena." When adequately described they are quite distinctive for each disease and may contribute part of the evidence on which diagnosis is based. This is evident in such terminology as epidemic typhus, scrub typhus, Rocky mountain spotted fever, trench fever, etc.

Although originally limited to disease in human populations, epidemiology has been applied to the study of disease in animal herds and in plant life. Such usage is justified by the derivation of the

word epidemic, which literally translated from the Greek means "upon the population." The population may consist of human beings, animals, plants or what not. Thus, a veterinarian may refer to the epidemiology of cattle plague, a plant pathologist to the epidemiology of stem rust of wheat. However, this article is limited to the original connotation, the study of the occurrence of disease in human populations.

Epidemiology may now be defined as that field of medical science concerned with the description of factors and conditions that are significantly associated with the occurrence of an infectious process, disease or abnormal physiological state in a human community, with elucidation of the manner in which these factors and conditions operate in the causative complex. To the extent that this knowledge is advanced and integrated with that of collateral fields of science, particularly of experimental pathology, it becomes possible for appropriate community agencies to take effective measures of prevention, control or eradication. It supplies the scientific background on which public health strategy is based.

But epidemiology is more than a descriptive science. It may be determinative or experimental. By observing and comparing the frequency of occurrence of a disease in different population groups, places, times and circumstances and by drawing upon collateral knowledge, it is possible to reason inductively about the causation; that is, not only about the nature and sources of the primary causative agent but also about the part played by secondary or contributory factors and conditions. Assembling evidence of this kind into a logical pattern leads progressively to a concept of forces that are favourable or unfavourable to the continued existence of a disease in a population.

While epidemiology was developed largely around investigations of infectious diseases, its scope has been extended to include diseases and abnormal physiological states due to noninfectious causes, such as nutritional deficiencies, cancer, essential hypertension, multiple sclerosis and accidents. On the other hand, it has been suggested that there is an epidemiology of health (New York Academy of Medicine, Health Education Council, *Epidemiology of Health* [1953]). Intriguing though this proposal may be to some, others would question this extension of meaning. Health is identified negatively with the absence of disease and abnormality and positively with the range of variation of normal physiological function, growth and development.

Sources of Data.—There are three sources of quantitative data: (1) official registration of deaths due to all causes; (2) official reports of cases of a selected list of communicable diseases; (3) field investigations of a specific disease conducted on selected groups or representative samples of a population. The first source has the advantage of being available for large population units all over the civilized world and, in many countries, over long periods of time. Specific death or mortality rates can be used as an index of the frequency of occurrence of fatal cases of a disease and their distribution according to sex, race, age, geography and time (long-term trends). Specific morbidity rates (attack rates), where available, are a more sensitive index of shifts in incidence and distribution over short time periods and for smaller population units. With mortality rates, attack rates provide the data for delineating the general epidemiological features of a disease.

Special investigations or field studies vary greatly in scope, methods and objectives. They may be historical (retrospective), observational or experimental. The investigation may be designed to reveal what proportion of persons have had or have a specific infection, disease or abnormal state at a particular time (prevalence ratio). It may be designed to characterize in some detail the kinds of people who acquire a disease and those who escape. It may seek to ascertain the cause of an epidemic. It may be set up as an experiment to test the protection afforded by an immunizing agent under natural conditions of exposure or to test the effectiveness of some other control measures. Whatever its purpose, an appropriate schedule of information is obtained about each case and about activities and environment from assembled records or personal interview, together with the results of such laboratory examinations, skin tests, X-rays, etc., as are deemed desirable and practical.

To ask intelligent questions and to interpret the data require the formulation of adequate hypotheses. Logically, this leads to a division of epidemiology into two categories, that of diseases due to living agents (infectious diseases) and that of diseases and states due to nonliving agents or deficiencies (noninfectious).

Basic Concepts of Infectious Disease.—The basic concept upon which our reasoning with regard to the infectious diseases rests was comprehensively formulated by Theobald Smith in 1934. Smith wrote, "Infectious disease is a manifestation of parasitism." (*See* PARASITOLOGY.) Simple though this concept seems now, it afforded the framework into which scattered observations could be fitted into a logical pattern. The dynamics of infectious disease in human communities are interpreted as expressions of the eternal struggle of living things for food by predation and parasitism, for shelter and for opportunity to propagate their kind. As a result of centuries of host wanderings, mutation and selective adaptation, certain worms, protozoa, fungi, bacteria, rickettsiae and viruses have been successful in maintaining their propagation in the biologic orbit of man and are responsible for some of his disabilities, plagues and pestilences.

Natural Host Range.—The first requisite of a rational explanation of the behaviour of infectious diseases is to understand to what extent man as a host bears responsibility for the continuous propagation of disease agents or shares this responsibility with other species, arthropod (insects, ticks, mites, etc.) or animal.

This concept affords the useful division of infections into two groups. In the first group are those caused by microparasites that are dependent upon the human host alone for survival. They are propagated from person to person by some form of direct or indirect contact. Hence, they are contagious infections. The second group includes those diseases caused by specific microparasites that are not directly transmissible by person-to-person contact or are transmissible only occasionally under unusual circumstances. Either they are arthropod-borne or spread by contact with animal or other extrahuman sources or some combination of these relationships. Propagation of these human noncontagious infections involves more than one host species and transmission chains of varying complexities, to which reference will be made later.

Human Host-Parasite Interaction.—A second requisite of a rational explanation of the behaviour of infectious diseases is knowledge of the characteristics of the reaction between the human host and a specific parasitic species. In clinical medicine, interest is centred upon a patient during the period when he is more or less incapacitated by the invasion of the pathogenic microorganism; *i.e.*, from onset of symptoms to clinical recovery or death. The epidemiology of a disease is determined by all phases of the host-parasite relationship—from the time and circumstances of exposure until the microparasite is suppressed or eliminated from the host's body. This includes the manner in which the microparasite gets into the body (portal of entry), the time lag before onset of symptoms (incubation period), the pathways of invasion and tissue localization (pathogenesis), the clinical manifestations and course of the illness (if any), how and when the progeny of the microparasite make their exit from the body (infectious period), how resistant are the microparasites to the conditions of external environment (extrahuman survival), the routes they may take to a new human host (modes of transmission).

In an individual, the outcome of an infectious process is determined by the balance between the devices of aggression of the infecting strain of microparasites (infectivity, virulence, pathogenicity) and the mechanism of defense of the human host (genetic resistance, susceptibility, immunity) (*see* IMMUNITY AND IMMUNIZATION). The reaction to an infection varies in severity and duration. If it is of sufficient intensity to permit clinical diagnosis it is classified as a "case" of a disease.

If the clinical manifestations are so indefinite or so slight as to pass unnoticed, it is classified as a subclinical or inapparent infection, identified by such procedures as cultural recovery of the infecting microorganism as, for example, by nose or throat cultures, for diphtheria bacilli; change in the response of the skin to a specific antigenic material, as in the tuberculin test; or change in a serologic reaction from negative (low titre) to positive (high

titre), as in the serum-neutralization test with the virus of influenza.

The term carrier (*q.v.*) includes two categories, those who continue to harbour and excrete pathogenic microparasites: (1) after recovery from a recognized attack; and (2) during or after recovery from a subclinical infection.

Thus, the source or reservoir of infectious agents causing contagious diseases consists of the number of cases, subclinical infections and convalescent carriers that exist in the population at any time. Spread through human populations is a highly visible phenomenon when the ratio between clinical and subclinical infections is high (about 19 to 1), as it is with measles, and largely invisible when it is low (about 1 to 100), as in poliomyelitis. The proportion of cases that terminate in death is an index of virulence. With each disease the case-fatality rate has a characteristic age distribution.

Group or Herd Susceptibility.—The human community is made up of individuals who vary not only in their genetic capacity to react but nearly always in previous experience with a particular species of microparasite and its close relatives. Some individuals have acquired immunity, some a partial immunity and some none. The proportion of a population that at any one time has little or no immunity determines the theoretic level of community, group or herd susceptibility for a particular contagious infection. If such an infection is commonly prevalent in a community, the proportion is constantly changing as susceptibles are infected, develop immunity and recover. If the immunity conferred by either clinical or subclinical infection is durable, susceptibility decreases with advancing age as more and more of the population have had experience with the agent; the age distribution of cases, consequently, is that of a children's disease. If the immunity conferred is temporary, as in some acute respiratory infections such as influenza, the same individual may be reinfected after a short interval; consequently, the disease may attack adults and old people as well as children. Thus, for each of the common contagious diseases there is a characteristic age distribution of cases from which inferences may be drawn in regard to past experience and durability of specific immunity.

Effect of Environment.—The spread of a contagious infection through a community is affected not only (1) by the nature of the interaction between the microparasite and the human host and (2) by the proportion of susceptibles but also (3) by the frequency of contact between individuals under the circumstances required for transmission—the effective contact rate. This is determined by such conditions as the density of population, house or room crowding, household cleanliness, personal hygiene, population movement, transportation facilities and places of congregation, such as schools, churches, theatres, barracks, etc. In communal living the potentialities of common water, milk and food supplies as vehicles of dissemination are well known. The downward trend of many of the contagious diseases, especially those transmitted by fecal dissemination, during the first half of the 20th century, was largely the result of improved sanitation.

That climatic factors—temperature, humidity, rainfall, sunlight—affect the incidence of contagious diseases is confirmed by the fact that each has a characteristic seasonal distribution. Whether such factors act directly by influencing the extrahuman survival of the microparasite or indirectly by affecting the resistance of the host population is not known.

Epidemics of Contagious Disease.—The dynamics of the complex interrelationship between a specific microparasitic population and a human host population, as modified by constantly changing environmental conditions, is indicated by fluctuation in the incidence rates with the passage of time. When the balance is stable it is manifested by an endemic prevalence, as in tuberculosis. When subject to violent disturbance it is manifested by epidemics, as with influenza. If the balance is in favour of the microparasite the trend is upward; if in favour of the human population the disease shows a downward trend and tends to disappear from the community, at least temporarily, until it is reintroduced.

Each of the common contagious diseases has an interannual variation, years of high incidence (epidemic years) being interspersed with a variable number of years of low incidence, some exhibiting a more or less regular and predictable periodicity. For example, epidemics of influenza A can be expected every two or three years, according to the Commission on Acute Respiratory Diseases. Each has a seasonal distribution that follows a more or less regular pattern, reaching maximum incidence about the same time each year, acute respiratory infections during colder months, alimentary, fecal-borne infections during the warmer months.

Properly speaking, an epidemic is an unusually high incidence of a disease defined in time, place and persons and compared with previous experience. Epidemics of contagious disease are of two types. A propagated or contact epidemic is one that results from increased frequency of person-to-person or chain transmission. A common vehicle epidemic is one that results from dissemination of the causative agent to a group of people through a common medium, such as water, milk or food.

If the onsets of all or nearly all of the cases fall within an interval no greater than that of the known variation in the incubation period for the disease, it can be assumed that they arose from simultaneous exposure to a common vehicle of dissemination. The problem is then resolved into discovering what common occasion or by what common medium the persons selected for attack were exposed. When the onsets of cases are scattered over an interval of time greater than that of the maximum incubation period, it can be assumed either that the common source continued to act over a period of time or that the cases are in part or wholly caused by transmission by person-to-person contact. To explain the increased incidence in a propagated epidemic is more difficult. The principal factors and conditions that affect the host-parasite balance can be enumerated, but to assess their relative importance in a particular situation is not easy.

Experimental Epidemiology.—An experimental approach to discovery of laws and general principles governing the behaviour of infectious diseases has been made through observations upon epidemics in animal colonies. Notable among the many contributions were those made by L. T. Webster and his associates (1946) and by W. W. C. Topley and his associates (1942). The general procedure was to assemble uninfected animals in unit cages, the arrangement of which could be altered to simulate a community of any desired size. An epidemic was started by introducing into the uninfected mouse colony a number of mice infected with microparasites to which they were naturally susceptible and which were spread from mouse to mouse by contact. The course of the subsequent epidemic was indicated by the occurrence of specific deaths proved by necropsy, cultures or both.

Effort was made to hold all of the important factors constant except the one under examination and to note the effect this variable had upon the course of an artificially produced epidemic. From these observations, investigators were inclined to believe that changes in the virulence of the infecting microparasites were of little or no importance in causing the rise and fall of epidemic waves. They emphasized particularly the accelerating effect upon incidence of an influx of susceptibles into an infected community and aggregation of individuals into large groups (crowding) and, on the other hand, the retarding effect of the accumulation of immune survivors and dispersion of the mice into small cage units. These experiments added support to some of the concepts derived from experiences with propagated epidemics in human populations under natural conditions.

Noncontagious Infections.—Some of the most devastating diseases of mankind are due to microparasites that (1) can make exit from the human body only through the intermediation of the bite of an arthropod vector or (2) cannot make an effective exit, human infection being a terminal event in transmission. In the first situation, man may share responsibility for the continued propagation of the microparasite with an arthropod vector in the transmission, as in louse-man-louse propagation of epidemic typhus fever or mosquito-man-mosquito in malaria. In the second situation, survival of the parasitic species is entirely independent of its relationship to man. Thus, *Pasteurella pestis*, the cause of bubonic plague, is continually propagated in rodents through the intermediation of ectoparasites. The principal transmission chain is rat-

flea-rat. The infection of man by an infected flea is an accident that is of no biological importance in the continuous passage of *P. pestis*.

To understand and effectively interfere with transmission of a disease agent to man requires knowledge of the underlying infection chains and the ecology associated with the arthropod, animal or bird hosts. How complicated these infection chains may be is suggested by the natural host range of *Bacterium tularense*, the cause of tularemia in man. By 1954 *B. tularense* had been isolated from naturally infected animals as follows: wild rabbits and hares and their ticks, widespread in the United States; California ground squirrels, often; ground squirrels of Utah, wild rats of Los Angeles, wild mice of California, quail, ruffed grouse, sharp-tailed grouse and gray foxes of Minnesota, once each; sage hens and their ticks in Montana, once; sheep of Idaho, three times; wild rabbits of Japan, Norway and Canada, the water rat of Europe (U.S.S.R.), wood ticks of Montana and Minnesota, often; the wood ticks of California, once.

The concise term "zoonoses" gained wide acceptance as denoting about 80 animal diseases transmissible from animals to human beings. One epidemiological characteristic which these diseases have in common is that none is world-wide in distribution. They are limited in their geography to localities or focuses where specialized ecology required in maintaining transmission can be realized. The second characteristic is that the immunity status of the human population is rarely an important limiting condition. Usually the seasonal variation and group selection are determined principally by exposure to the bite of the arthropod vector or contact with infected animals or their products or both. Prevention and elimination of the human disease depend in large part on control of these infections in the animal or vector hosts.

Noninfectious Diseases and Abnormal Physical States.—The epidemiological approach can be applied to the problems presented by diseases due to dietary deficiencies, to industrial hazards, to disturbances of growth and metabolism, etc. Observation is limited to human populations and the factors or deficiencies that may be injurious. A primary requisite is that the morbid condition that is being studied be clearly differentiated by clinical examination and laboratory tests, and the criteria upon which diagnosis is based be satisfactory. The objective is to ascertain what proportion of people have this disease or state and how they differ significantly from those who do not have it, and whether these differences are causally related.

Field studies were particularly productive in advancing knowledge of nutrition. Notable examples were the discovery of the relation of dietary deficiency (niacin) to causation of pellagra; of drinking water low in iodine content to endemic goitre; of drinking water low in fluoride content to dental caries; and of drinking water high in fluoride content to mottled enamel.

See BACTERIAL AND INFECTIOUS DISEASES; IMMUNITY AND IMMUNIZATION; PARASITOLOGY; PREVENTIVE MEDICINE; PUBLIC HEALTH; *see* also references under "Epidemiology" in the Index.

BIBLIOGRAPHY.—New York Academy of Medicine, Health Education Council, *Epidemiology of Health* (1953); W. H. Frost, *Papers; a Contribution to Epidemiological Method*, ed. by K. F. Maxcy (1941); J. E. Gordon and H. LeRiche, "The Epidemiologic Method Applied to Nutrition," *Am. J. Med. Sc.*, 219:321–345 (March 1950); A. Hirsch, *Handbook of Geographic and Historical Pathology*, vol. i, ii, iii (1883–85); H. R. Leavell and E. G. Clark (eds.), *Textbook of Preventive Medicine* (1953); M. J. Rosenau, *Preventive Medicine and Hygiene*, 7th ed., ed. by K. F. Maxcy (1951); K. F. Meyer, "The Animal Kingdom: a Reservoir of Human Disease," *Ann. Int. Med.*, 29:326 (1948); W. W. C. Topley, "The Biology of Epidemics," *Proc. Roy. Soc.*, London, s.B. 130:337 (1942); L. T. Webster, "Experimental Epidemiology," *Medicine*, 25:77 (1946). (K. F. M.)

EPIDOSITE, in petrology, a typical member of a family of metamorphic rocks composed mainly of epidote and quartz; they are pale or greenish yellow, hard and somewhat brittle, and are derived from several kinds of rock. Some have been epidotic grits and sandstones; others are limestones which have undergone contact-alteration; but the majority are probably allied to epidiorite and amphibolite and are local modifications of rocks which were primarily basic intrusions or lavas. The sedimentary epidosites occur with mica-schists, sheared grits (sandstones) and granulitic gneisses; they often show, on minute examination, the remains of clastic, that is, pre-existing, structures. The epidosites derived from limestones may contain a great variety of minerals such as calcite, augite, garnet, scapolite, etc., but their source may usually be inferred from their close association with calc-silicate rocks in the field. The third group of epidosites may form bands, veins or irregular streaks and nodules in masses of epidiorite and hornblende-schist. In microscopic section they are often merely a granular mosaic of quartz and epidote with some iron oxides and chlorite, but in other cases they retain much of the structure of the original rock though there has been a complete replacement of the former minerals by new ones. Epidosites when streaked and variegated have been cut and polished as ornamental stones.

See also METAMORPHISM.

EPIDOTE is a group of complex calcium and aluminum orthosilicate minerals and also is the most commonly occurring mineral of the group, which includes zoisite (*q.v.*), clinozoisite, piedmontite and allanite. Well-developed crystals are of frequent occurrence and have been found at many localities, the best being from Knappenwald in the Untersulzbachthal in Salzburg, and Prince of Wales Island, Alaska. Well-known occurrences of epidote include those in Chaffee county, Colo., at Haddam and Willimantic, Conn., and at Franconia, N.H. Perfectly transparent dark-green crystals obtained from the Knappenwald and Brazil have occasionally been used as gem stones.

Epidote crystallizes in the monoclinic system and crystals are usually elongated along the b axis (perpendicular to the plane of symmetry). A perfect basal cleavage is present. Many of the characteristics of this mineral, as for instance the colour (yellow to dark green), the hardness (6 to 7) and the specific gravity (3.3 to 3.5), vary with the amount of iron present in the form of ferric oxide (Fe_2O_3), which ranges from 5% to 17%.

Epidote is always found as a secondary mineral. It occurs most abundantly in low-grade metamorphic rocks, such as marble and schist (*qq.v.*). It is fairly common in igneous rocks as an alteration product of ferro-magnesian minerals and calcic feldspars (*see* FELDSPAR).

The epidote minerals have the general formula $R_2^{2+}\ R_3^{3+}\ Si_3O_{12}(OH)$. Substitution of the cations R_2^{2+} and R_3^{3+} results in the formation of the different mineralogical species:

R^{2+}	R^{3+}	Mineral name
Ca	Al	zoisite, clinozoisite
Ca	Fe,Al	epidote
Ca	Al,Mn	piedmontite
Ca,Ce,Fe	Al,Fe	allanite

The minerals epidote and zoisite (*q.v.*) have been synthesized under hydrothermal conditions. A fine-grained aggregate of epidote or zoisite formed from alteration of plagioclase is known as saussurite. The term epidosite (*q.v.*) has been given to highly altered igneous rocks containing large amounts of epidote and quartz. (E. G. E.)

EPIGRAM, originally an inscription suitable for carving on a monument, but since the time of the *Greek Anthology* applied to any brief and pithy verse, particularly if astringent and purporting to point a moral. By extension the term is also applied to any striking sentence in a novel, play, poem or conversation which appears to express a succinct truth, usually in the form of a generalization.

The Greek epigram is treated in the article ANTHOLOGY. Two lines of Catullus initiate the Latin:

> Odi et amo: quare id faciam, fortasse requiris.
> Nescio, sed fieri sentio et excrucior.

("I hate and I love: perhaps you ask why. I do not know, but I feel it, and I am in torment.")

The form was given final shape by the Spanish-born Latin epigrammatist Martial (*q.v.*), whose pungent and often indecent verses served as models for French and English epigrammatists of the 17th and 18th centuries. He provided, among many pieces frequently translated or adapted, the original of Thomas Brown's "I do not love thee, Dr. Fell":

Non amo te, Sabidi, nec possum dicere quare;
Hoc tantum possum dicere, non amo te.

The middle ages can show few epigrams, though the Spanish rabbi Sem Tob's sententious *Proverbios* (mid-14th century) are certainly epigrammatic. The form was revived by Renaissance scholars and the French poets Clément Marot and Mellin de Saint-Gelais, and epigrams were written both in Latin and the vernacular. In England the form took shape somewhat later in the hands of Ben Jonson and his school, among whom Robert Herrick wrote with particular charm and variety. "The Amber Bead" and the brief "No difference i' th' dark" may stand as samples of his two contrasting styles:

I saw a Flie within a Beade
Of Amber cleanly buried:
The Urne was little, but the room
More rich than Cleopatra's Tombe.

This looks forward to the lapidary Landor, whereas the second example offers a bucolic truth with something of Martial's directness:

Night makes no difference 'twixt the Priest and Clark;
Jone as my Lady is as good i' th' dark.

As the 17th century progressed the epigram grew more astringent and closer to Martial both in England and France. French practitioners, among whom were François de Malherbe, Boileau, J. B. Rousseau and Voltaire, used the iambic quatrain, varied occasionally by the couplet, and generally set a sting in the tail. Boileau's lines to his enemy Charles Perrault exemplify the merciless wit of the time:

Ton oncle, dis-tu, l'assassin,
M'a guéri d'une maladie;
La preuve qu'il ne fut jamais mon médecin
C'est que je suis encore en vie.

("That murderer your uncle, you say, has cured me of my illness. The proof that he was never my doctor is that I am still alive.")

English epigrams of the early 17th century were, on the whole, clumsy and scabrous. But in the hands of Abraham Cowley, Edmund Waller and Dryden they began to take on a French point and delicacy. As in Boileau's *Satires*, the epigram began to invade the longer poem. Dryden's characterization of "the true old enthusiastic breed" in *Absalom and Achitophel* could stand detached as an epigrammatic distich:

Born to be saved, even in their own despite;
Because they could not help believing right.

In the case of Alexander Pope, the epigrammatic style is even more succinct. Not only are the satires studded with epigrams, but others written in the conventional quatrain are among the best in English. A memorable example is that "On Certain Ladies":

When other fair ones to the Shades go down
Still Chloe, Flavia, Delia stay in Town:
Those ghosts of beauty wandering here reside,
And haunt the places where their honour died.

Pope, Swift, Matthew Prior and many another gentleman at the court of Anne and George I contributed to the bright summer of the English epigram. Prior, in particular, whose longer works are now little read, equalled and surpassed his French exemplars as an epigrammatist, bringing like Herrick a certain homely English coarseness to a medium which was growing over-refined. His advice to lovers in the English climate remains eternally apposite:

Poor Hall caught his death standing under a spout
Expecting till midnight when Nan would come out.
But fatal his patience, as cruel the Dame,
And curst was the weather that quenched the man's flame.
"Whoe'er thou art, that read'st these moral lines,
Make love at home, and go to bed betimes."

The 18th-century epigram proliferated in a number of collections and miscellanies up to the time of Burns and Byron, the only romantic poets to use the form successfully, and Thomas Hood, who invented a punning variety which looks forward to the clerihew (*q.v.*) and is at its best in "On the Death of the Giraffe":

They say, God wot!
She died upon the spot:
But then in spots she was so rich,—
I wonder which?

The chief epigrammatist of the early 19th century was Walter Savage Landor, who substituted sentiment for satire and looked to the *Greek Anthology* for models. His "Love and Time" introduces his characteristic elegiac note:

Love flies with bow unstrung when Time appears,
And trembles at the assault of heavy years;
A few bright feathers bear him on his flight
Quite beyond call, but not forgotten quite.

Victorian poets in general were not succinct enough to be epigrammatic. On the other hand, George Meredith in his novels, and Wilde and Shaw in their plays, placed frequent epigrams in the mouths of their characters. The style of the *Yellow Book* generation as a whole strives for the epigrammatic.

Among more modern writers Hilaire Belloc and his group are remarkable for their attempt to revive the satirical convention of the 18th century. But their epigrams have an air of pastiche; their shafts tickle but do not draw blood.

While the Martial tradition has been dominant in England and France, the *Sinngedicht* or sententious epigram engaged German taste in the 17th century—the time of Friedrich Logau and the mystical Angelus Silesius (*see* SCHEFFLER, JOHANNES). With G. E. Lessing and the Enlightenment there was a brief incursion of the Gallic style, but with Goethe's *Zahme Xenien* (1820) sententiousness resumed its sway. Other countries have shown less interest in the epigram, which appears to have yielded in the 20th century to the off-beat utterance of such individual masters of absurdity as Ogden Nash.

BIBLIOGRAPHY.—J. W. Mackail, *Select Epigrams From the Greek Anthology*, 2 vol. (1938); G. R. Hamilton (ed.), *The Soul of Wit* (1924) and *Wit's Looking Glass* (1934), which contain English and French epigrams respectively, the latter with translations.

(Jo. M. C.)

EPIGRAPHY, a term used to denote (1) the study of inscriptions collectively; (2) the science connected with the classification and explanation of inscriptions; and (3) in a more contracted sense, the paleography in inscriptions. Epigraphy in relation to coins is discussed in the article NUMISMATICS.

See INSCRIPTIONS; PALEOGRAPHY.

EPILEPSY, a chronic clinical disorder or syndrome characterized by recurring paroxysmal attacks or fits in which consciousness is lost or impaired in varying degrees with or without a succession of tonic or clonic muscular spasms. The affliction may appear at any age but has its onset most frequently in early childhood. While the tendency to recurrence of seizures usually continues throughout life, a small percentage of patients experience spontaneous recovery with little or no evidence of mental or emotional deterioration. In areas where its incidence has been determined, approximately 0.3% to 0.5% of the population was found to suffer from epilepsy.

Because of the bizarre and dramatic character of its symptoms, epilepsy is described in the most ancient medical records. It apparently puzzled and plagued mankind in all parts of the world from the dawn of human existence. Its history prior to the middle of the 19th century was a veritable epitome of man's pathetically futile attempts to explain and ameliorate his more serious bodily ailments without the benefit of knowledge concerning natural phenomena. At the time of Hippocrates (460 to 370 B.C.), it was commonly referred to as "the falling sickness" or "the sacred disease," presumably because of its grotesque manifestations and its intractability to treatment. Hippocrates' efforts to change the popular attitude from one that regarded the affliction as evidence of demoniacal possession to one that attributed the symptoms to natural causes was the first notable step toward solution of the problem. The first truly significant advance in its treatment, however, was not achieved until more than 20 centuries later (1857) when the English physician Thomas Laycock introduced bromide therapy.

Types of Seizure.—Individual seizures or fits vary in severity from almost imperceptible momentary lapses of consciousness or mental clouding to violent generalized convulsions lasting for several minutes. The involuntary muscle spasms may be sufficiently violent at times to fracture a vertebra or other bone. The character of the seizures varies according to the part of the brain involved and the nature of the underlying disturbance in brain cell

function. Attacks tend to be predominantly motor, sensory or psychic in character, although various combinations of these different types may occur.

The most common and most terrifying motor seizures are the generalized convulsions (*grand mal* epilepsy). In a typical attack, the afflicted person suddenly loses consciousness with or without a brief preliminary warning (*aura epileptica*) of sensory or motor character. Simultaneously with the loss of consciousness there may be a sharp, loud cry when the muscles of the larynx, like those of the trunk, head and extremities, suddenly go into a state of tonic or stiff contraction. The victim, if standing at the time, may fall to the ground forcibly; as the tongue is protruded between the gnashing teeth, it may be bitten. The limbs stiffen and the head is turned. The pupils dilate and the eyeballs roll upward or to one side. The face first becomes pale but, when breathing is suspended by closing of the glottis and spastic fixation of the respiratory muscles, it becomes livid or purplish in colour. This phase of the seizure ends more or less abruptly after 20 to 30 seconds and the second or clonic (*i.e.*, tumultuous) phase immediately supervenes.

The clonic phase is characterized by violent though rhythmic jerking spasms that involve the entire muscular system and usually last from 30 seconds to more than 100 seconds. During this period the breathing becomes deep and stertorous and foamy saliva (often bloodstained) issues from the mouth. In unusually severe attacks, control of the rectum and bladder may be lost, resulting in fecal and urinary incontinence. Following this phase, the patient regains consciousness for a short time but because of sheer exhaustion is prone to lapse into a deep sleep that may last for an hour or more. Such *grand mal* seizures vary greatly in frequency, duration and severity. In exceptional cases they occur with such frequency that one attack is barely ended when the next begins. Such an episode is designated as *status epilepticus*. Death may occur in the latter condition as a result of exhaustion, asphyxia or hemorrhage into the brain, unless suitable treatment is administered promptly.

A type of focal or localized motor seizure without impairment of consciousness occurs less frequently. This form of seizure is unilateral or it involves the muscles of one extremity or one side of the face only (Jacksonian or partial epilepsy). A still less common type of motor seizure involves some visceral motor mechanism, such as that of mastication (masticatory seizure) or swallowing (deglutitory seizure). A rare form of involuntary motor discharge is that in which an individual group of muscles abruptly goes into a state of violent jerking. This is usually designated as myoclonic epilepsy.

Sensory seizures are recurrent episodic attacks of disturbed sensation, usually focal in character. They may take the form of visual, olfactory, auditory, gustatory, tactual or painful experience or appear as a vague, diffuse feeling of discomfort or shakiness in an area such as the upper part of the abdomen. They occur most frequently as premonitory symptoms (*aura epileptica*) in association with the onset of *grand mal* seizures.

Petit mal seizures, which occur more frequently in childhood than at any other age, consist of a transient loss of consciousness with very minor or no motor manifestations other than staring, upward rolling of the eyes, rhythmic nodding of the head, moving of the lips or very slight quivering of the trunk muscles. Such episodes are frequently referred to by parents, teachers or classmates of the afflicted child as "staring attacks," "little absences," "fainting turns," "dizzy spells" or "little blackouts." They are distinguishable from true fainting or syncopal attacks by the abruptness of their onset, brief duration and occurrence without association with some painful or distressing experience. Activity is temporarily interrupted during the period of unconsciousness but is resumed immediately after termination of the attack. The patient usually drops objects held in his hand or mouth at the onset of the spell, but he seldom falls unless the spell is unduly prolonged. He may or may not be aware of having had an attack when it is over. Such seizures may occur infrequently during the daytime or at night only and so may go unrecognized for years, or they may occur with a frequency numbering several hundred in a single day. As the patient approaches adulthood, the seizures tend to change in character to resemble more closely mild *grand mal* attacks or may cease entirely.

In rare instances, a type of seizure occurs in which the victim, while standing, suddenly collapses or falls to the floor in a heap without any premonitory warning (akinetic seizure). This is thought to be caused by momentary widespread loss of muscular tonus, which may or may not be accompanied by a lapse of consciousness.

Another more or less well-defined type of seizure, the psychomotor attack or epileptic equivalent, may occur in a patient as the sole manifestation of epilepsy or, more commonly, in association with other types of seizures. This form of attack is characterized by varying degrees of impairment of consciousness in which the patient without provocation may wander from home or school, become destructive or negativistic, go into temper tantrums unexpectedly or become loquacious, apathetic or sullen. The victim may suddenly break into an aimless run up and down stairs or in a straight line and as suddenly stop with the return of consciousness. He may have no recollection of what transpired during his period of amnesia. The duration of such attacks may be many minutes. Misconduct of varying degrees of seriousness is at times attributable to this comparatively uncommon type of epileptic reaction.

Causes and Diagnosis.—While much remains to be learned regarding the basic causes of epilepsy, a large number of factors are known to contribute to the provocation of seizures in susceptible persons. Certain abnormal physiological conditions, such as an alkaline state of the blood and other body fluids, excessive water-drinking without an equivalent increase in the intake of common salt, decreased oxygen supply to the brain and various forms of emotional disturbance (*e.g.*, conflict, anxiety and boredom) have this deleterious effect. On the other hand, a state of mild acidosis and dehydration of the body, as well as good emotional status and mental preoccupation, inhibit the seizure tendency.

Injuries of the brain before, during or at any time after birth by mechanical violence, certain toxic agents, infection, hemorrhage, asphyxia or impairment of blood supply may alter the cells of the brain in such a manner as to make them function abnormally thereafter. The terms "symptomatic," "organic" and "secondary" are used interchangeably to designate the type of epilepsy or convulsive state in which a deleterious condition such as brain trauma, exposure to a toxic agent, intracranial infection or hemorrhage or prolonged oxygen deprivation is known to have preceded the onset of recurring seizures. When there is no clinical history of brain injury of any kind and no detectable sign of organic disease of the central nervous system in a patient who suffers from recurrent seizures or fits of the types already described, the terms "idiopathic," "essential," "cryptogenic" or "genuine" epilepsy are used interchangeably in recording the diagnosis. It has long been claimed by some epileptologists that heredity plays a prominent role in the causation of idiopathic epilepsy but equally eminent authorities deny this.

By means of the electroencephalograph, a vacuum-tube amplifying device with oscillographic writing pens for recording changes in the electrical potentials of the brain, Hans Berger in 1931 demonstrated that epileptic seizures are the outward manifestations of a profound disturbance in the electrical activity within the central nervous system. Subsequent studies revealed a close correlation between the character of the abnormal electrical reaction and the type of seizure observed. Whereas the predominating electric wave pattern of normal persons shows a frequency of 8 to 12 per second and potentials varying between 20 and 60 μv, the tracing (electroencephalogram) recorded immediately before and during a *grand mal* seizure is characterized by spikelike waves of high amplitude (100 to 200 μv) occurring at the rate of 15 to 40 per second. *Petit mal* seizure patterns show a sharp spike followed immediately by a dome-shaped, slow wave of high amplitude (100 to 250 μv). These complexes occur at the rate of approximately three per second.

The typical psychomotor seizure pattern is one showing exceptionally broad flat-topped waves of high amplitude with superim-

posed shallow saw-toothed waves. These broad serrated waves are frequently accompanied by large, irregularly occurring slow waves with smooth contours. Even during the free intervals between seizures of all types, abnormal tracings of short duration are found in 85% of epileptic persons. In addition to furnishing information about the physiological disturbance responsible for epileptic fits or seizures, therefore, the electroencephalogram became a highly valued aid in diagnosis. (*See* ELECTROENCEPHALOGRAPHY.)

Various methods are in use for determining the existence in the brain of gross lesions, such as atrophy, old scars, tumours or vascular anomalies, which might contribute to the causation of convulsive attacks. The most helpful adjuncts to the neurological examination in detecting and localizing such abnormalities are electroencephalography and pneumoencephalography. The latter consists of X-ray visualization of the brain and the ventricular and subarachnoid spaces after the cerebrospinal fluid has been replaced temporarily by air, oxygen or other suitable gas. Information so derived assists the physician in estimating the prognosis for the afflicted person. Certain types of lesion detected by these means can be treated surgically.

The greater the extent of organic damage present, the greater is the tendency to mental deterioration with the passage of time. The epileptic patient who displays no evidence of organic brain disease shows little tendency to deterioration if adequately treated.

Treatment.—The guiding principle in treatment of epilepsy is to restore normal brain function without causing deleterious side effects. In many instances the abnormal electrical activity of the brain, as well as the outwardly observed seizures, can be controlled by fasting or by a special ketogenic diet (one with high fat and low carbohydrate content) and moderate restriction of water intake. Good mental hygiene is a highly important factor in the treatment of all cases of epilepsy. The most simple and spectacular therapy, however, is that with drugs—phenobarbital and diphenylhydantoin sodium—for prevention of motor seizures (*grand mal*), which are associated with excessively rapid electrical brain waves, and synthetic oxazolidine compounds and amphetamine sulfate for control of the *petit mal* type, which is associated with an excessively slow electrical rhythm. (*See* NEUROPHARMACOLOGY AND PSYCHOPHARMACOLOGY.) When under suitable treatment persons afflicted with epilepsy should be permitted or encouraged to live essentially normal lives, as far as this is possible. School administrators should provide educational opportunities for children, and employers should be encouraged to find positions for afflicted adults.

See also references under "Epilepsy" in the Index volume.

BIBLIOGRAPHY.—Hans Berger, "Ueber das Electrenkephalogram des Menschen," *Arch. Psychiat. Nervenkr.*, vol. xciv (1931); Wilder Penfield and Theodore Carl Erickson, *Epilepsy and Cerebral Localization* (1941); William Gordon Lennox, *Science and Seizures*, 2nd ed. (1946); Frederic Andrews Gibbs and Erna Gibbs, *Atlas of Electroencephalography* (1952); S. Livingston, *The Diagnosis and Treatment of Convulsive Disorders in Children* (1954); Irvine McQuarrie, "In Pursuit of Convulsive Mechanisms in Clinical Medicine," *Quart. Bull. Northw. Univ. Med. Sch.*, vol. 33, no. 4 (1959). (I. MCQ.)

EPILOBIUM, a genus of plants of the family Onagraceae, popularly called willow herbs. *See* WILLOW HERB.

EPILOGUE. The appendix or supplement to a literary work, and in particular to a drama in verse, is called an epilogue, from *epilogos*, the name given by the Greeks to the peroration of a speech. As in Shakespeare's *A Midsummer Night's Dream*, the epilogue was generally treated as the apology for a play; it was a final appeal made to encourage the good nature of the audiences and to deprecate attack. Ben Jonson made it a feature of his drama and may almost be said to have established in the English theatre the tradition of its regular use. He employed the epilogue either to assert the merit of the play or to deprecate censure of its defects. In *Cynthia's Revels* (1600), for example, the actor went off and immediately came on again, saying:

> Gentles, be't known to you, since I went in
> I am turned rhymer, and do thus begin:—
> The author (jealous how your sense doth take
> His travails) hath enjoined me to make
> Some short and ceremonious epilogue.

He then explained to the audience what an extremely interesting play it had been. Beaumont and Fletcher used the epilogue sparingly, but after their day it came more and more into vogue, and the form was almost invariably that which Ben Jonson had brought into fashion; namely, the short complete piece in heroic couplets. The heyday of the epilogue, however, was the Restoration, and from 1660 to the decline of the drama in the reign of Queen Anne scarcely a play, serious or comic, was produced on the London stage without a prologue and an epilogue. It became the custom for playwrights to ask their friends to write these poems for them, and the publishers would even come to a prominent poet and ask him to supply one for a fee. It gives us an idea of the seriousness with which the epilogue was treated that Dryden originally published his valuable "Defence of the Epilogue; or An Essay on the Dramatic Poetry of the Last Age" (1672) as a defense of the epilogue which he had written for *The Conquest of Granada*. (E. G.)

EPIMENIDES (6th century B.C.?), a Cretan from Cnossus who was reputed to be a prophet and an author of poetical and historical works. None of the few fragments extant under his name is certainly genuine. The line St. Paul quotes from a Cretan prophet has been taken to be his (Tit. i, 12). The details of Epimenides' life are mostly legendary. According to Diogenes Laertius he fell asleep in a cave and awoke after 57 years. Later the Athenians sent for him to purify their city from a plague caused by the murder of Cylon. He is credited with living to the age of 157 or 299, and figures in some lists of the seven wise men. His supposed writings included oracles, a theogony, an epic on the voyage of the "Argo," works on purifications, the mysteries and Cretan history, and two letters to Solon.

BIBLIOGRAPHY.—Fragments, etc., in H. Diels and W. Kranz, *Die Fragmente der Vorsokratiker*, 7th ed. (1954). *See* also H. Démoulin, *Épimenide de Crète* (1901); E. R. Dodds, *The Greeks and the Irrational*, pp. 141–146 (1951).

ÉPINAL, a town of eastern France, capital of the *département* of Vosges, lies on the Moselle 69 km. (43 mi.) S.S.E. of Nancy by road. Pop. (1954) 26,288; (1957 est.) 38,000. The town proper—the Grande Ville—stands on the right bank of the Moselle, which there divides into two arms forming an island where the Petite Ville is. The lesser of these two arms, which is canalized, separates the island from the suburb of Hospice on the left bank. The basilica of St. Maurice preserves a 12th-century tower, and the museum of the *département* has an interesting picture gallery with a magnificent Rembrandt (portrait of an old woman), one of Georges de la Tour's great candlelight scenes, "The Prisoner," and a splendid collection of 18th-century paintings and engravings. The International Museum of Popular Imagery is unique in France. Épinal is the seat of a prefecture, a court of assize, a tribunal of the first instance, a *Tribunal de Commerce* and *Conseil des Prud'hommes* (courts for settling matters of commerce and trade), and a chamber of commerce.

The town is a centre for the weaving, printing and spinning of cotton and also manufactures artificial fibres. There is a school of spinning and weaving and one of industry and commerce. An industry peculiar to the town is the production of cheap images, lithographs and engravings. There is also trade in wine, grain and livestock and in the starch products made in the vicinity.

Épinal grew round a 10th-century monastery founded by Thiery de Hamelant, bishop of Metz. The translation there of the miraculous relics of St. Goëry assured it a rapid development. In the 12th century John of Lorraine surrounded it with walls flanked by towers of red sandstone and built a splendid castle, the ruins of which are now in a park. The founder's successors ruled the town until 1444, when it passed to Charles VII of France who ceded it to the dukes of Lorraine in 1465. In 1671 it was dismantled by the king's order and only parts of the walls and ruins of the castle remained. In 1766 it was incorporated finally into France on the death of the last duke of Lorraine. It was occupied by the Germans during the Franco-German War and again in World War II from June 1940 to Sept. 1944. (AL. A.)

ÉPINAY, LOUISE FLORENCE PÉTRONILLE TARDIEU D'ESCLAVELLES, DAME D' (1726–1783), French writer remembered chiefly for her intimate association

with Rousseau, Diderot and Grimm, was born at Valenciennes on March 11, 1726. On Dec. 23, 1745, she was married to her cousin, Denis Joseph de La Live d'Épinay, a financier, but he proved a bad husband and she was formally separated from him in 1749. After an initial amour with L. C. Dupin de Francueil, she made her château of La Chevrette, near Montmorency, the centre of a circle of distinguished writers, which soon included Diderot, J. J. Rousseau, the baron von Grimm, J. F. de Saint-Lambert and the latter's mistress, Mme d'Houdetot. Grimm began a lifelong liaison with Mme d'Épinay; but Rousseau, whom in 1756 Mme d'Épinay had installed in a specially built "Hermitage" near her château, fell in love with Mme d'Houdetot, quarreled with Grimm and Diderot and withdrew to Geneva. (*See* ROUSSEAU, JEAN JACQUES.) The quarrel was never healed and had constant repercussions in the writings of all concerned. In 1759 Mme d'Épinay began a lasting intellectual friendship with the *abbé* Ferdinando Galiani.

Closely collaborating with Grimm in his famous correspondence, Mme d'Épinay also published books of her own: *Mes moments heureux* (anon., 1758); *Lettres à mon fils* (anon., 1758); and *Conversations d'Émilie* (1774, 1775 and 1781), a treatise on the primary education of girls which received the gold medal for "public utility" from the French Academy in 1783. She died in Paris on April 15, 1783. Much of her celebrity, however, is due to a collection of literary material first published in an uncritical edition 35 years after her death. From analysis of the manuscripts from which this collection was composed, it seems that about 1755 she was engaged on a novel, to be called *Histoire de Madame de Rambure* and to reflect her relations with members of her circle, who were to appear in it under disguised names (Rousseau = René; Grimm = Volx; etc.); that about 1758 or later the title of the projected novel was changed to *Histoire de Madame de Montbrillant;* and that Mme d'Epinay, Diderot and Grimm subsequently rewrote much of the text in a thoroughly anti-Rousseau sense, so as to have it ready as a counterblast to Rousseau's *Confessions*, which they expected to attack them. As the first volumes of Rousseau's *Confessions* (1781) did not cover the years of the quarrel, the *Histoire* was still unpublished at Mme d'Épinay's death. It passed with other manuscripts into Grimm's possession and thence into the French state libraries. The substance of the novel, conflated with genuine letters and other material, was eventually published as *Mémoires et Correspondance de Madame d'Épinay*, 3 vol. (1818). The circumstances of this work's genesis were elucidated in the 20th century.

See G. Roth (ed.), *Les Pseudo-mémoires de Madame d'Épinay*, 3 vol. (1951); L. Perey and G. Maugras, *Une femme du monde au XVIII^e siècle*, 2 vol. (1882–83).

EPINEPHRINE, a substance produced, along with norepinephrine, by the medulla of the adrenal gland. It is a powerful vasopressor drug; *i.e.*, one that increases blood pressure and stimulates the heart.

See ADRENAL GLANDS; ADRENALINE AND NORADRENALINE; HORMONES, VERTEBRATE.

EPIPHANIUS, SAINT (*c.* 315–403), Greek Church Father who wrote a work on heresies, was born at Bezanduca near Eleutheropolis in Palestine. After studying in Egypt, where he became familiar with the monasticism of that country, he returned to Bezanduca and founded a monastery of which he became the superior. In 367 he was made bishop of Constantia (Salamis), the metropolis of Cyprus. He devoted himself to the spread of monasticism and to the confutation of heretics, of whom he regarded Origen and his followers as the chief representatives. In about 390 he visited Jerusalem, where he met Jerome and angered the local bishop by a violent attack on Origen. Later he came into conflict with John Chrysostom, bishop of Constantinople, who was accused by Theophilus, bishop of Alexandria, of sheltering four Nitrian monks expelled from Alexandria on a charge of Origenism.

On arrival at Constantinople Epiphanius realized that he had been misled by Theophilus, and started home but died before reaching Cyprus. His feast day is May 12. A zealous bishop and a man of austere life, Epiphanius was lacking in moderation and judgment, defects reflected in his writings, which have no literary grace.

His chief work, *Panarion,* an account of 80 heresies including Greek philosophical schools, Jewish sects and Samaritans, is particularly useful for its long quotations from lost works of earlier authors. It ends with a statement of orthodox doctrine. He also wrote *Ancoratus* on the true faith. His book on the weights and measures of the Bible is not fully extant in Greek. It begins with an account of the various translations of the Old Testament into Greek, and ends with an explanation of biblical place names found only in the Syriac version. A treatise on the 12 precious stones of Aaron's breastplate survives almost complete in an early Latin version, the Greek text being excerpts only. Two letters, to John of Jerusalem and to Jerome, are extant in Latin versions. Five homilies attributed to Epiphanius date from a later period. His name is also found in connection with a recension of the *Physiologus,* an early text of the *Notitia episcopatuum* (probably of the 7th century) and a list of the prophets, apostles and disciples (probably of the 8th century). Some critics have attributed works wrongly ascribed to Epiphanius to one or other of the two later bishops of Constantia who bore the same name (7th and 10th centuries).

BIBLIOGRAPHY.—Works in J. P. Migne, *Patrologia Graeca*, vol. xli–xliii (1858). Critical edition of *Ancoratus* and *Panarion* by K. Holl, 3 vol. (1915–33) in *Griechische christliche Schriftsteller.* *See* also O. Bardenhewer, *Geschichte der altkirchlichen Literatur*, vol. iii, pp. 293–302 (1923). (K. MCNA.)

EPIPHANY, on Jan. 6, marking the end of the Christmas season, is one of the three principal and oldest festival days of the Christian Church, the other two being Easter and Pentecost. Epiphany (Gr. *epiphaneia,* "manifestation") commemorates the first manifestation of the Lord Jesus Christ to the Gentiles, in the persons of the Magi, and the manifestation of his divinity, as it occurred at his baptism in the Jordan river and at his first miracle at Cana in Galilee. The Greek Orthodox Church still retains the other very old name for this same feast, Theophany (which means a manifestation of God).

The feast of the Epiphany had its early origin in the Eastern Churches where it also included a more or less explicit commemoration of the nativity of Christ, his manifestation in the flesh (*see* CHRISTMAS). The Philocalian or Liberian calendar of the Church of Rome, for the year 354, among other feast days mentions that of the birth of Christ, on Dec. 25: "Christ born in Bethlehem of Judaea." No mention of the Epiphany is found in the Roman calendar at this period. When later, in the same century, the Eastern Churches began to adopt the date of Dec. 25 for Christ's nativity, Rome and the west began to introduce the feast of the Epiphany, on Jan. 6, retaining the principal characteristics it had in the east, but giving primary importance to the visit of the Magi. Because of this new element, the Feast of the Epiphany became popularly known as the Feast of the Three Holy Kings, especially among Germanic people. The Magi, who came to adore the infant Jesus as the newly born king of the Jews, offering him gifts of gold, frankincense and myrrh (Matt. ii), belonged to an old and powerful caste of ancient Persia, having a sacred character and a dominating religious influence on the kings of that land, which they themselves had ruled at one time in the 6th century B.C. Originally, the Magi had been disciples of Zoroaster (*q.v.*), and they were still powerful at the time of the birth of Christ.

The Gospel narrative makes no mention of the number of the Magi that came to visit the child Jesus in Bethlehem. Popular tradition has set their number anywhere from 2 to 12; however, it has favoured, especially in the west, the number 3, no doubt because of the three gifts they offered. The same western tradition has named them Gaspar, Melchior and Balthasar. The east has different names and a different number for the Magi, and Epiphany in the Eastern Churches is concerned mainly with commemorating the baptism of Christ and his first miracle at Cana. Both these events are commemorated in the traditional custom of consecrating baptismal water on the eve of the Epiphany, and in that of blessing rivers, lakes and seas by dipping a crucifix in their waters. *See* also CHURCH YEAR.

BIBLIOGRAPHY.—King A. Archdale, *The Rites of Eastern Christen-*

dom, 2 vol. (1948); *The Catholic Encyclopaedia* (1907–22); Raymond Janin, *Les Églises orientales et les rites orientaux* (1926); K. A. H. Kellner, *Heortologie* (1906); Ricciotti-Zizzamia, *Life of Christ* (1952); L. Duchesne, *Christian Worship*, trans. by M. L. McClure (1903); P. Guéranger, *The Liturgical Year*, 15 vol. (1925). (P. P. P.)

EPIPHENOMENALISM is the theory of the body-mind relation according to which mental events and processes are by-products of physical processes, the necessary results of them, but do not themselves have any effects. Thus a certain thought would be produced by a certain brain state and be a necessary part of a total process, but the whole process would subsequently continue as if the brain state had occurred alone. To make epiphenomenalism intelligible, therefore, we must distinguish between cause and invariable antecedent, for mental events of certain kinds are well known to be invariable antecedents of physical events of certain kinds. This distinction cannot be made empirically, so that epiphenomenalism is a metaphysical rather than a strictly scientific theory, and its function is to direct attitudes and interests. It allows explanatory force to physical laws and concepts only, discouraging the attempt to explain phenomena psychologically or teleologically.

As it cannot be established by empirical means, so epiphenomenalism cannot be decisively refuted by them; but it would be rendered practically untenable if psychical research produced really good evidence for human survival after death. It is in any case an incomplete conceptual scheme, in which mental phenomena, playing no part in explanation, are themselves inexplicable. *See* further BODY AND MIND. (M. KE.)

EPIPHYTE, the term used in botany to denote free-living plants that grow perched upon other plants or a support. Epiphytes are a characteristic feature of tropical forests. They include notably certain orchids and Spanish moss (*qq.v.*). A usual feature of epiphytes is the development of aerial roots. *See* also ROOT.

EPIRUS (EPEIROS), an area of the Balkan peninsula lying across the northwestern frontier of modern Greece and the southern frontier of Albania. In ancient geography, Epirus was the district extending along the Ionian sea from the Acroceraunian promontory (Cape Linguetta or Kep i Gjuhëzës) on the north to the Ambracian gulf (Gulf of Arta) on the south and conterminous on the landward side with the ancient Illyria, Macedonia and Thessaly. The name signifies "mainland" and was originally applied to the whole west coast north of the Corinthian gulf, in contradistinction to the neighbouring islands; the political use of the term is first found in Pindar. The country is mountainous, especially where the rivers Achelous, Arachthus and Aous rise in Mt. Lacmos, the backbone of the Pindus range. The climate is damp, with cold winters. In ancient times Epirus did not produce grain sufficient for its inhabitants, but it has always been celebrated for its cattle and horses.

According to Theopompus (4th century B.C.), the Epirotes were divided into 14 independent tribes, of which the principal were the Chaones (perhaps akin to the Chones in the heel of Italy) along the Acroceraunian shore, the Molossians inland round Lake Pambotis (Ioannina) and the Thesprotians north of the Ambracian gulf. In spite of distance and its reputation for barbarism, Epirus exerted early no small influence on Greece, through the oracle of Dodona (*q.v.*). Herodotus knew Epirus as the home of the Dorians, and to Aristotle it was the original seat of the Hellenes.

Numerous Greek colonies on the coast formed steppingstones toward the Adriatic and the west. One of the earliest and most flourishing was the Corinthian colony of Ambracia, which gives its name to the gulf. Elatria, Bucheta and Pandosia, in Thesprotia, originated from Elis. Other towns of some importance in Chaonia were Palaeste and Chimaera, fortified posts to which country folk could retire in war; Onchesmus (Sarandë or Santi Quaranta) opposite Corcyra; Phoenice, wealthiest of all and after the fall of the Molossian kingdom the centre of an Epirote league; Buthrotum (Butrinto); Phanote, important in the Roman campaigns in Epirus; and Hadrianopolis, named from its founder. In Thesprotia the chief town was Cassope, important seat of the most powerful Thesprotian clan, the Ephyra (Cichyrus). In Molossia, Passaron, where the kings were wont to receive their people's allegiance, and Tecmon, Phylace and Horreum were important centres.

The portion of Epirus in modern Greece comprises the prefectures of Arta, Thesprotia, Ioannina and Preveza with administrative centres at Arta, Igoumenitsa, Ioannina and Preveza respectively. The principal town in the sector included in southern Albania is Gjinokastër (Argyrokastro).

History.—The chieftains of the Molossians, who ultimately dominated all Epirus, claimed to be descended from Pyrrhus, son of Achilles, who settled there after the sack of Troy and transmitted his kingdom to Molossus, his son by Andromache. The early history of the dynasty is obscure, but in the 5th century Pindar celebrated the claims of the royal house, and Admetus is remembered for his hospitable reception of the banished Themistocles, though he was his personal enemy. Admetus was succeeded, about 429 B.C., by his son or grandson, Tharyps or Tharypas. Tharyps, who was educated at Athens and received Athenian citizenship, carried further the hellenization which appears already to have begun about the time of the Greco-Persian Wars.

Under Tharyps' son, Alcetas, from 385 B.C., the Molossians probably took over the control of Epirus from the Chaones. Alcetas established close relations with Jason of Pherae and joined the Athenian confederacy. Alcetas' son, Arybbas, ruled with prudence and equity and gave encouragement to literature and the arts. To him Xenocrates of Chalcedon dedicated his four books on the art of governing; and he bestowed great care on the education of his nieces. One of them, Troas, he married; Olympias, the other, was married to Philip II of Macedonia and became the mother of Alexander the Great. A clash between Philip and Arybbas ended in the expulsion of the latter from Epirus (343) and his subsequent death in exile. In his place Philip enthroned Alexander, the brother of Olympias, who married Philip's daughter Cleopatra and took the title of king. Asked by Tarentum for aid against the Samnites and Lucanians, Alexander landed at Paestum in 332 B.C. and reduced several Lucanian cities, but was defeated and killed near Pandosia.

After an interlude during which Olympias controlled Molossia, Aeacides, a son of Arybbas, occupied the throne, but he was deposed, took refuge in Aetolia and fell in battle against the Macedonians (*c.* 314). He had one son, the celebrated Pyrrhus (*q.v.*). His brother and successor, Alcetas II, fought against Cassander, king of Macedonia, who soon after recognized him; but in 307 Alcetas was murdered by his own subjects and succeeded by Pyrrhus. Pyrrhus, after a period of exile from 302 to 297, established himself as co-ruler with Neoptolemus II (perhaps a grandson of Alcetas II), whom he speedily removed. For six years Pyrrhus fought against the Romans in southern Italy and the Carthaginians in Sicily, giving Epirus its sole moment of influence in world history.

Between 329 and the time of Pyrrhus, Epirus was organized in a federation. This is attested by inscriptions and is probably the Epirote constitution treated by Aristotle. This federation was dominated by Molossia, but there is no evidence that the Molossian kings were ever "kings of the Epirotes."

Alexander, the son of Pyrrhus, who succeeded in 272, defeated Antigonus Gonatas but was himself expelled from his kingdom by Demetrius II of Macedonia. He recovered it, however, and spent the rest of his days in peace. Two other insignificant reigns brought the family of Pyrrhus to its close. Shortly before 231 a revolution swept away the remnants of the dynasty, and a republic was set up with an assembly meeting at Phoenice, which had replaced Ambracia as capital. Epirus joined the Greek league set up by Antigonus Doson of Macedonia, fought beside Philip V against Aetolia, but was able to maintain a *de facto* neutrality in the Roman wars against Philip and Antiochus. Having imprudently supported Perseus of Macedonia in 168, Epirus was devastated by Aemilius Paullus after the battle of Pydna; 150,000 citizens were enslaved and the area never recovered. The confederacy continued to exist for a time, but the organization of Epirus after Rome's final victory in Greece in 146 is unknown. In 27 B.C. most of it formed part of the province of Achaea, and

the rest belonged to Macedonia. Roads were built, a Roman colony was founded at Buthrotum, and a new Greek city, Nicopolis Actia, was built.

On the division of the Roman empire Epirus fell to the east, and so remained until the taking of Constantinople by the Latins in 1204 when Michael Angelus Comnenus, a Byzantine prince, seized Aetolia and Epirus. On the death of Michael these countries fell into the hands of his half-brother Theodore. Thomas, the last of the direct line, was murdered in 1318 by his nephew and his dominions were dismembered. Not long after, Epirus was overrun by Serbs and Albanians. Charles II Tocco, count of Cephalonia, was recognized as despot of Epirus by the emperor Manuel II Palaeologus in the beginning of the 15th century, but his family was deposed in 1431 by the Ottoman sultan Murad II. In 1443 Skanderbeg (*q.v.*) made himself master of a considerable part of Epirus, but on his death it fell to the Venetians, and from these passed to the Turks.

The southeastern sector of Epirus was incorporated in modern Greece by the conference of Constantinople (1881); the southwestern sector by the treaty of London (1913) after the First Balkan War. The independence of Albania, which includes the northern sector of Epirus, was also recognized in 1913. Having been annexed by Italy with the rest of Albania in April 1939, Gjinokastër (Argyrocastro) was temporarily occupied by Greek troops after the Italian attack on Greece in World War II.

See also references under "Epirus" in the Index volume.

BIBLIOGRAPHY.—G. N. Cross, *Epirus* (1932); S. I. Oost, *Roman Policy in Epirus . . . in the Age of the Roman Conquest of Greece* (1934); P. R. Franke, *Alt-Epirus and das Königtum der Molosser* (1955). (F. W. WA.)

EPISCOPACY, the office of a bishop, the system of church government which comprises three distinct orders, bishops, priests and deacons. The origins of episcopacy are obscure (*see* MINISTRY, CHRISTIAN), but by the 2nd century A.D. it was beginning to be accepted in the main centres of Christianity and, with it, the idea of apostolic succession.

Apostolic succession, as it was first understood, was primarily historical; it meant succession to a particular chair, to the bishopric of a particular church, and not simply to the possession of episcopal orders. The bishop of any community had been entrusted with the cathedra or teaching chair in succession to a predecessor who in his turn had been the recognized holder of the chair and so on back to the foundation of the community. If the church had been founded or organized by an apostle, then there was a direct formal connection between its episcopate and the apostles; if the church was the offshoot of a mother church, then the daughter church would trace its succession back via its own bishops and those of the mother church to the apostles. Hence the appearance in the 2nd century of the various episcopal lists. Toward the end of the century the historical succession was given a sacramental interpretation, to the effect that the succession was primarily the transmission of a peculiar gift to the individual by means of a particular sign—that is, episcopal consecration—whereby the gift and the power to transmit had been handed down in an unbroken line from one of the original apostles.

The 2nd-century bishop was charged with the spiritual welfare of his flock; he was the chief liturgical minister; he baptized, celebrated the Eucharist, ordained, absolved, controlled finances and settled matters of dispute. With state recognition of Christianity in the 4th century, the bishop came to be regarded not only as a leader of the church but also as an important figure in secular affairs, and this latter aspect of his work received more and more emphasis.

As congregations grew in size and number, it became necessary either to have more bishops or to delegate some of their functions to others. In certain areas the first of these expedients was adopted. So in north Africa a bishop of full rank was appointed wherever a Christian community existed. In Syria and parts of Asia Minor each country church was provided with a resident staff supervised by *chorepiscopoi* or itinerant bishops who were answerable to the bishop of the nearest large city. Elsewhere the second of the possible courses was followed: so in the districts around Alexandria the villages were entrusted to presbyters under the supervision of the bishop of Alexandria, and it was this practice that was most widespread in western Europe, where the bishop retained as his exclusive right only the power to confirm, ordain and consecrate (*see* HOLY ORDERS). As the middle ages advanced, the system of delegation became excessively organized and an ecclesiastical bureaucracy came into being. A complex hierarchy of subordinate officials acted on the bishop's behalf and, in many instances, the bishop himself was no more than a prisoner of the system. Although his contribution to the medieval state was immense, this was achieved only by a corresponding loss of spiritual efficiency. Indeed the bishop tended to become more of a civil servant or temporal lord than a shepherd and overseer of his people.

At the Reformation, episcopacy was repudiated by many of the Protestant churches, partly on the grounds of its corruption and its transformation into prelacy but also upon the thesis that it could claim no justification from the New Testament. Its preservation, within the historic succession, was and is claimed by the Roman, Old Catholic, Eastern Orthodox, Oriental and Anglican Churches. Within these communions, however, the influence of the Reformation is to be noted in the changed conception of episcopacy, so that the bishop is no longer regarded as a civil administrator but as a pastor whose primary duty is the spiritual oversight of his people. Episcopacy is also accepted by certain members of the Lutheran family of churches, by the Moravians and by the Methodist Episcopal bodies in the United States.

In Sweden episcopacy within the historic succession would seem to have been preserved. The first reformed archbishop, Laurentius Petri, was consecrated by Petrus Magni, bishop of Västerås, and Magnus Sommar, bishop of Strängnäs, in 1531, and the succession has been maintained since. In 1593, however, the Lutheran Confession of Augsburg was accepted; this states that no particular form of church organization is necessary. Thus the church of Sweden is Lutheran and acknowledges no one form of ministry as essential but has retained the historic episcopate.

Episcopacy is also to be found in the Lutheran Churches of Denmark, Finland, Iceland and Norway; in these, however, there is no question of succession, in the generally accepted sense, since the first reformed "bishops" were consecrated by simple priests.

Episcopacy within German Lutheranism postdates the Reformation. The title was revived in certain areas as a consequence of the creation of the Evangelical Union by Frederick William III of Prussia in 1817; but it was not until the reorganization of the German Lutheran Churches after World War I that territorial bishops (*Landesbischöfe*) came into being in such territories as Saxony, Hanover, etc.

The Moravian Church or Unitas Fratrum, claims to be an episcopal church, but the bishops are excluded from governing and their authority extends only to the performance of ordinations. The origin of their succession is to be sought in the 15th century, when a certain Roman priest named Michael Bradacius obtained authorization from a Waldensian "senior" and consecrated Matthias of Kunwald as their first "bishop."

Episcopacy found a place in American Methodism from its beginnings. John Wesley believed that "bishops and presbyters are essentially of one order." Consequently when R. Lowth, bishop of London, refused to ordain men for work in the United States, Wesley took upon himself to do so, and on Sept. 1, 1784, in company with Thomas Coke and another Anglican clergyman, he made two deacons. On the following day these two were ordained elders and Coke was made a "superintendent, for the Church of God, under our care in America." Whatever Wesley's intention, Coke regarded himself as a bishop. He himself consecrated Francis Asbury and the Methodist Episcopal Church came into existence. This body recognizes two orders, that of deacons and elders; their bishops, although elected and consecrated, are regarded as elders merely, but possessed of the function to ordain. So the ministry is Presbyterian in origin, tripartite in function and monepiscopal in practice.

In the 20th-century ecumenical movement (*q.v.*) episcopacy inevitably becomes a focus of debate between churches seeking

reunion. There are those who maintain that it is necessary to the church, those who think it is beneficial to the church and those who consider it to be neither necessary or beneficial. Most Christians would agree that *episcope*, in its original Greek sense of "oversight," is essential to the church, but they differ as to who may be said to exercise it. So, for example, Methodists in England locate *episcope* in their annual conference, while those churches that have the historic episcopate see the bishops as the sole organs duly authorized to exercise it.

See C. Jenkins and K. D. Mackenzie (eds.), *Episcopacy Ancient and Modern* (1930); H. B. Swete (ed.), *Essays on the Early History of the Church and the Ministry* (1918). *See* also bibliography to MINISTRY, CHRISTIAN. (J. G. DA.)

EPISCOPAL CHURCH: *see* ANGLICAN COMMUNION; ENGLAND, CHURCH OF; PROTESTANT EPISCOPAL CHURCH.

EPISCOPIUS, SIMON (SIMON BISCOP, BISCHOP or BISSCHOP) (1583–1643), Dutch theologian, the systematizer of Arminianism (*q.v.*), was born at Amsterdam on Jan. 1, 1583. In 1600 he entered the University of Leiden, where he studied theology under and came under the influence of Jacobus Arminius. In 1610, the year in which the Arminians presented the famous Remonstrance to the states of Holland, he became pastor at Bleyswick, near Rotterdam; in the following year he supported the Remonstrants conference. In 1612 he was made professor of theology at Leiden. Episcopius was spokesman of the 13 representatives of the Remonstrants before the synod of Dort in 1618; but he was refused a hearing, the Remonstrant doctrines were condemned and he and the other Arminian representatives were banished. After the death (1625) of Prince Maurice, the Arminian controversy abated and Episcopius was permitted in 1626 to return to his own country. He was appointed preacher at the Remonstrant college in Amsterdam, and died there on April 4, 1643. In theology, Episcopius protested against the tendency of Calvinists to stress abstract dogma, especially predestination and original sin, and argued that Christianity was practical rather than theoretical. He was author of the Arminian confession of faith (published 1622). His *Opera Theologica*, edited posthumously by S. Curcellaeus and P. van Limborch, was published in two volumes at Amsterdam (1650–65).

Episcopius' biography was written by P. van Limborch (1701). *See* also F. Calder, *Memoirs of Simon Episcopius* (1835).

EPISCOPI VAGANTES (Lat. "wandering bishops"), bishops without canonical authority who intrude themselves into the jurisdiction of other bishops. The problem of such bishops arose early in church history as is shown by the attempts of the Council of Antioch (341) to control them. Various causes contributed to the problem, chiefly the consecration of bishops without diocesan status or jurisdiction. Moreover the theological controversies of the 4th and 5th centuries often resulted in bishops' being deprived of their sees; they retained their orders but had to wander to seek a livelihood. Vagancy also could arise from misfortune.

In later times the number of episcopi vagantes was augmented by wandering Irish bishops and bishops driven out of their sees by war, especially in Spain and East Prussia; this state of affairs continued during the period of the crusades, when titular bishops were often consecrated for sees under infidel occupation. It was not until after the Council of Trent (1545–63) that the activity of episcopi vagantes was restricted in the Western Church.

In modern times episcopi vagantes have appeared for quite other reasons outside the control of any ecclesiastical authority; they claim valid episcopal orders but are not recognized by any church body outside their own sects. Emphasis was placed upon episcopal orders partly as a result of the disputes about valid orders in the 19th century, which culminated in the papal condemnation of Anglican orders in 1896. A few Anglican clergy received clandestine episcopal consecration, but the serious increase in consecrations began with the consecration by Old Catholic bishops in 1908 of Arnold Harris Mathew, which was later officially described as having been obtained by misrepresentation. Mathew consecrated eight or nine persons who in turn consecrated others.

Another source of bishops was Joseph René Vilatte, consecrated in Ceylon in 1892, who claimed succession from a Syrian bishop of south India. Dozens of bishops now claim episcopal orders from these or other sources. Consecrations continue to take place, and some of the later candidates have gone to great pains to unify the various lines of succession at their own consecrations.

Episcopi vagantes have founded or taken over more than 20 churches whose titles often contain one or more of the following words: Old, Catholic, orthodox, episcopal, sometimes combined with protestant or evangelical. Many of their "churches" are mere paper organizations, and the rest have no more than a handful of adherents. Their high-sounding names often lead people to suppose that they are either Orthodox Eastern or Old Catholic. In fact they are recognized by none of the traditional churches of Christendom, even though their orders might be held to be technically valid according to the Latin view of orders. Ordination *sub conditione* is required in the case of such episcopi vagantes desiring to become Anglican clergy.

See H. R. T. Brandreth, *Episcopi Vagantes and the Anglican Church* (1961); A. J. Macdonald, *Episcopi Vagantes in Church History* (1945). (H. M. W.)

EPISTAXIS: *see* NOSEBLEED.

EPISTEMOLOGY (Greek, *episteme*, "knowledge"; *logos*, "account"), that branch of philosophy which is concerned with problems of the nature, limits and validity of knowledge and belief. The term "epistemology" was first used in the 19th century, as a synonym for "theory of knowledge," but western philosophical thought on the subject has a history reaching back to Plato and Aristotle. Traditionally the range of its problems has served to contrast it on the one hand with metaphysics, concerned with the nature of reality, and on the other hand with logic, concerned with the forms of valid argument. But, with the modern recognition that much that had previously been regarded as philosophy's function properly belonged to the natural sciences, what was left of metaphysics tended to be absorbed into epistemology. This process began with Descartes and was carried on by Locke and the British empiricists and, above all, by Kant, whose investigations into the extent and conditions of the possibility of knowledge by the human understanding were intended to determine the boundaries between knowledge and rational belief on the one hand and what he regarded as the illusions of metaphysics on the other. *See* KNOWLEDGE, THEORY OF. (A. D. W.)

EPISTLE, in its original sense a word meaning simply a letter, has come to be used only of formal letters written in ancient times, or of an elaborate kind of literary production, akin to the ode and the elegy, which, while addressed to a particular person, is concerned with public rather than personal affairs or expresses a universal feeling on a particular occasion. The letter is essentially a spontaneous, nonliterary production, although, by the eminent, it may be written with an eye to publication, and may give pleasure, when published, to many for whom it was not primarily intended. It is a substitute for spoken conversation, and its quality, even when published, depends on its impression of intimacy and ease and its power to convey the writer's personality. The epistle bears much the same relation to it as does the Platonic dialogue to the conversation between friends: it is written for an audience, with a conscious artistry and elaboration of style, to develop an argument or theme.

Even in classical times, the word *epistola* came to have the additional meaning of an imperial decree, and the epistles which have come down from the ancient world are usually the public utterances of great men, akin to speeches in subject matter and formality. The posthumous value placed on the letters of a great man led naturally to the production of epistles written under his name to set forth the views of a particular school of thought, or to express what were thought likely to be his views on some topic of moment. Pseudonymous epistles of this kind were numerous under the early Roman empire, and mainly attached themselves to the names of Plato, Demosthenes, Aristotle and Cicero. Other writers of epistles which have been preserved from classical times are Isocrates and Epicurus, Plutarch, Seneca and the younger Pliny.

The Pauline Epistles are of particular interest, and in classifying them, and the other letters and Epistles in the New Testament, some biblical scholars have sought to develop the idea of the divi-

sion between letters and Epistles, and to distinguish a new category "the general letter addressed to a whole class of persons or to the whole Church of Christ." It is notable that even in his formal, expository or Pastoral Epistles, St. Paul includes personal messages characteristic of the letter. These Epistles resemble the "imperial edicts and rescripts by which Roman law grew, documents arising out of special circumstances but treating them on general principles." From the reading of a passage from one of the New Testament Epistles, as well as from one of the Gospels, at the Roman Catholic Mass and the Anglican communion service derives one of the commonest applications of the word today: the Missal and the Book of Common Prayer contain "epistles" and "gospels" for every Sunday and other feast day in the Christian year. Papal encyclical letters, addressed to the whole church, on questions of faith and practice, have carried on the apostolic epistolary tradition.

An interesting development of the epistle in the post-apostolic age was the communication between branches of the church, rather than individuals, such as the *First Epistle of Clement* (Rome to Corinth), the *Martyrdom of Polycarp* (Smyrna to Philomelium), and the *Letters of the Churches of Vienne and Lyons* (to the congregations of Asia Minor and Phrygia), describing the Gallican martyrdoms of A.D. 177. In the following centuries epistles on doctrine and practice were written by the Christian Fathers—Cyprian, Ambrose, Chrysostom, Augustine and Jerome. In the secular world, at the same period, epistles developed increasingly into a branch of literature, and the ten books of Symmachus' *Epistolae*, so highly esteemed in the 4th century, are models of elegant style. They are in the line connecting Cicero's letters (many of them, strictly speaking, epistles for general publication) with the Renaissance literary epistles, supremely exemplified by Petrarch, which are modeled on Ciceronian style. The literary epistle developed as a vehicle for satire, as in Swift's *Drapier Letters* and the *Letters of Junius*, for religious and contemplative ideas, as in Pascal's *Les Provinciales*, and for polemics, as in the epistles of Martin Marprelate.

Poetic Epistles.—A branch of poetry bears the name of the epistle, and is modeled on those pieces of Horace which are almost essays on moral or philosophical subjects, and are chiefly distinguished from other poems by being addressed to particular patrons or friends. Examples of the work in this direction of Ovid, Claudian, Ausonius and other late Latin poets have been preserved, but it is particularly those of Horace which have given this character to the epistles in verse which form so very characteristic a section of French poetry. Clément Marot, in the 16th century, first made the epistle popular in France. Nicolas Boileau's 12 epistles (composed 1668–95) are the classic examples of this form of verse in French literature. In the 18th century Voltaire enjoyed a supremacy in this grateful and sparkling species of writing; the *Epitre à Uranie* is perhaps the most famous of his verse letters.

In England the verse epistle was first prominently used by Samuel Daniel in his "Letter from Octavia to Marcus Antonius" (1599), and later on, more legitimately, in his *Certain Epistles* (1601–03). His letter, in *terza rima*, to Lucy, countess of Bristol, is one of the finest examples of this form in English literature. It was Daniel's deliberate intention to introduce the epistle into English poetry, "after the manner of Horace." He was supported by Ben Jonson, who has some fine Horatian epistles in his *Forests* (1616) and his *Underwoods*. *Letters to Several Persons of Honour* form an important section in the poetry of John Donne. At the close of the 17th century Dryden greatly excelled in this class of poetry, and his epistles to Congreve (1694) and to the duchess of Ormond (1700) are among the most graceful and eloquent that we possess. During the age of Queen Anne various Augustan poets essayed the epistle with more or less success, and it was employed by John Gay for several exercises in his elegant persiflage. But the great writer of epistles in English is Pope. The "Epistle to Dr. Arbuthnot" has not been surpassed, if equaled, in Latin or French poetry of the same class. Pope excelled not only in the voluptuous or didactic epistle but in that of compliment, of which a graceful example is his letter to Robert, earl of Oxford (1721). After Pope the epistle again fell into desuetude, or occasional use, in England. It revived in the charming naïveté of Cowper's lyrical letters in octosyllabics to his friends, such as William Bull and Lady Austin (1782). At the close of the 18th century Samuel Rogers endeavoured to resuscitate the neglected form in his "Epistle to a Friend" (1798). Shelley's "Letter to Maria Gisborne" (1820), Keats's "Epistle to Charles Clarke" (1816), and Landor's "To Julius Hare" (1836), in spite of their romantic colouring, are genuine Horatian epistles and of the pure Augustan type. This type, in English literature, is commonly cast in heroic verse. But Daniel employs rime royal and *terza rima*, while some later epistles were written in iambic rhymed measures or in blank verse.

EPITAPH, an inscription upon a tomb, or by extension, anything written as if to be inscribed on a tomb. Probably the earliest surviving are those of the ancient Egyptians, written on the sarcophagi and coffins. Those that have been deciphered are all very much in the same form, beginning with a prayer to a deity (generally Osiris or Anubis) on behalf of the dead person, whose name, descent and office are usually stated. Ancient Greek epitaphs, unlike the Egyptian, are often of great literary interest, deep and tender in feeling, rich and varied in expression, and epigrammatic in form. They are usually in elegiac verse, though many of the later epitaphs are in prose. Among the most familiar are those, ascribed to Simonides of Ceos (*c.* 556–469 B.C.), on the heroes of Thermopylae, the most famous of which has been translated:

> Go tell the Spartans, thou that passest by,
> That here, obedient to their laws, we lie.

In Sparta epitaphs were inscribed only upon the graves of those who had been especially distinguished in war, but in Athens they were applied more generally.

Roman epitaphs, in contrast to the Greek, contained as a rule nothing beyond a record of facts. The inscriptions on the urns usually present little variation: the letters D. M. or D. M. S. (*Diis Manibus* or *Diis Manibus Sacrum*, "dedicated to the souls of the departed") are followed by the name of the person whose ashes are enclosed, his age at death and sometimes one or two other particulars. Roman tombs being often by the wayside, inscriptions usually begin *Siste* (or *Aspice*) *Viator*, which survives in the common formula "Stop, passer-by"; another Roman tombstone phrase which continued in frequent use is *Sit tibi terra levis*, "may the earth lie light upon thee." A satirical inversion of this is seen in the epitaph by Abel Evans (1679–1737) on the architect Sir John Vanbrugh:

> Lie heavy on him, earth, for he
> Laid many a heavy load on thee.

Many Roman epitaphs included a denunciation on any who should violate the sepulchre; a similar later denunciation is found on Shakespeare's tomb (probably not by the poet himself):

> Good frend, for Jesus sake forbeare
> To digg the dust enclosed heare;
> Blest be ye man yt spares thes stones,
> And curst be he yt moves my bones.

The oldest existing British epitaphs date from the Roman period, and are in Latin; a collection of these is to be found in vol. vii (ed. by E. Hübner) of the *Corpus Inscriptionum Latinarum* (1863–). The earliest are those of soldiers of the legions of occupation, but the Roman form was soon adopted for native Britons. Latin, indeed, continued for many centuries to be a preferred language for epitaphs: in 1776 Dr. Johnson, asked to write an English epitaph for Goldsmith, replied that he would "never consent to disgrace the walls of Westminster abbey with an English inscription." Famous epitaphs in Latin include Sir Christopher Wren's, in St. Paul's—*Si monumentum requiris, circumspice* ("if you seek his monument, look around you"); and that of Swift, written by himself, in St. Patrick's, Dublin, which speaks of him as lying *Ubi saeva indignatio cor ulterius lacerare nequit* ("where fierce indignation can no longer tear the heart").

The earliest epitaphs in English churches are invariably in Latin, and are usually a simple statement of name and rank, with a *Hic jacet* ("here lies"). In the 13th century French began to be used —on, for example, the tomb of Henry III (d. 1272) at Westminster, and on that of Edward the Black Prince (d. 1376) at Canterbury; indeed, most of the epitaphs of celebrated persons between 1200 and 1400 are in French. Such epitaphs generally conclude

with an urgent request for prayers for the deceased. English began to be used around the middle of the 14th century, the inscriptions being of the same general character as those in French.

Until the Reformation most of the inscriptions that have survived were upon brasses (*see* METALWORK, DECORATIVE). By Elizabethan times, however, epitaphs upon stone monuments, in English, became much more common, and began to assume a more literary character. Thomas Nashe tells how, by the end of the 16th century, the writing of verse epitaphs had become a trade, and George Puttenham, in *The Arte of English Poesie* (1589), treats of the epitaph as a separate verse form, akin to the epigram —"an inscription such as a man may commodiously write or engrave upon a tombe in few verses, pithie, quicke and sententious, for the passer-by to peruse and judge upon without any long tariaunce." Many of the best-known epitaphs are, in fact, primarily literary memorials, not necessarily ever intended to be placed on a tomb. Among the finest may be mentioned that by William Browne (1591?–1645?) on the countess of Pembroke, Ben Jonson's on the boy actor Salomon Pavy, Herrick's *Upon a Child*, Milton's on Shakespeare, and R. L. Stevenson's *Requiem* for himself. Pope wrote several epitaphs, but they are not among his best work; they inspired, however, one of the few monographs on the subject—Dr. Johnson's examination of them in *The Universal Visiter* for May 1756. They may also have prompted Wordsworth's three essays on epitaphs (1810), for much of his space is devoted to an attack on Pope for his "insincerity" and the "artificiality" of his diction. It is true to say, however, that the great majority of serious epitaphs are extremely banal, despite the sincerity of the feeling behind them. On the other hand, an unstudied simplicity sometimes produces an epitaph which is genuinely moving, even when it has unintentionally comic overtones, such as this from Gloucestershire:

> Here lies two babbies, dead as nits,
> Who died in agonising fits;
> They were too good to live with we,
> So God did take to live with He.

Semiliteracy often produces epitaphs which are comic through grammatical accident—for example, "Erected to the memory of/ John MacFarlane/Drowned in the Water of Leith/By a few affectionate friends"; far more common, though, are the deliberately witty epitaphs, which abound in Britain and America. Acrostics, palindromes, riddles, puns on names and professions, are found: Benjamin Franklin's epitaph for himself plays on his trade as a printer, hoping that he will "appear once more, in a new and more beautiful edition, corrected and amended by the Author"; the antiquary Thomas Fuller has the inscription "Fuller's Earth"; Alexis Soyer, the great French cook, has "Soyer Tranquille," and so on. Many offer some wry comment, such as John Gay's

> Life is a jest, and all things show it;
> I thought so once, and now I know it.

The epitaph was also seen as an opportunity for epigrammatic satire, as in the earl of Rochester's lines on Charles II, "Who never said a foolish thing/Nor ever did a wise one."

Looking at any collection of epitaphs we may wonder, with Charles Lamb, "but where are all the *bad* people buried?"; nevertheless, within the conventions imposed by the general necessity of providing a brief memorial tribute, epitaphs provide a fascinating guide to the interests and tastes of a nation at any time, and exemplify every variety of sentiment, from pathos and dignity to scurrility.

BIBLIOGRAPHY.—J. Weever, *Ancient Funerall Monuments* (1631; W. Tooke's ed., 1767); P. Labbe, *Thesaurus epitaphiorum* (1666); J. Hackett, *Select and Remarkable Epitaphs* (1757); P. de La Place, *Recueil d'Épitaphes sérieuses, badines, satiriques et burlesques* (1782); W. Pulleyn, *Church-yard Gleanings* (*c.* 1830); L. Lewysohn, *Sechzig Epitaphien von Grabsteinen des israelitischen Friedhofs zu Worms* (1855); T. J. Pettigrew, *Chronicles of the Tombs* (1857); S. Tissington, *Epitaphs* (1857); E. le Blant, *Inscriptions Chrétiennes de la Gaule antérieures au VIII^e siècle*, 2 vol. (1856–65); J. Kenrick, *Roman Sepulchral Inscriptions* (1858); J. McCaul, *Christian Epitaphs of the First Six Centuries* (1869); H. J. Loaring, *Quaint, Curious and Elegant Epitaphs* (1872); J. R. Kippax, *Churchyard Literature, a Choice Collection of American Epitaphs* (1877); W. Wordsworth, *Prose Works*, ed. by W. Knight, vol. 2 (1896); W. H. Beable, *Epitaphs; Graveyard Humour and Eulogy* (1925); R. W. Ketton-Cremer, *Lapidary Verse* (1960). (R. P. C. M.)

EPITHALAMIUM, a nuptial song or poem. In ancient Greece the singing of songs to the bride and bridegroom at their wedding was a traditional way of invoking good fortune on the marriage and often of indulging in some indecencies. By derivation the epithalamium should be sung at the marriage chamber, but the word is also used to include the song sung during the wedding procession which contained repeated invocations to Hymen (Hymenaeus), the Greek god of marriage. The earliest evidence for literary epithalamia are the fragments from Sappho's seventh book. The *Peace* and the *Birds* of Aristophanes conclude with epithalamia, and in Euripides' *Troades* Cassandra sings her own unhappy epithalamium. Theocritus' 18th idyll is an epithalamium in honour of Helen and Menelaus. The earliest surviving Latin epithalamia are Catullus' 61st, 62nd and 64th poems, the last being known as the epithalamium of Peleus and Thetis. Later writers of epithalamia include Statius, Claudianus and the Christian poets Paulinus of Nola and Venantius Fortunatus. Epithalamia based on classical models began to be written at the Renaissance, for instance in French by Ronsard and in Italian by Tasso. The best-known epithalamium in English literature is that by Spenser, for William Ponsonby's marriage, but various other poets have composed epithalamia including Donne, Herrick and Shelley. No special metre has been associated with the epithalamium in either antiquity or modern times.

EPITHELIUM, in anatomy, a layer of cells closely bound to one another to form continuous sheets that cover every surface of the body that may come into contact with foreign substances. By the formation of outgrowths or ingrowths from these surfaces, structures, consisting largely or entirely of cells directly derived from the surface epithelium, may be formed. In this way originate the central nervous system, the sensitive surfaces of the special sense organs, the glands, and the hairs, nails, etc. The epithelial cells possess typical microscopical characteristics that enable them to be readily distinguished from all others. Thus the cell outline is clearly marked, the nucleus large and spherical or ellipsoidal. The protoplasm of the cell is usually large in amount and often contains large numbers of granules.

FIG. 1.—CILIATED COLUMNAR EPITHELIAL CELLS, HIGHLY MAGNIFIED, FROM THE INTESTINE OF A FROG

Function.—Epithelium may be protective, absorptive or secretory. It may produce special outgrowths (hairs, nails, horns on animals, etc.), and for such purposes it may manufacture within itself chemical material best suited for that purpose (*e.g.*, keratin); here the whole cell becomes modified. In other instances it contains fat droplets, granules of various kinds, protein, mucin, watery granules or glycogen, etc.

FIG. 2.—COLUMNAR EPITHELIAL CELLS, CILIATED, RESTING ON A BASEMENT MEMBRANE. HIGHLY MAGNIFIED

In a typical absorbing cell, granules of material being absorbed may be seen. A secreting cell of normal type forming specific substances stores these in its interior until wanted; *e.g.*, fat (as in sebaceous and mammary glands), ferment precursors (salivary, gastric glands, etc.), and various excretory substances (as in the renal epithelium) (*see* GLANDS).

Varieties.—The cells forming an epithelial membrane may be columnar, cubical, squamous (flattened), irregular or ciliated. The membranes formed by these cells may be only one cell thick, as in the major part of the gastro-intestinal tract, or may consist of several layers, as in the epidermis of the skin.

FIG. 3—MOSAIC APPEARANCE OF COLUMNAR EPITHELIUM AS SEEN FROM ABOVE. MAGNIFIED

Columnar Epithelium.—This variety (fig. 1 and 2) covers the intestinal tract from the end of the esophagus to the be-

ginning of the rectum. It also lines the ducts of many glands. In a highly typical form it covers the villi of the small intestine (fig. 1). The close apposition of these cells to form a closed membrane is well seen when a surface covered by them is examined from above (fig. 3).

Cubical Epithelium.—This differs from the former in that the cells are shorter. It is found in many glands and ducts (*e.g.*, the kidney), in the middle ear, choroid plexuses of the brain, etc.

Squamous or Flattened Epithelium.—In this variety (fig. 4), the cell is flattened, very thin and irregular in outline. It occurs as the covering epithelium of the alveoli of the lung, of the kidney glomeruli and capsule, etc. The surface epithelial cells of a stratified epithelium (*see* below) are also of this type.

FIG. 4.—CELLS OF SQUAMOUS EPITHELIUM FROM THE MOUTH. HIGHLY MAGNIFIED

FIG. 5.—CILIATED COLUMNAR EPITHELIUM CELLS FROM TRACHEA. HIGHLY MAGNIFIED

Ciliated Epithelium.—The surface cells of many epithelial membranes carry fine protoplasmic processes or cilia. Most commonly the cells are columnar (*see* fig. 5), but other shapes also are found. During life the cilia are always in movement, and set up a current tending to drive fluid or other material on the surface in one direction along the membrane or tube lined by such epithelium. Ciliated epithelium lines the trachea, bronchi, parts of the nasal cavities and the uterus, oviduct, vas deferens, epididymis, etc.

If there is only one projection from the exposed surface of the cell it is usually large and long, and is known as a flagellum. Flagellated cells are common on the surface of many simple animal organisms.

Stratified Epithelium.—When the cells of an epithelial surface are arranged several layers deep, other types can be distinguished (fig. 6 and 7). Stratification is found in the epithelium of the skin and of many mucous membranes (mouth, esophagus, rectum, conjunctiva, vagina, etc.). Here the surface cells are very much flattened, those of the middle layer are polyhedral and those of the lowest layer are cubical or columnar. This type of epithelium covers surfaces exposed to friction. The surface may be dry, as the skin, or moist, as, for example, the mouth.

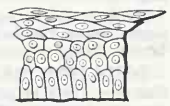


FIG. 6.—STRATIFIED EPITHELIUM FROM A MUCOUS MEMBRANE. HIGHLY MAGNIFIED

The surface cells are constantly being rubbed off and are replaced by new cells growing up from below. Hence the deepest layer is formative, and in successive stages from this can be traced the gradual transformation of these protoplasmic cells into scaly cells, which no longer show any sign of being alive. In the moist mucous surfaces the number of cells forming the epithelial layer is usually much smaller than in a dry stratified epithelium.

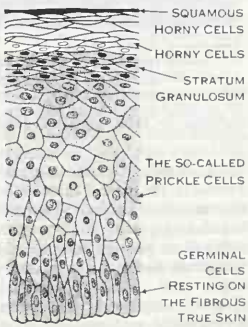


FIG. 7.—STRATIFIED SQUAMOUS EPITHELIUM OF SKIN. HIGHLY MAGNIFIED, SHOWING KINDS OF CELLS

Stratified Ciliated Epithelium.—In this variety the superficial cells are ciliated and columnar; between the bases of these are found fusiform cells, and the lowest cells are cubical or pyramidal. This epithelium lines parts of the respiratory passages, the vas deferens and the epididymis.

Transitional Epithelium.—This variety of epithelium (fig. 8) lines the urinary bladder, and its appearance depends upon the contracted or distended state of the bladder from which the preparation was made. If the bladder was contracted the form seen in fig. 8 is obtained. If the bladder is distended before the preparation is made, the cells are stretched transversely. The surface cells then become very flattened.

FIG. 8.—TRANSITIONAL EPITHELIUM FROM THE URINARY BLADDER. HIGHLY MAGNIFIED

The epithelium is in three or more layers, the superficial one being very characteristic. The cells are cubical and fit over the rounded ends of the cells of the next layer. The latter are pear-shaped, the points of the pear resting on the basement membrane. Between the bases of these pear-shaped cells lie those of the lowermost layer. These are irregularly columnar.

See ENDOTHELIUM; SKIN: *Epidermis; see* also references under "Epithelium" in the Index volume.

EPODE, in classical literature, the name both of the third section of a choral ode (*q.v.*), following the strophe and antistrophe, and also of a complete poem (composed for example by Archilochus and Horace) consisting of alternate long and short lines unless written in elegiac couplets.

EPONA. This goddess, as her name implies (*epo-*, Gaulish equivalent of Latin *equo-*), was patroness of horses, and also of asses and mules. The majority of inscriptions and images bearing her name have been found in Gaul, Germany and the Danube countries; of the few that occur in Rome most have been found on the site of the barracks of the *equites singulares*, a foreign imperial bodyguard recruited mainly from the Batavians. Her cult does not appear to have been introduced into Rome before imperial times, when she is often called Augusta and invoked on behalf of the emperor and the imperial house. The Romans used to place the image of the goddess, crowned with flowers on festive occasions, in a sort of shrine in the centre of the architrave of the stable. In art she is generally represented seated, with her hand on the head of the accompanying horse or ass.

EPPING, a market town and urban district of Essex, Eng., 18 mi. N.N.E. of Charing Cross, London, by road. Underground trains of the Central line go to Epping. Pop. (1961) 9,998. The town lies high and is picturesquely situated at the northern outskirts of Epping forest; it is mostly strung along the sides of the London-Cambridge road. Epping has an iron foundry and makes agricultural machinery and conveyers. There is a Monday market; a charter for a market was granted by Henry III. (X.)

Epping forest was formerly part of the great forest of Essex (*q.v.*) and served as a royal hunting ground. By degrees most of the land was enclosed for agriculture but, as a result of opposition from the holders of common rights, this process was stopped by the Epping Forest act of 1871. The 5,600 ac. that remained were placed in the care of the Corporation of the City of London. The forest extends southward from Epping town for seven miles, forming a rather narrow belt along a ridge of clay soil between the valleys of the rivers Roding and Lea. It is traversed by the high-road to Walthamstow and toward the south becomes hemmed in by the London suburbs. The principal tree is the hornbeam, which was formerly lopped at a height of about six feet to provide firewood for the commoners and "browse" or winter feed for their cattle. There are also many fine beeches and some old oaks. A herd of dark-coloured fallow deer, survivors from the days when Queen Elizabeth I used the still-standing hunting lodge, haunts the woods, which are also rich in bird life. The hornbeam groves are diversified with open meadows, still grazed by the commoners' cattle, and with ornamental lakes used for boating. The aim of the forest management is to provide natural woodland, open for recreation. (H. L. EN.)

EPSOM AND EWELL, a municipal borough (1937) of Surrey, Eng., on the edge of Banstead downs, 15 mi. W.S.W. of London on the main road to Dorking. Pop. (1961) 71,177. The manor of Epsom (Ebbes Ham in 973) once belonged to Chertsey abbey, but the place first came into notice when mineral springs, from which Epsom salts were named, were discovered in about 1618. As a spa Epsom reached its zenith a century later. Horse

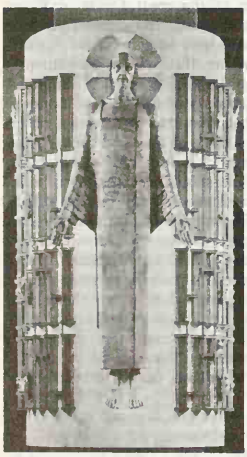
races appear to have been established there as early as James I's residence at Nonsuch, but they did not assume a permanent character until 1730. The spring races (which include the Metropolitan and the City and Suburban) are held toward the close of April while the great Epsom meeting, at which the principal races are the Derby (1780) and the Oaks (1779); named after the 12th earl of Derby and his seat, the Oaks, in the neighbourhood), takes place during the last week in May or the first week in June. There are numerous training stables and several 18th-century houses including the Durdans, house of the earl of Roseberry. The parish church of St. Martin of Tours is a 19th-century restoration. The Epsom and Ewell School of Arts and Crafts was opened in 1896. Close to the town are the extensive buildings of Epsom college, founded in 1855. Epsom downs (387.5 ac.) and Epsom common (435.5 ac.) have been common land from time immemorial. Nonsuch park (300 ac.) is a part of the London "green belt." The borough maintains recreation grounds and owns certain of the stables. Fruit and other soft drinks are manufactured. Epsom market was granted in the reign of Charles II and is held on Saturdays; the July fair dates to the same period. At Ewell, which has also springs of pure water and has been inhabited since about 300 B.C., there is a large county technical college.

EPSOM SALTS, small, colourless, needle-shaped crystals, white when powdered, often used as a cathartic, promoting bowel evacuation. Chemically the salt is heptahydrated magnesium sulfate, $MgSO_4.7H_2O$, the *magnesii sulphas* of pharmacy. It occurs dissolved in sea water and in most mineral waters, especially in those at Epsom, Eng. (from which place it takes its name) and at Seidlitz and Saidschitz, Czech. It also occurs in nature in fibrous excrescences, constituting the mineral epsomite or hairsalt; as compact masses (reichardite); or in association with other minerals such as gypsum and limestone.

In medicine pure Epsom salts is sometimes employed orally as a cathartic, especially valuable in febrile diseases and in the obstinate constipation of painters' colic; and as an antidote for poisoning by barium, barbiturates, lead and other agents. It possesses the advantage of exercising but little irritant effect upon the bowels. Its nauseous bitter taste may to some extent be concealed by fruit juices. Dosage must be carefully prescribed since overdoses may be toxic and even fatal. Epsom salts is used topically, in a saturated solution in water, for various inflammations and, in a dilute solution, for treatment of burns.

Other uses for this salt are in the manufacture of leather, textiles, paper, fire-proofing material, mineral waters, explosives, fertilizers, cosmetics, etc.

See also MAGNESIUM: *Magnesium Compounds.*



STANLEY TRAVERS AIBP, ARPS

"CHRIST IN MAJESTY" BY SIR JACOB EPSTEIN, CAST IN UNPOLISHED ALUMINUM: 1957. IN LLANDAFF CATHEDRAL NEAR CARDIFF, WALES

EPSTEIN, SIR JACOB (1880–1959), one of the most influential sculptors of his time, was born of Russian-Polish parents in New York city on Nov. 10, 1880. His artistic training at the Art Students league in New York city and the École des Beaux-Arts in Paris encouraged the development of his independence in style and technique. He was influenced by the advanced art movements of his time and early in his career explored the possibilities of abstraction in sculpture as in his group "Two Doves" (1913) and the marble "Venus" (1914). By this time he had settled in England where he was to live for most of his life.

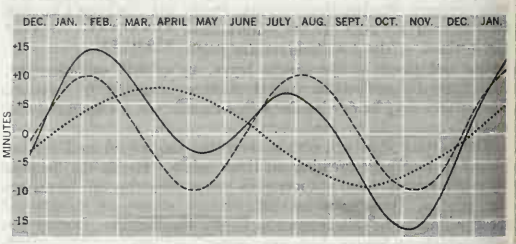
He was among the first sculptors in England to work in the idiom of the leading French artists and in this way he introduced new forms into English art. An artist of keen perception and of high artistic integrity he never permitted himself to be swayed by the criticism or controversy his work sometimes roused although he was always sensitive to public reaction. In most of his work he was able to express a deep feeling in subject matter but at the same time he never concealed the basic quality of the material in which he worked, whether this was stone, clay or metal. In this respect his work recalls that of Auguste Rodin. However, instead of the physical energy seen in Rodin, Epstein revealed a nervous tension and a penetrating emotional sympathy. In his religious figures this often aroused protests as in his life-sized bronze figure of Christ and his monumental stone "Ecce Homo." This same penetration, however, resulted in his success in portraiture.

Among his more important portraits are those of "R. B. Cunninghame Graham," "Kathleen," "Albert Einstein," "George Bernard Shaw," and "Somerset Maugham." It is significant that most of his portraits were in bronze where the plasticity of his modeling material was retained by the metal. This impressionistic surface treatment combined with a cautious emphasis on distinctive features produced a marked and refreshing vitality in his work.

Among his larger commissioned works in London were the sculpture for the British Medical association building in the Strand, for which he made 18 large figures (1908), the Hudson memorial in Hyde Park (1925), two groups "Day" and "Night" on the London Transport building (1929), "Madonna and Child," at the convent of The Holy Child Jesus, Cavendish square (1951), "Jan Christiaan Smuts," in Parliament square (1956), and the bust of William Blake in Westminster abbey (1957). In 1957 he made the figure of "Christ in Majesty," of unpolished aluminum, for Llandaff cathedral, near Cardiff. A retrospective exhibition of his work was held in the Tate gallery, London, in 1952. In 1931 he published together with Arnold Haskell *The Sculptor Speaks,* and in 1940 appeared his autobiography, *Let There Be Sculpture.* He was knighted in 1954. Epstein died in London on Aug. 19, 1959.

BIBLIOGRAPHY.—Bernard van Dieren, *Epstein* (1920); Hubert Wellington, *Jacob Epstein* (1925); L. B. Powell, *Jacob Epstein* (1932); *Epstein,* photographs by G. Ireland (1958). (I. S. McN.)

EQUATION OF TIME, the difference between apparent time (*e.g.*, the time given by a sundial) and mean time (given by a standard clock). The earth rotates on its axis once every 23



——— EQUATION OF TIME
..... EQUATION DUE TO VARIABLE MOTION OF SUN
– – – – – EQUATION DUE TO OBLIQUITY OF ECLIPTIC

FROM H. N. RUSSELL, R. S. DUGAN, V. O. STEWART, "ASTRONOMY"; REPRODUCED BY PERMISSION OF GINN & CO., 1945

EQUATION OF TIME, SHOWING MEAN TIME MINUS APPARENT TIME (*see* TEXT)

hours, 56 minutes; the interval from noon to noon is 24 hours, 0 minutes, the extra 4 minutes being necessary to "catch up" with the sun, which has moved on in the sky since it makes a complete circuit in a year. This variation (strictly 3 minutes, 55.91 seconds) is an average allowance; the sun moves sometimes faster, sometimes slower, reflecting the varying motion of the earth in its elliptic orbit. The sun's motion is fastest at the time of perihelion (Jan. 3); accordingly at that time of the year the clock allowance is insufficient, and apparent noon by the actual sun becomes later and later according to the uniform clock. It is for this reason that after the shortest day the afternoons lengthen faster than the

mornings. Another cause of difference between the sun and the clock is the inclination of the ecliptic (the sun's apparent path in the sky) to the celestial equator. The two causes combined give the values indicated by the solid line in the accompanying figure. *See* ASTRONOMY: *Modern Descriptive Astronomy.*

EQUATIONS, THEORY OF. The theory of equations is concerned with determining the solvability of algebraic equations, and with the development of techniques for the solution of equations which are theoretically solvable.

An equality that holds for all values of the unknowns is an identity. Thus $3x + 2 = 5$, true only for $x = 1$, is an equation; $x^2 - y^2 = (x + y)(x - y)$, holding for all values of x and y, is an identity. To distinguish identities from equations the symbol $\equiv$ is sometimes used, as in $x^2 - y^2 \equiv (x + y)(x - y)$.

The theory of equations is concerned chiefly with the properties of a single algebraic equation of the type

$$c_0x^n + c_1x^{n-1} + \ldots + c_n = 0, (c_0 \neq 0)$$

in which n is a positive whole number called the *degree* of the equation and the $c_0, c_1, \ldots, c_n$'s are either given numbers or numbers that are not specified but are assumed known, called the coefficients of the equation. Roughly speaking, the theory of equations discusses this problem: the coefficients being specified, find all values of x that make the equation true. Such values are called roots of the equation, and finding them is called solving the equation. More generally, there are often sets of simultaneous equations in several variables $x, y, z, \ldots$ to be solved, but the solution of these may be reduced (by eliminating variables) to solving a number of equations of the above type.

The earliest known equivalents of algebraic equations occur in the Rhind papyrus, evidently compiled from earlier works by the Egyptian Ahmes about 1650 or 1700 B.C. For example, he proposes this problem: "A quantity and its seventh added together become 19. What is the quantity?" The problem, therefore, is to solve the equation $x + \frac{1}{7}x = 19$, as mathematicians would now express it. Lacking a convenient algebraic notation, he proceeded by a cumbersome method later known as that of "false position."

Although the Greeks are sometimes credited with solving equations of the second degree, in general neither the Egyptians nor the Greeks made any progress that is significant from a modern point of view, and neither people rose to the abstract conception of a theory of equations as a fruitful field of mathematical science. The Indians and the Arabs achieved more, but it was not until the Renaissance, when the Italian mathematicians of the 15th and 16th centuries succeeded in solving by radicals the general equations of the third and fourth degree, that any work of lasting interest was done. Similar attempts to solve equations of higher degree led nowhere, however; deeper insight into the nature of the roots of an algebraic equation was required, and this did not come until the work of Joseph Louis Lagrange in 1770–71 and, more decisively, that of two young mathematicians half a century later, Abel and Galois.

Niels Henrik Abel (1802–29) at the age of 22 almost proved that equations of degree greater than 4 could not be solved by radicals. His attempt contained two oversights, however, and the definitive proof of this important fact must be credited to Évariste Galois (1811–32). Killed in a duel at the age of 21, Galois spent the night before his death outlining a comprehensive theory of the roots of equations in a hastily scrawled letter to a friend. This theory, which now goes by the name of Galois theory, is a prime source of our information about algebraic equations and has served as an inspiration to much of the modern work in algebra and number theory.

In general, the theory of equations as presented in this article dates from the early part of the 19th century. The approximate solution methods developed then have remained largely unchanged, and although modern algebra has enormously simplified the theory of Galois, the basic ideas are the same.

Statement of the Problem.—The theory of algebraic equations discusses what is essentially a two-fold problem. Recall that in addition to the ordinary real numbers, there are also the complex numbers, *i.e.*, those of the form $a + bi$, where a and b are real and $i = \sqrt{-1}$; the real numbers are, then, those complex numbers for which $b = 0$ (*see* COMPLEX NUMBERS; NUMBER). The algebraic equation given above can be written as $f(x) = 0$, and its coefficients can be taken to be either real or complex numbers; if all the coefficients are real, the equation is called real. This is the important case, since otherwise, multiplying $f(x)$ by $f(x)$, in which the coefficients have been changed into their complex conjugates $(a - bi)$, gives a new equation $F(x) = 0$ which is real, of twice the degree and such that all the roots of the original equation are also roots of this new equation. The term "root" in this case means a complex number α such that $f(\alpha) = 0$. Even if the equation is real, it may well have complex roots: consider $x^2 + 1 = 0$, for instance.

The central problems of the theory are these:

1. The coefficients being given as definite numbers, find the roots of the equation—exactly, if possible, otherwise determine how to calculate them to as many decimal places as needed.
2. Find general formulas that express the roots in terms of the coefficients. If possible the formulas should require only the operations of addition, multiplication, subtraction, division and extractions of roots, this being called a "solution by radicals." If such formulas are not possible, find the simplest functions of the coefficients that do satisfy the equation.

After the above-mentioned work of Abel and Galois, Charles Hermite in 1858 succeeded in solving the general equation of degree 5 by means of elliptic functions (*q.v.*). Modern work in this direction, originating with Henri Poincaré about 1880, solves the general equation of degree n in terms of Fuchsian functions. For a time, problem (1) was considered exhausted, but new methods have been developed that are especially suited to modern high-speed computers. Problem (2) now involves group theory and algebraic number theory, as well as special functions of a complex variable.

The Fundamental Theorem of Algebra.—A basic result applicable to both (1) and (2) is the so-called fundamental theorem of algebra, which states that every algebraic equation has a root (*see* ALGEBRA). It is proved either from this, or almost in one step from an application of an integral formula of Augustin Louis Cauchy in the theory of functions of a complex variable, that an equation of degree n has precisely n roots. The remarkable feature of this theorem is that, in order to solve completely any algebraic equation, it is unnecessary to go beyond the domain of complex numbers. It is by no means obvious that this should be so. The novice in algebra regards the theorem as a truism, the seasoned mathematician sees in it a species of fortunate miracle, while the sophisticated critic views it with suspicion. The first allegedly satisfactory proof was given in 1799 by Carl Friedrich Gauss, who subsequently added three more.

The first useful consequence of the theorem is the following. If $x_1, x_2, \ldots, x_n$ are the n roots of $f(x) = 0$, where

$$f(x) \equiv c_0x^n + c_1x^{n-1} + \ldots + c_n,$$

then

$$f(x) = c_0(x - x_1)(x - x_2)\ldots(x - x_n)$$

Conversely, if $f(k) = 0$, then $x - k$ is a factor of $f(x)$. More generally, if t is any complex number, $f(t)$ is equal to the remainder obtained on dividing $f(x)$ by $x - t$, and hence $f(t)$ can be calculated by division, a result of importance in the numerical solution of equations. This is called the remainder theorem.

Since $c_0, c_1, \ldots, c_n$ are any complex numbers independent of x, and $c_0 \neq 0$, the equation $f(x) = 0$ may be divided throughout by c_0. It then becomes $x^n + a_1x^{n-1} + \ldots + a_n = 0$, where $a_1, \ldots, a_n$ are complex numbers. This form is precisely as general as the original and is convenient to use. If the roots are $\alpha_1, \alpha_2, \ldots, \alpha_n$, the linear (first degree in x) factors of

$$x^n + a_1x^{n-1} + \ldots + a_n$$

are

$$(x - \alpha_1)(x - \alpha_2)\ldots(x - \alpha_n),$$

and hence, on comparing coefficients of like powers of x, it is seen that $a_1 = (-1)$ times the sum of the roots and $a_n = (-1)^n$ times the product of all the roots. This result frequently is useful in testing for rational roots an equation whose coefficients are rational numbers; by this means all the rational roots may be found.

Approximate Location of Roots.—Consider now problem (1); in particular the calculation of the roots by an approximative method. The present discussion will be confined to the real roots of real equations; this is the significant case, because the determination of the complex roots may be reduced to locating the real roots of an auxiliary real equation. In fact, attention may even be confined to locating just the positive real roots, since the negative real roots of $f(x) = 0$ are the positive roots of $f(-x) = 0$, which is the equation obtained from $f(x) = 0$ by changing the signs of all the terms of odd degree.

The problem divides naturally into two steps. First comes the approximate location of the real roots. This includes both determining their total number and isolating them. (A real root is said to be isolated if two real numbers α and β are known such that the equation has exactly one real root between α and β.) Once the root is known approximately, a method of calculating it to any desired degree of accuracy must be determined.

Some elementary properties of real equations which are helpful in connection with the first step will now be mentioned.

Imaginary roots, if any, occur in conjugate pairs, $a + bi$, $a - bi$ (a and b are real numbers and $i \equiv \sqrt{-1}$). Hence, if the degree n is odd, the equation has at least one real root. Furthermore, $f(x)$ being a real, continuous function of x (*see* FUNCTION), it follows that if $f(r)$ and $f(s)$ have opposite signs, where r and s are real numbers, the curve whose equation is $y = f(x)$ cuts the axis of x an odd number of times between r and s. If among the n roots of $f(x) = 0$ there are precisely h each equal to α, and if $h > 1$, the root α is said to be of multiplicity h. Let $f'(x)$ be the first derivative of $f(x)$. Then the theorem of Michel Rolle states that between two consecutive real roots of $f(x) = 0$ there is an odd number of real roots of $f'(x) = 0$, provided a root of multiplicity h is counted as h roots.

When $f(x) = 0$ is real, it is possible, by measuring where the graph of $y = f(x)$ crosses the x-axis, to ascertain how many real roots lie within given limits; by sufficiently enlarging the scale a particular real root may be located with any desired degree of accuracy. The graphical method is a valuable adjunct to the arithmetical processes, because in all numerical solutions the initial difficulty is in approximately locating the root to be calculated.

More powerful tools must usually be applied, however. Descartes' rule of signs gives an upper estimate for the number of positive roots in terms of the successive signs of the coefficients of the equation. Although easily applied, it gives rather crude information, and frequently none at all. The conclusive isolation of the roots is given by a method discovered in 1829 by C. F. Sturm. For an equation of degree n, n auxiliary polynomials called the Sturm functions must be computed (a rather laborious task); the number of real roots (counted with multiplicities) of $f(x) = 0$ between α and β is then ascertained by inspection from the signs of the Sturm functions evaluated at α and β. Budan's theorem (1807) derives a different set of auxiliary polynomials which are used in the same way; although they are easier to compute, they give less precise information.

Computation of Roots.—When a real root of a real equation $f(x) = 0$ has been isolated by any of the methods suggested, it may be calculated, digit by digit, by any one of several arithmetical processes, the commonest of which are those named after W. G. Horner (1819), known to the Chinese mathematicians of the 13th century, and Sir Isaac Newton's method of about 1675.

Newton's method has the great advantage of being applicable to equations other than algebraic. It can be explained by using his own example, $x^3 - 2x - 5 = 0$. This has only one root between 2 and 3. To find this root, replace x by $2 + h$, where h is necessarily between 0 and 1. The terms in h^2 and h^3 being negligible in comparison with h, in the new equation

$$h^3 + 6h^2 + 10h - 1 = 0$$

$h = 0.1$ is clearly a first approximation to h, and hence the root is roughly 2.1. For the next approximation, replace h by $0.1 + k$ in the h-equation, and obtain

$$k^3 + 6.3k^2 + 11.23k + 0.061 = 0$$

Omit the k^3 and k^2 terms, which are small in comparison with k; then k is approximately -0.0054. From $h = 0.1 + k$ the result $h = 0.0946$ is obtained, giving 2.0946 as an approximate value of the required root. By repeating this process a sufficient number of times, any desired degree of accuracy is attainable.

In Horner's method (the one usually given in elementary texts on algebra, except those most popular on the continent of Europe) a set of equations, one for each digit of the required root, is obtained successively from the given equation. The process is an arithmetical restatement and refinement of the crude graphical method, and amounts to successive shifts of the origin and magnifications of the scale. As already pointed out, it is sufficient to discuss the method for positive real roots only. The essence of the method is the successive diminution of the roots of the given equation by the smaller member in successive pairs of positive real numbers; *i.e.*, a gradual creeping up on the root from behind. Suppose, for example, that a real root of $f(x) = 0$ has been isolated between 200 and 300. If an equation $f_1(x) = 0$ whose roots are those of $f(x) = 0$ each diminished by 200 is constructed, then $f_1(x) = 0$ will have one and only one root between 0 and 100. By any of the proposed methods locate this root; *e.g.*, say it is between 60 and 70. Then $f(x) = 0$ has a root between 260 and 270. Construct an equation $f_2(x) = 0$, whose roots are those of $f_1(x) = 0$ diminished by 60, and repeat the argument. Then $f_2(x) = 0$ has a root between 0 and 10, which may be located as before; *e.g.*, say it is between 5 and 6. Then 5 is the third digit of the required root of $f(x) = 0$. Next, construct $f_3(x) = 0$, whose roots are those of $f_2(x) = 0$ diminished by 5; $f_3(x) = 0$ then has a root between 0 and 1, say 0.2. The required root so far is 265.2, and the process can be continued to any prescribed number of digits.

Many devices for shortening the labour of Horner's method are explained in treatises on equations; in particular the last digit (at least) that is required can, in general, be obtained by simple division. Horner's method is, beyond any question, the most practical yet devised for the numerical solution of equations having numerical coefficients. (E. T. B.; A. P. M.)

Formulas for Solving Equations of Low Degree.—Consider now problem (2), the solution of equations by means of exact formulas for the roots in terms of the coefficients. The three equations of low degree for which such formulas exist are the quadratic (of degree 2), cubic (degree 3) and biquadratic (degree 4). Thus, the quadratic equation

$$ax^2 + bx + c = 0$$

has the solutions

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

and so a single square-root extraction is needed to solve it. In a sense this formula, the famous quadratic formula, is only an approximate solution, since the square root can be only approximately calculated; the significant thing, however, is that the formula is universal: one formula takes care of all possible choices for the coefficients.

Similar formulas for the roots of the cubic equation were first published in the *Ars Magna* (1545) by Geronimo Cardano, who obtained them from Tartaglia ("the stutterer," true name Niccolò Fontana) by questionable means; they will not be given here because they are rather complicated and because Horner's method is in general preferable. In particular, if all three roots are real, the formulas give them in terms of complex numbers, a fact which makes the formulas quite impracticable; alternative formulas give the roots as trigonometric functions of the coefficients. Even more impracticable is the general solution to the quartic equation, found by Lodovici Ferrari at about the same time.

The Galois Theory.—The general theory of the roots of equations alluded to earlier will now be discussed. The Galois theory proves the impossibility of such general formulas as the above whenever the equation is of degree greater than 4.

A rational number (or common fraction) is a real number which is the quotient of two integers, so that, in particular, the sum, product, difference and quotient of rational numbers are also ra-

tional numbers. Suppose for simplicity that the equation $f(x) = 0$ has rational numbers for coefficients, and suppose also that $f(x)$ cannot be factored as a product of polynomials of lower degree with rational coefficients. Let $x_1, x_2, \ldots, x_n$ be the roots of this equation; these roots are complex (or perhaps real) numbers, and all different, since $f(x)$ cannot be factored. The crucial notion now is that of an admissible permutation (or substitution) of these roots. A permutation $(x_1, \ldots, x_n) \rightarrow (x_{i_1}, \ldots, x_{i_n})$ of the roots, *i.e.*, a scrambling of them, is said to be admissible if every algebraic relation holding among the roots continues to be valid after the permutation is made. Only those relations expressible in terms of sums of products of the roots, with rational numbers as coefficients, will be considered here. For example, consider the equation $x^4 - 2 = 0$. Letting x_1 be the real positive fourth root of 2, it is easy to see that the other roots are $x_2 = ix_1$, $x_3 = -x_1$ and $x_4 = -ix_1$. The permutation $(x_1, x_2, x_3, x_4) \rightarrow (x_2, x_1, x_3, x_4)$ is not admissible, because the valid algebraic relation $x_1x_3 + x_2x_4 = 0$ goes by the permutation into $x_2x_3 + x_1x_4 = 0$, which is false. On the other hand, the permutation $(x_1, x_2, x_3, x_4) \rightarrow (x_2, x_1, x_4, x_3)$ is admissible, and it may be easily checked, for instance, that it does preserve the validity of the above given relation, although this check in itself would not be sufficient to prove that the permutation is admissible. In this case, of the 24 possible permutations, exactly 8 are admissible.

The set of admissible permutations is called the Galois group of the equation. Since there are n roots, the total number of all permutations is $n!$; thus the Galois group can be no larger than $n!$ and it may in fact be much smaller. If it is of this maximal size, however (*i.e.*, if every permutation is admissible), then the equation is said to have the symmetric group. Loosely speaking, if an equation is written down at random, the chances are that it will have the symmetric group; the ones with smaller Galois groups are more special. If permutations are multiplied in the usual way by performing one after the other, then the admissible permutations form a group in the technical sense (*see* GROUPS).

This group measures what might be called the algebraic symmetries that the roots possess, much as the group of the six possible rigid motions carrying an equilateral triangle onto itself (three rotations and three reflections) measures its sixfold geometric symmetries.

In general a finite group G is called solvable if it has subgroups $G = G_1, G_2, \ldots, G_r = \{1\}$ such that each G_{i+1} is a normal subgroup of G_i and such that G_i has p_i times as many elements as G_{i+1}, where p_i is a prime number. Such groups are very special kinds of groups. The heart of Galois's theory is the theorem that an algebraic equation is solvable by radicals if and only if its Galois group is a solvable group.

Roughly speaking, the p_i entering above into the structure of G corresponds to the solution of the equation by extraction of p_1th, p_2th, . . . , p_rth roots; the possibility of solving the equation by simple root extractions is the reflection of the possibility of elucidating the structure of its Galois group by finding a sequence of subgroups of the sort indicated. Now, since it is easy to show that the group of all permutations on n symbols (there are, as we have remarked, $n!$ of them) is not solvable when $n \geqq 5$, and since there are equations with rational coefficients which have these groups as Galois groups, it follows that these equations cannot be solved by radicals. Thus there cannot be any general formula for the roots in terms of the coefficients when $n \geqq 5$, because such a formula would have to work regardless of what the coefficients were, and in particular would have to work for the special unsolvable equations as well.

See also references under "Equations, Theory of" in the index.

BIBLIOGRAPHY.—E. Artin, *Galois Theory* (1946); G. Birkhoff and S. MacLane, *A Survey of Modern Algebra* (1941); L. E. Dickson, *New First Course in the Theory of Equations* (1939); J. V. Uspensky, *Theory of Equations* (1948); B. L. van der Waerden, *Modern Algebra* (1948); L. Weisner, *Introduction to the Theory of Equations* (1938).

(A. P. M.)

EQUATOR, in geography, the great circle around the earth, equidistant from the north and south poles, lying in a plane perpendicular to the axis of the earth. This "geographical equator" or "terrestrial equator" thus forms the imaginary reference line on the surface of the earth from which latitude is reckoned (*see* LATITUDE; LONGITUDE).

In astronomy, the "celestial equator" is the name given to the great circle in which the plane of the terrestrial equator intersects the celestial sphere; it is consequently equidistant from the celestial poles. The celestial equator is sometimes referred to as the equinoctial, since when the sun lies in the plane of the equator, day and night are of the same length. *See* EQUINOX.

(R. L. DE.)

EQUATORIA (AL ISTIWĀ'IYAH), the southernmost province of the Republic of the Sudan, lies between about latitude 3° and 7° N. and is bounded east by Ethiopia and Kenya, south by Uganda and the Republic of the Congo and west by the Central African Republic. Pop. (1956 census) 903,503. Area 76,495 sq.mi. Equatoria was called Mongalla until 1936 and from 1936 to 1956 included Bahr el Ghazal province. The name Mongalla derived from the administrative post that was founded when Khedive Ismail, viceroy of Egypt, sent Sir Samuel Baker in 1870, followed by Col. C. G. Gordon in 1874, to observe and curb the activities of the Arab slavers whose fortified posts had hitherto loosely dominated the region.

Equatoria is traversed northward by the Bahr el Jebel or White Nile, which from the Uganda frontier at Nimule runs nearly to Juba (100 mi.) through a wide gorge containing the Fola rapids and dominated by the escarpments of the central African plateau. The southeast border region consists of crystalline blocks of highland, including the magnificent Imatong (Kinyeti, 10,456 ft.) and Didinga mountains, and descends north by abrupt steps and huge valley basins to the Torit plateau level, below 3,000 ft. The Torit plateau, tilting northward again and passing under a swampy plain which receives most of its drainage, resembles and is geologically one with the Western plateau draining to the Bahr el Ghazal river. (*See* BAHR EL GHAZAL, articles on river and province.)

The climate is hot and equable. In the nominally "dry" season (November–March), showers and thunderstorms occur at times and the minimum and maximum temperatures are mostly within the range 68°–100° F. Annual rainfall at Juba averages below 40 in., but in the cloudy southeast border highlands is about 60 in. There, above 6,000 ft., are coniferous forests, and temperate crops are possible. Lower down the broadleaf woodland has clearings suitable for tea and coffee and there is rich evergreen forest in the ravines. Exploitation of the highland region is hindered by transport problems. The Torit plateau, with sparse bush and grass on the divides and green parkland along the rivers, is noted for game, especially giraffes. Other big game of all kinds abounds, particularly in the reserve at Nimule.

The grazing in the east of the province is not rich; unreliable rainfall yields little surplus grain or oil and may mean famine. Cotton, ginned at Torit (pop. 2,353), the district headquarters town, has been successful only on the black foothill soils. Eastern tribes are Bari, Lotuko and Toposa. The west of the province is more productive in the Zande district, where an experiment was begun in 1946 to promote the welfare of the Azande (*q.v.*) tribe, numbering 212,000 and previously diminishing. The project includes bush clearance, resettlement, improvement of crops and livestock, housing and hygiene. At Nzara, near the district headquarters at Yambio (pop. 3,890), mills have been built for processing cotton, sugar and oil. Smaller western centres, Maridi and Yei, lie in areas of shifting cultivation, partly resettled along main roads.

The district of the main Bari-speaking tribes is Juba, astride the Nile. On the left bank, the town of Juba (pop. 10,660) is the southern capital and headquarters of the province and district. It is the head of Sudan navigation on the Nile, 1,090 mi. above Khartoum, and is linked by good roads with the end of Uganda navigation at Nimule. All-season roads with permanent bridges also reach Stanleyville and the railway in the Republic of the Congo. Juba is linked by air services with Khartoum and with Entebbe, Uganda. Most of the Sudanese Arabs in Equatoria, together with about 1,000 officials, teachers, clerks and traders and

about 1,000 foreigners, live in Juba, Yambio, Nzara and Torit. The chief exports are mostly shipped north and include fruit, oilseeds, timber, bamboo, hides and skins, and woven mats, but the volume and value are far below the potentialities of Equatoria. (M. T. P.)

EQUATOR PROVINCE (PROVINCE DE L'EQUATEUR), **DEMOCRATIC REPUBLIC OF THE CONGO,** in the northwestern part of the country, is bounded west and north by the Congo and Ubangi rivers and is traversed by the equator. Pop. (1958 est.) 1,808,264, of whom 6,632 were nonindigenous. Area 155,258 sq.mi. Except in the extreme north, the climate is typically equatorial. The province contains an important part of the central Congo basin, with low lateritic plateaus and flat-bottomed valleys which are swampy or periodically flooded.

The Tertiary and Quaternary sedimentaries of the basin contrast with Precambrian metamorphic or crystalline formations which outcrop in the northern part of the province. The tropical rain forest covers the basin region almost continuously. The Congo there flows sluggishly, strewn with islands and in a valley of great width (8½ mi. near the Mongala confluence). Near the Ubangi confluence is the vast Giri swamp. A little downstream is Lake Tumba (193 sq.mi.) into which the Congo overflows at the flood season.

North of Gemena and in the northeast is a wide savanna country. Except in the north and northwest, most of the province is below 1,500 ft.

European exploration and occupation of the region began seven years after the second African journey of H. M. Stanley (*q.v.*), who descended the Congo in 1877. C. A. Coquilhat visited the Bangala country in 1844, and Edmond Hanssens, A. van Gèle and George Grenfell explored the Ubangi in 1884–89.

The province is unevenly populated, the densest concentrations being around Lake Tumba, the Bwaka country northwest of Gemena, the Bumba region and the upper Tshuapa region. The chief town, port and administrative centre is Coquilhatville (pop. 37,587); other administrative and trading posts are Gemena, Libenge, Banzyville, Bumba, Boende, Lisala and Bikoro. In the south Bantu languages (Mongo, Bangi, Ngombe, Budja, Bangala) are spoken; in the north the languages are Sudanic or Cameroun (Bwaka, Ngbandi). The dialect of the Bangala or Ngala, a small group of riverain people, is partly the basis of Lingala, one of the Congolese common languages. The province is divided into four administrative districts: Equator (headquarters Coquilhatville), Ubangi (Gemena), Mongala (Lisala) and Tshuapa (Boende).

The province is essentially agricultural and no minerals are worked. The chief food crops are maize, rice and cassava. Plantations founded by Europeans furnish more than half of the Congolese production of rubber but the Africans have begun to plant rubber trees (*Hevea brasiliensis*). The province yields about one-quarter of Congolese palm-oil production and contains the largest oil-palm plantations in the republic. Plantations founded by Europeans yield coffee (*Coffea robusta*) and cocoa; the northern regions grow cotton. There are no railways and few all-weather roads within the province and most goods travel by river. Air services connect Coquilhatville, Libenge and other centres with Léopoldville and Stanleyville. (HE. NI.)

EQUIDAE, a family of hoofed mammals including the living and fossil horses and asses, onagers and zebras. The Equidae are placed in the order Perissodactyla (*q.v.*), the "odd-toed" hoofed mammals, to which the tapirs and rhinoceroses and several families of extinct mammals also belong. The closest known ancestors of the horses were the palaeotheres, a group of early and rather unprogressive hoofed mammals that once lived in Europe and Asia. The palaeotheres became extinct about 35,000,000 years ago. Other extinct relatives of the horses were the titanotheres and chalicotheres. The horse family has the most complete fossil record of any group of mammals. The history of Equidae can be followed in considerable detail from almost the very beginning of the line, more than 50,000,000 years ago. All living members of the horse family (horses, asses, onagers, zebras) are placed together in the single genus *Equus*, but at least 19 additional genera of extinct horses are known. The horse and its allies, which passed through all of their significant evolutionary history in North America (although horse fossils are known from other parts of the world), are among the most abundant fossils found in the western part of the United States.

Modern Equidae.—There are six main species of modern equids. These can interbreed (although they do not do so in nature), but such crosses almost always produce sterile offspring, showing that they are indeed distinct species.

Domestic Horse and Its Allies.—The domestic horse, *Equus caballus*, no longer exists in the wild state. (The "wild" horses of the western U.S. are really feral; that is, they are descended from domestic horses that escaped.) The last truly wild representative of this species is Prjevalsky's (or Przhevalski's) horse—sometimes given separate species rank as *E. przewalskii*—a small and rather coarse horse with a short erect mane like a zebra's. Prjevalsky's horse once roamed central Asia in large numbers, but apparently is now extinct in the wild, although specimens are often exhibited in zoos. The tarpan is another type of *Equus caballus*, a small horse with a long flowing mane like that of the domestic horse. The tarpan once flourished from Europe through central Asia and survived in the wilder parts of central Europe through the middle ages. It interbred with domestic horses and therefore never really became extinct. The tarpan was "restored" in the Munich zoo by careful selective breeding of domestic horse stocks of partial tarpan ancestry, and specimens from the Munich herd may now be seen in the larger zoos in the U.S. and in Europe. For the domestic horse, its care and breeds, *see* HORSE.

Onager.—The onager (*Equus hemionus*), a small horse with moderately large ears and a short erect mane, once ranged through the steppes and deserts of Asia from China to Palestine. Its range is now much restricted and the onager is probably approaching extinction. This horse was domesticated in ancient Mesopotamia and was used for about 1,000 years, but was given up in favour of donkeys and domestic horses.

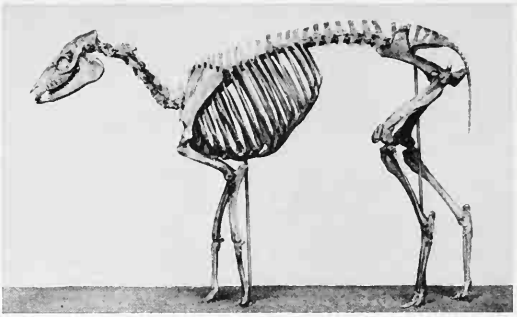
Ass.—The wild ass (*Equus asinus*) is the ancestor of the domestic donkey, and has large ears and a short erect mane. It was once widely distributed in northeastern Africa but is now extinct over most of its former range. (*See* Ass.)

Zebras.—All the remaining living equids are striped and are often lumped together as zebras. This assemblage is a confusing one, and there is some disagreement on the number of species. There appear to be three distinct species: the mountain zebra (*Equus zebra*), Grévy's zebra (*E. grevyi*), and the quagga (*E. quagga*), the last of which is now extinct in nature. Burchell's zebra, apparently closely allied to the quagga, is considered a separate species (*E. burchelli*) by some zoologists. All zebras are African and, as with all other wild horses, their ranges and numbers have been tremendously reduced by modern man. (*See* ZEBRA; QUAGGA.)

Evolution of the Equidae.—The direct ancestor of the horse family is unknown, but unquestionably it was one of the Condylarthra, a group of very primitive mammals, long extinct, that were ancestral to all the latter hoofed mammals. The known history of the Equidae begins in the Eocene epoch, about 50,000,000 years ago, with the eohippus (*q.v.; Hyracotherium*), the well-known "dawn horse" or "four-toed horse" (*see* fig. 1). Several species of eohippus, ranging in size from a terrier dog to a Shetland pony, lived in North America and Europe during Early Eocene times. In some parts of western U.S. fragments of these little horses are among the most abundant fossils, but complete skeletons, even complete skulls, are very rare. The eohippus was a very primitive creature, not yet much different from the condylarths from which it had descended, and so different from modern horses that it was not recognized as a horse when first discovered. The skull was like that of the condylarths, with a short snout and the eye socket not completely enclosed by bone. All 44 teeth characteristic of primitive mammals were still present (in living horses the normal number is 40 in the male and 36 in the female). The cheek teeth were small and low crowned in the eohippus, adapted to crushing leaves and other soft vegetation; they could not have been used to eat grass, which is the main diet of modern horses. The back was arched instead of straight as in

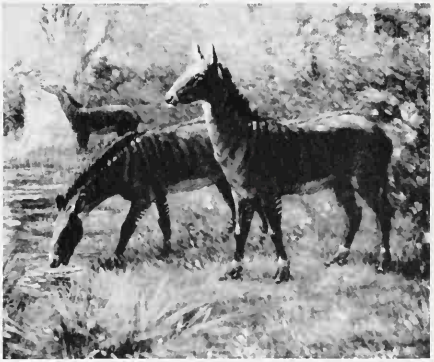
Restoration of a three-toed horse, *Mesohippus*, extant during the Oligocene epoch, about 30,000,000 years ago

Fossil skeleton of *Mesohippus* found in the rock of the Big Bad Lands of South Dakota



Mounted skeleton of *Mesohippus*. This genus represents the stage in evolution during which equids became more like modern horses

Restoration of the earliest one-toed horse, *Pliohippus*, extant during the Pliocene epoch, about 5,000,000 years ago

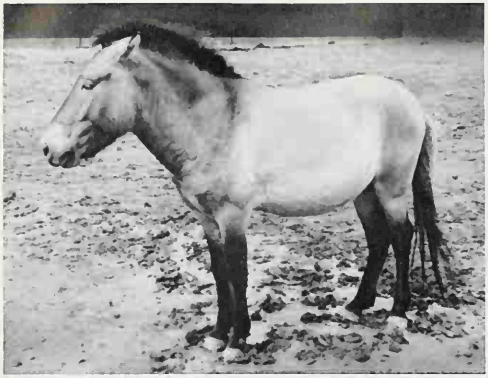


Restoration of an anchithere, *Hypohippus*, extant during the Miocene epoch, about 20,000,000 years ago

REPRESENTATIVE EXTINCT EQUIDS

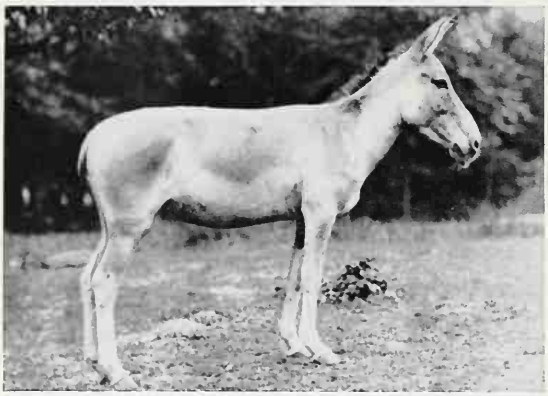
BY COURTESY OF (TOP, CENTRE LEFT AND RIGHT, BOTTOM LEFT) THE CHICAGO NATURAL HISTORY MUSEUM, (BOTTOM RIGHT) THE AMERICAN MUSEUM OF NATURAL HISTORY, DETAIL FROM A PAINTING BY CHARLES R. KNIGHT

EQUIDAE



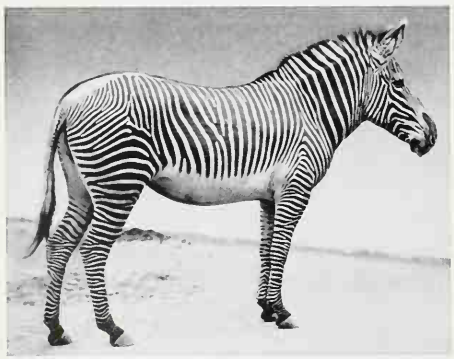
Prjevalsky's (or Przhevalski's) horse (*Equus przewalskii*), the last representative of the wild horse, it is now apparently extinct in a wild state

Tarpan (*Equus caballus*) "restored" in the Munich zoo by careful selective breeding of domestic stocks of partial tarpan ancestry

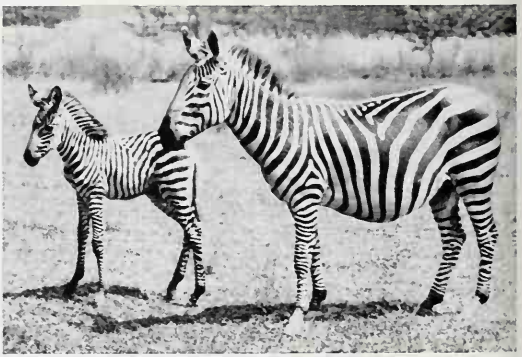


Onager (*Equus hemionus*). Once found widely distributed through the steppes and deserts of Asia from China to Palestine, its range is now greatly restricted

Wild ass (*Equus asinus*). The stock from which the domesticated donkey was derived, it is found in small numbers in northeast Africa



Grévy's zebra (*Equus grevyi*) of Somaliland, northern Kenya and parts of Ethiopia



Quagga (*Equus quagga*), once widely distributed in Africa and represented by several races. The South African races have been exterminated by man

REPRESENTATIVE EXTANT EQUIDS

BY COURTESY OF (TOP LEFT AND RIGHT) D. DWIGHT DAVIS, (OTHERS) THE NEW YORK ZOOLOGICAL SOCIETY

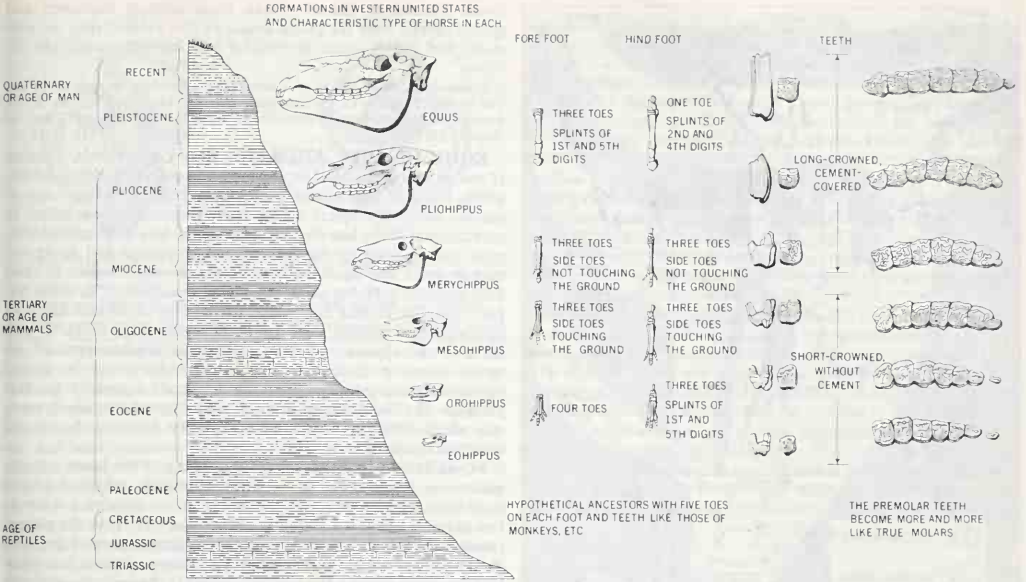


FIG. 1.—EVOLUTION OF THE HORSE FROM THE EOCENE EPOCH TO THE PRESENT

modern horses. The foot bones had already begun to elongate, although they were still much shorter than in later horses. There were four toes on the front foot and three on the hind, each ending in a tiny hoof. The brain was at a low level of development, about like that of a modern opossum. The eohippus was evidently a forest animal, and probably looked and behaved much like the small chevrotain *(q.v.)* or mouse deer that lives in tropical forests, although it was far less intelligent.

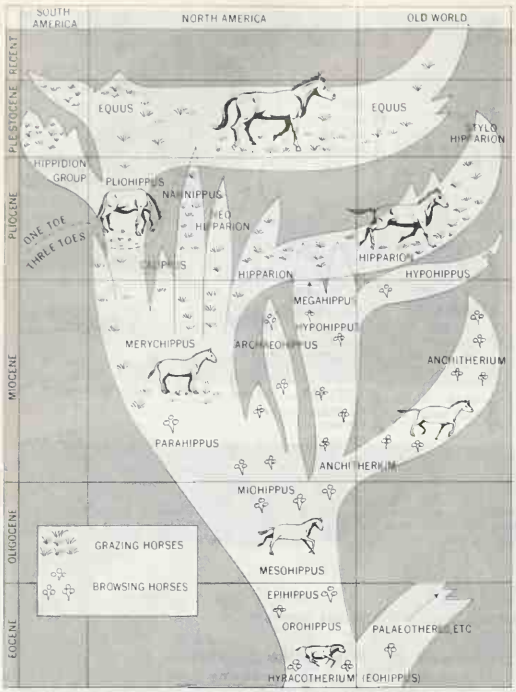
During the remainder of the Eocene the most progressive changes in equids were in the teeth. In modern horses the premolar teeth are very similar to the molars, and this change (called "molarization of the premolars") had already begun in such Late Eocene horses as *Orohippus* and *Epihippus*, although all the cheek teeth were still low crowned as in the eohippus. In all other respects *Orohippus* and *Epihippus* were very similar to *Hyracotherium*. During the next epoch, the Oligocene, equids became more like modern horses in several ways. There was a tendency toward increased size, and the back became straighter and the legs longer. One toe was lost on the forefoot, leaving three toes on both front and hind feet. Molarization of the premolars was carried to completion, although the cheek teeth were still low crowned for crushing soft foods. This stage in equid evolution is represented by *Mesohippus;* remains of *Mesohippus*, one of the best known of all fossil equids, are abundant in the Big Bad Lands of South Dakota. In the Middle and Late Oligocene, in the genus *Miohippus*, the evolution of the teeth and feet was carried a step farther in the direction of modern horses.

The history of the Equidae is often described as if there had been a straight-line evolution leading directly from the eohippus up to modern horses, but this is not true. In North America equid evolution was in fact a straight-line lineage for the first 25,000,000 years, in the line *Hyracotherium—Orohippus—Epihippus—Mesohippus—Miohippus*, each successive group becoming more like modern horses than its ancestors were. But after *Miohippus* the history of the horse family becomes much more complicated (*see* fig. 2). The descendants of *Miohippus* divided into two great subfamilies: Anchitheriinae, the anchitheres, and the Equinae, or "true" horses. *Mesohippus* and *Miohippus* included the ancestors of both the anchitheres and the Equinae, but for convenience they are formally classed among the anchitheres.

The anchitheres remained browsing forest-inhabiting animals throughout their history. Their teeth remained relatively low crowned as in the earlier horses, and they retained three toes on each foot. The best known anchithere is *Anchitherium*, a form that originated in North America but then spread to Europe, from where it is best known. More typically American anchitheres were *Hypohippus* and *Megahippus*. The anchithere line became extinct in the Early Pliocene, about 10,000,000 years ago.

The Equinae were more progressive than the anchitheres, and eventually more successful. Grasses became common on the open plains of North America during the Miocene, about 20,000,000 years ago, and this provided a rich and abundant food supply. Grass is a harsh substance, however, and since it is always contaminated with abrasive dust and sand, it can be eaten only by animals with specially modified teeth and digestive organs. The simple low-crowned teeth of the anchitheres could not have withstood the constant abrasion of grass and its gritty contaminants; they would have worn out while the animal was still young. Grass-eating horses have tall prism-shaped teeth (the hypsodont condition) that gradually wear down with use, but are tall enough to last a lifetime despite continuous wear. These teeth have deep folded enamel ridges surrounded by cement. As the tooth wears down, the folded ridges continuously provide many sharp blades that can cut grass into small pieces. Such teeth began to appear among the Equinae during the Miocene, first in *Parahippus*, and later (and better developed) in *Merychippus*. In the later species of *Merychippus* the teeth are practically the same as in all later horses. Other less dramatic changes in the direction of modern horses were taking place in the skull and limbs. Even the most progressive species of *Merychippus* still retained three toes on each foot, but otherwise these creatures, 15,000,000 years ago, must have looked very much like modern ponies.

By the end of the Miocene the grazing horses descended from *Merychippus* had split up into at least six lines. Several of these were short-lived and apparently not very successful, but two lines led to large and successful groups. One of these is typified by *Hipparion*, the best-known representative of a large assemblage of three-toed grazing horses with high-crowned teeth. *Hipparion*



FROM "HORSES: THE STORY OF THE HORSE FAMILY IN THE MODERN WORLD AND THROUGH SIXTY MILLION YEARS OF HISTORY" BY G. G. SIMPSON. COPYRIGHT 1951 BY OXFORD UNIVERSITY PRESS, INC., AND REPRINTED BY PERMISSION

FIG. 2.—THE SEVERAL LINEAGES OF THE HORSE FAMILY

itself became extinct in the Late Pliocene, but a closely related genus (*Stylohipparion*) survived in Africa until well into the Pleistocene, only a few thousand years ago. *Hipparion* originated in North America, whence it spread to Asia by way of Alaska and pushed on into Europe and even Africa.

A second line that originated from *Merychippus* is characterized by the disappearance of the side toes, leaving only the single middle toe as in modern horses. This condition appeared in *Pliohippus*, which flourished in North America during Early to Middle Pliocene times, 5,000,000 to 10,000,000 years ago. The earliest representatives of *Pliohippus* still had vestiges of the side toes, although these were functionless little remains that no longer touched the ground. In later representatives the side toes were reduced to splint bones lying beneath the skin of the lower leg, as in modern horses. Such species of *Pliohippus* are almost indistinguishable from *Equus*, the genus of modern horses, and *Equus* in fact arose directly from *Pliohippus* at the end of the Pliocene, about 1,000,000 years ago. *Equus* quickly spread to all continents except Australia, which it could not reach because there was no land route. Its fossil bones occur abundantly in Pleistocene deposits in all parts of the world.

Horses first reached South America in the late Pliocene, soon after the Isthmus of Panamá arose. Some species of *Pliohippus* then crossed into South America and gave rise to a special stock of peculiarly South American horses typified by *Hippidion*. These were one-toed horses, but were not closely related to *Equus*. *Equus* itself reached South America during the Pleistocene and soon replaced the older *Hippidion* stock.

After living through their whole 50,000,000 year history in North America, horses abruptly became completely extinct throughout the new world not more than 10,000 years ago. The reason for this remarkable extinction is an unsolved mystery. When European colonists reintroduced horses into North and South America in historic times, these animals flourished and spread rapidly over the plains areas; so it is evident that the new world had not in some mysterious way become unsuitable for horses.

BIBLIOGRAPHY.—F. Harper, *Extinct and Vanishing Mammals of the Old World* (1945); G. G. Simpson, *Horses; the Story of the Horse Family in the Modern World and Through Sixty Million Years of History* (1951). (D. D. D.)

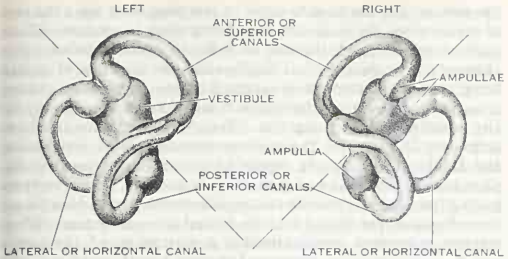
EQUILIBRIUM, ANIMAL. Nearly all animals, whether at rest or in motion, tend to assume some definite bodily position, which is called the normal position, in respect to the force of gravity and other forces acting upon them. An orientation to gravitational force has obvious advantages to a land animal that must run or creep on the surface, and is perhaps just as important to aerial and aquatic creatures whose very freedom in their fluid medium calls for some stabilizing conditions. In these adjustments the animal may make use of any of several kinds of information, including cues from the eyes, from contacts with surfaces and objects and from mechanical displacements of internal organs or appendages; but in particular it may be guided by cues from a sensory apparatus developed especially for this purpose—the sense of equilibrium. The structures which serve this sense differ in form and complexity in invertebrate and vertebrate animals.

Equilibrium in Invertebrates.—Many of the lower animals possess an organ called a statocyst, a roughly spherical vesicle lined with sensory cells and containing in the interior a more or less movable mass, the statolith. When the body is in the normal position the statolith rests against the lowermost part of the vesicle, in contact with a particular group of sensory cells; but when by some circumstance the bodily position is changed, the statolith is displaced to some new region of the sensitive wall and stimulates other cells. This stimulation sets up impulses in the nerve fibres belonging to these cells, and the effects are carried through the nervous system to the muscles of the legs, wings or fins, whose actions then tend to restore the animal to the normal position. Equilibrium thus is automatically maintained. When the organ is richly supplied with sensory cells, the adjustments are extraordinarily precise; thus experiments on the slug show that this animal can detect a tilt of its body of as little as one-third of a degree.

Statocysts of the type described are found in the ctenophores, or sea walnuts; in mollusks, such as bivalves, snails and squids; and in crustaceans. Sometimes only a single organ is present, but more often there are two, arranged symmetrically on the two sides of the head. In some animals, such as the strong-swimming crustacean *Mysis*, each of the paired statocysts works independently; while in others, like the prawn, the statocysts are set at an angle to one another so as to be stimulated differently by a change in position, and hence are in some degree complementary in function.

The simplest method of proving that the statocysts are organs of equilibrium is to destroy or remove them. Y. Delage was the first to do this. He destroyed the organs with a needle, and afterward observed serious disturbances of orientation and movement: the animal lay on its side or on its back, turned somersaults, ran in circles and the like. Often the disturbances are not found after simple elimination of the statocysts, but appear after the animal is also blinded, though blinding alone has no such effect.

A further method of investigating the action of the statocysts is available in the crustaceans, in which the vesicle is open at one place to the exterior, and the statoliths, instead of being secreted by the animal, are grains of sand and the like obtained from the outside. When the animal molts and casts off its chitinous covering, it loses the lining of the statocyst and its contents as well. The statoliths then must be replaced. V. Hensen observed this curious circumstance in the prawn, and reported that the animal used its pincers to pick up sand grains and carry them to the statocysts. A. Kreidl took advantage of this practice in an ingenious experiment. He placed the newly molted animals in filtered sea water and supplied them with fine iron particles which they could do no other than accept as statoliths. In consequence, the balancing organs became sensitive to magnetic in-



FROM T. G. ANDREWS, "METHODS OF PSYCHOLOGY" (1948); REPRODUCED BY PERMISSION OF JOHN WILEY & SONS, INC.

FIG. 1.—SEMICIRCULAR CANALS AS SEEN FROM BEHIND AND ABOVE

The three canals on one side lie in planes at approximately right angles to one another. The two lateral, the right anterior and left posterior, and the left anterior and right posterior canals lie in approximately the same planes

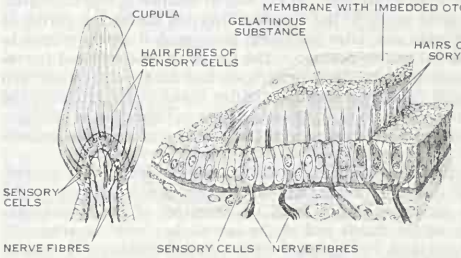
fluences. When, with the animal in a normal position, a strong electromagnet was held over it, the "iron statoliths" were attracted upward, and the organ was stimulated just as if the animal had been turned on its back. Accordingly it turned over and remained with ventral side uppermost so long as the magnet was kept in position, for then the statoliths were pulled against the part of the vesicle wall that was usual for the normal posture. In general, when the magnet was given various positions, the orientation assumed by the animal corresponded to the resultant of the two forces, magnetic and gravitational, acting on the statocyst organs.

Yet another form of experiment in which a new force was made to act on the statoliths was carried out by G. Fraenkel on a turbellarian flatworm that has a single statocyst located in the "brain." Ordinarily when these worms are placed on a vertical surface they creep downward, with almost perfect vertical orientation of their bodies. Fraenkel placed them on a glass plate that was held vertically in a centrifuge and rapidly rotated. The animals on the axis of rotation were unaffected, but all others assumed positions inclined to the vertical and crept in sweeping arcs downward and outward. They were influenced by the centrifugal force produced by the rotation, and were affected the more the farther they were from the centre. As in the preceding experiment, the orientations were determined by the combined actions of the two forces exerted on the statocyst organs.

The results of this last experiment have a more general significance. As here was demonstrated experimentally, so in many situations of constant occurrence in ordinary life animals are exposed to centrifugal forces, and their orientations are affected accordingly. If an animal while in motion changes its direction suddenly, or if it speeds up or slows down, the organs of equilibrium are stimulated. A sudden increase of speed causes the statolith to lag backward, and the animal responds by leaning forward. A decrease of speed arouses the contrary adjustment. When a corner is rounded suddenly, the statolith is thrown to the outside, and the animal is caused to lean inward. All these postural changes are such as to orient the body in the most favourable way under the circumstances, and it is a fortunate fact that these dynamic adjustments are altogether consistent with the simpler, static functions of the statocyst organs.

In some forms, like the jellyfish *Aurelia*, the statocysts have a different structure, and are called tentaculocysts. They are clublike projections weighted at the end with crystals, and are eight in number, arranged around the margin of the umbrella. It is believed that they assist in the maintenance of an upright position during swimming.

Equilibrium in Vertebrates.—In the vertebrates the organ of equilibrium (often referred to as the vestibular end organ) attains an extensive development. It consists of a number of sensory endings within the membranous labyrinth, an enclosure of complicated form usually surrounded by bone or cartilage and located within the inner ear (*see* fig. 1 and 2). In general, there are two sacs, the utricle and saccule, and three semicircular canals, all filled with a fluid, the endolymph. The utricle and saccule contain one type of ending, the macula, which bears some resemblance to the statocyst of invertebrates; while the semicircular canals have a different type, the ampullar crista (*see* fig. 3). The utricle,



FROM T. G. ANDREWS, "METHODS OF PSYCHOLOGY" (1948); REPRODUCED BY PERMISSION OF JOHN WILEY & SONS, INC.

FIG. 3.—(A) CROSS SECTION OF AMPULLAR CRISTA; (B) MACULA IN THREE-DIMENSIONAL CROSS SECTION

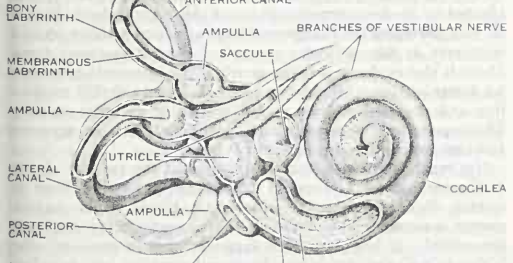
saccule and semicircular canals are referred to as the vestibular division of the inner ear. The remaining part, the cochlea, is the end organ for hearing. (*See* EAR, ANATOMY OF.)

The macular ending is a broad, oval patch on the inner surface of the membranous cavity, and is made up of epithelial cells liberally provided with short cilia which extend toward the interior. Over the cells and firmly embedding them is a dense gelatinous blanket, the statolithic membrane, which is weighted with numerous calcareous particles, the statoliths. (In certain fishes the statoliths are fused into a single body.)

The stimulation of the macula is a little different from that of the invertebrate statocyst. Its statolithic mass is not free to move to different regions of its cavity, but is fixed in one place. As the head takes different positions it hangs differently, exerting varying tensions on the cells at its base and so stimulating them. In effect, then, its action is similar to that within the statocyst.

The second type of ending, the crista, is found in an enlargement (ampulla) at one end of each semicircular canal. It differs from the macula in several respects. The epithelial cells lie on the surface of a ridge and have long hairs extending into a gelatinous mound called a cupula. There are no statolithic granules. The cupula extends all the way across the ampullar cavity to the opposite wall, but is not attached to this wall.

The ampullar endings are stimulated by rotations of the head or, more strictly, by changes in the rate of rotation (angular acceleration). When a rotation begins in the plane of a given canal, the endolymph within it tends to lag behind, and only after



FROM T. G. ANDREWS, "METHODS OF PSYCHOLOGY" (1948); REPRODUCED BY PERMISSION OF JOHN WILEY & SONS, INC.

FIG. 2.—LATERAL VIEW OF RIGHT LABYRINTH OF THE HUMAN EAR

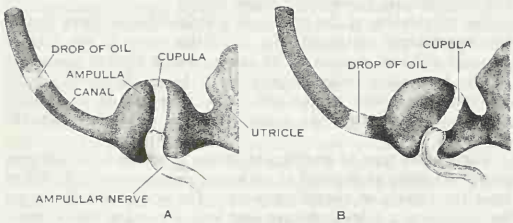
The bony labyrinth is partly cut away to reveal the membranous labyrinth within. The utricle, saccule and semicircular canals make up the vestibular part of the inner ear, which serves as the end organ for the sense of equilibrium. The cochlea is the end organ for hearing

some time does it follow the motion fully. Meanwhile there is relative movement between the fluid and the surrounding tissues, including the cupula. The fluid exerts pressure on the cupula in its path, pushes it over and bends the hairs embedded in it. Thereby the sensitive cells at the base of the organ are excited.

The details of this action were brought to light in an experiment by W. Steinhausen in 1931. He worked mostly on fishes such as the pike, whose canals are large and relatively accessible. After exposing a canal he drilled a minute hole in it and injected a drop of India ink. As the ink penetrated the ampulla, the form of its cavity became visible as a dense shadow against a light directed from behind. When the animal was kept motionless the ink diffused only up to the cupula and there stopped short, proving that the cupula when in its normal position forms a tight barrier to fluid movement. When the fish's head was rotated, however, the cupula was seen to bend over from time to time and allow a globule of fluid to pass.

G. Dohlman, using similar techniques, confirmed and extended the findings of Steinhausen. With the advantage of more modern instrumentation he was able to observe and time the movements of the cupula of a semicircular canal during rotation about an axis perpendicular to the plane of the canal. Starting from a resting position (fig. 4[A]) the cupula is displaced by movement of the endolymph during angular acceleration (fig. 4[B]). If, after maximal displacement of the cupula, rotation is maintained at constant speed (no acceleration), the cupula returns to its resting position within 20 to 30 seconds. If the rate of rotation is now decreased, the cupula bends in the opposite direction during the period of deceleration and, after movement is stopped, it slowly returns to its original resting position. The length of time required for return to resting position depends upon how far the cupula has been displaced during deceleration. After maximal displacement the time required is 20 to 30 seconds. Rate and direction of acceleration or deceleration determine the amount of cupular displacement.

Another kind of evidence bearing upon the function of the vestibular system has come from experiments in which activity in the nerve pathways leading from the receptors of the semicircular canals, of the utricle and of the saccule has been recorded. In experiments on cartilaginous fishes (Chondrichthyes), O. Löwenstein and his colleagues recorded by electrophysiological techniques the nerve impulses which are set off when the temporal bone containing the vestibular receptors is subjected to movement. Löwenstein was able to isolate a single nerve fibre from the ampulla of one semicircular canal of the ray. When the canal was at rest he found that the nerve fibre fired "spontaneously," that is, nerve impulses occurred at a rate of about 6 to 12 per second in the absence of any intentional movement. During rotation at an accelerating speed (angular acceleration) in the plane of the canal the rate of firing of nerve impulses increased when rotation was in one direction and decreased when rotation was in the opposite direction. If, after angular acceleration, rotation was maintained at a constant speed, the rate of firing of nerve impulses returned to the resting rate, the time to come to rest depending upon the rate of acceleration, the maximum time being 20 to 30 seconds. When rotation was stopped, the rate of firing of nerve impulses again changed during the period of deceleration, the direction of change being opposite to that which occurred during acceleration. After cessation of movement, the rate of firing of nerve impulses gradually returned to the resting rate. The results of Löwenstein's experiments fit well with those of the experiments of Steinhausen and Dohlman. When the cupula of a semicircular canal is displaced in one direction during angular acceleration, the nerve fibres leading from the receptors in the ampulla are excited and send more nerve impulses per second into the central nervous system. When rotatory movement is maintained at a constant speed, the cupula returns to its resting position, and the number of impulses in the nerve returns to the base rate, the same rate as that of the canal at rest. During deceleration, the cupula is deflected in a direction opposite to that of acceleration, and now there is a decrease in rate of firing of nerve impulses. After cessation of movement the cupula slowly returns to its normal resting position, and the rate of nerve impulses also returns to that of the resting state.



ADAPTED FROM GOSTA DOHLMAN BY COURTESY OF THE HONORARY EDITORS OF THE PROCEEDINGS OF THE ROYAL SOCIETY OF MEDICINE, 1935

FIG. 4.—MOVEMENT OF CUPULA AND ENDOLYMPH DURING ANGULAR ACCELERATION

(A) Position of cupula with the canal at rest. A drop of oil has been injected into the canal in order to obtain a record of fluid movement. (B) Position of cupula during angular acceleration. The fluid in the canal has moved toward the ampulla, as indicated by the change in position of the drop of oil, and the cupula has been displaced toward the utricle by the fluid movement

Since animals, including man, which have highly specialized organs of equilibrium (the vestibular system) are constructed in a bilaterally symmetrical fashion, it is of interest to note that the end organs on the two sides of the body behave in a reciprocal fashion. For example, the horizontal semicircular canals of the two sides are stimulated by rotation about an axis perpendicular to their plane (*see* fig. 2). When the rotation (angular acceleration) is toward the right side (clockwise direction), there is an increase in rate of firing of nerve impulses in the nerve fibres from the ampulla of that side, but a decrease in rate of firing of impulses in the nerve fibres from the left horizontal ampulla. When rotation is toward the left side (counterclockwise), there is an increased rate of firing in nerve fibres on the left and a decrease in fibres on the right. Similarly, for the other semicircular canals that work in pairs, when rotation occurs about other axes of the body, there will be an increased number of nerve impulses set off on one side and a decrease on the other.

Löwenstein also recorded the impulses in nerve fibres from the receptors in the maculae of the utricle and saccule. In terms of their responses to movements of the head, nerve fibres of several types may be described. One type, which comes from what Löwenstein calls positional receptors, has a resting rate of discharge of 10 to 20 impulses per second. When the head is tilted in one direction the rate increases; when it is tilted in the opposite direction the rate decreases. If the head is maintained in a tilted position, the rate of discharge remains different from that of the normal rate for an indefinite period of time. Nerve fibres of a second type are connected with "out-of-position" receptors. These show an increased rate of discharge to any change in position of the head and a quick return to the normal resting rate after movement has ceased. Still a third type of nerve fibre from the utricular macula tends to fire at a steady rate at all times and is little affected by head movements; fibres of this type probably have to do with providing the background level of neural activity which is important in the vestibular control of general muscular tonus. There is also evidence that the utricle is the principal sense organ for linear motion. Stimulation takes place only during acceleration and deceleration, not during motion at constant velocity. The cupulae of the semicircular canals may have an accessory function as receptors for linear movement.

Not much is known about the function of the saccule. In the cartilaginous fishes, Löwenstein showed that nerve fibres from the saccule discharge when the anterior two-thirds of the saccular macula is stimulated by vibrations at frequencies below 120 cycles per second. It has often been suggested that in man and other higher animals the saccule may be an end organ not only for vibration but also for the hearing of sounds of low frequency. There is little evidence to support this notion, but it remains an interesting hypothesis.

Other experimenters have used methods similar to those of Löwenstein to study responses of vestibular end organs or of the neural pathways of the vestibular system. The results found in experi-

ments on the frog and the cat, in general, confirm those reported by Löwenstein for the cartilaginous fishes.

A third line of investigations of vestibular function dates back to 1875 when Ernst Mach described the reactions of human subjects to controlled stimulation of the vestibular end organs. In the many studies which followed those of Mach, a number of different responses of man were observed and measured during linear and angular motion and during maintained changes of position. The responses most often examined were (1) subjective experiences of movement or of position; (2) eye movements; and (3) changes in appearance of objects seen in space. There is surprisingly good agreement between the results of these studies done with human subjects and the experiments described above in which lower animals served as subjects.

The threshold rate of acceleration for subjective experience due to rotation is approximately 0.5° per second per second. Acceleration rates somewhat higher than this are typically used in experiments examining human responses of the vestibular system. In the absence of visual cues, rotation at constant velocity ceases to be perceived by the human subject; he feels himself to be at rest. The time for cessation of subjective experience depends upon the previous rate of angular acceleration, the maximum time being 20 to 30 seconds. When rotation is stopped (angular deceleration), the subject has the experience of being rotated in an opposite direction. This "aftersensation" lasts for a time which is determined by the rate of deceleration and the speed and duration of rotation at constant velocity prior to deceleration. Again, the maximum time of the aftersensation is about 30 seconds.

Eye movements, called nystagmus, may be induced by rotation (*see* EYE, HUMAN). These movements, which consist of a slow drift of the eyes in one direction and a fast shift in the opposite, have been studied in many experiments on human subjects. During rotation, visual cues also affect eye movements, so it is necessary to eliminate visual cues by blindfolding the subject or by rotating the whole room in which he is placed. When the situation is thus controlled, it is found that the time relations of the eye movements during angular acceleration, rotation at constant velocity and deceleration are similar to those for subjective experience. Both eye movements and subjective experiences can be explained adequately in terms of the movements of the cupula described by Dohlman and the consequent flow of nerve impulses as observed and measured by Löwenstein.

From a practical standpoint, there has perhaps been more reason for man to be interested in his reactions to linear motion than in those due to angular motion. The nausea and sickness which he suffers on a ship, in a plane or in an automobile are due primarily to stimulation of the receptors which are affected by linear acceleration (*see* MOTION SICKNESS). In the laboratory, experiments using such devices as a swing, a hydraulic elevator and a giant centrifuge have shown that movement will affect the macular endings in the utricle primarily, but may also stimulate the semicircular canals to some extent. In addition to producing motion sickness, certain kinds of motion, such as that on the centrifuge or in maneuvers of an airplane, may cause objects in space to be falsely perceived. An object in space may be mistakenly seen as moving or as having shifted its position.

Space Flight.—Initial laboratory observations, supported by manned orbital flights in the early 1960s, gave experimental evidence of what was already hypothesized, that man's vestibular receptors will be subjected to unusual stimulation under such conditions. There is extremely high linear acceleration during launching, and high deceleration during return from orbital space into the earth's atmosphere. After escape from the earth's gravitational field, a condition of zero gravity, or weightlessness, exists. It has been shown that it is possible for man to maintain orientation and to carry on tasks requiring reasonably complex muscular movements under the conditions of zero gravity. Severe disturbances of equilibrium are more likely produced by excessive acceleration and by unusual combinations of rotatory and linear motion than by weightlessness. All these effects will be intensified when man becomes a passenger in vehicles traveling in outer space (*see* also SPACE EXPLORATION).

Other Senses.—The control of body posture and locomotion is dependent not only upon the special organs of equilibrium but upon other senses as well. In the absence of vestibular end organs, man can maintain an upright posture and can move about with reasonable co-ordination if he is able to see objects surrounding him and if he is receiving normal sensory cues from touch receptors and from the kinesthetic receptors which are stimulated by movements of muscles and joints. (*See* SENSATION.) If blindfolded or if in total darkness, he will have difficulty remaining upright.

Lower animals which move on four feet rather than two do better in the absence of both vestibular and visual cues; man likewise would find it advantageous under these circumstances to adopt the posture of the quadruped.

If, in an experimental animal, the sensory systems for touch and kinesthesis were destroyed in addition to the elimination of the visual and vestibular end organs, the animal would collapse and become a limp, disorganized mass.

In the intact animal, the activities of the separate sensory systems that are important in equilibrium are co-ordinated and integrated in the central nervous system, principally in brain stem centres and in the cerebellum. The maintenance of equilibrium is done on a more or less automatic or reflex basis; there is little conscious awareness or intentional guidance of the muscular responses which occur.

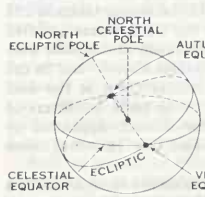
See BRAIN; *Hindbrain and Roof Plate.*

BIBLIOGRAPHY.—S. S. Maxwell, *Labyrinth and Equilibrium* (1923); M. Camis, *The Physiology of the Vestibular Apparatus,* trans. by R. Creed (1930); G. Rademaker, *Réactions labyrinthiques et équilibre* (1935); W. D. Neff, "Studying Proprioception," in T. G. Andrews (ed.), *Methods of Psychology,* ch. 11, pp. 289–318 (1948); G. R. Wendt, "Vestibular Functions," ch. 31 in S. S. Stevens (ed.), *Handbook of Experimental Psychology* (1951); O. Löwenstein, "Labyrinth and Equilibrium," in *Physiological Mechanisms in Animal Behaviour,* "Symposia Series" of the Society for Experimental Biology (1950), "The Equilibrium Function of the Vertebrate Labyrinth," *Biol. Rev.,* vol. ii, 113 (1936), "Comparative Physiology of the Otolith Organs," *Brit. M. Bull.,* 12:110–114 (1956) and "Peripheral Mechanisms of Equilibrium," *Brit. M. Bull.,* 12:114–118 (1956); W. Steinhausen, "Die lebende Cupula in den Bogengangsampullen," *Verh. dtsch. Zool. Ges.,* 1:81–98 (1934); E. Huizinga, "On the Tonic and the Dynamic Function of the Cristae," *Acta Oto-laryng.,* 24:82–95 (1936); G. Dohlman, "Some Practical and Theoretical Points in Labyrinthology," *Proc. Roy. Soc. Med.,* 28:1371–1380 (1935); O. Löwenstein and A. Sand, "The Mechanism of the Semicircular Canal. A Study of the Responses of Single-Fibre Preparations to Angular Accelerations and to Rotation at Constant Speed," *Proc. Roy. Soc.,* 129B:256–275 (1940); A. A. J. van Egmond, J. J. Groen and L. B. W. Jongkees, "The Mechanics of the Semicircular Canal," *J. Physiol.,* 110:1–17 (1949); C. S. Hallpike and J. D. Hood, "Fatigue and Adaptation of the Cupular Mechanism of the Human Horizontal Semicircular Canal: an Experimental Investigation," *Proc. Roy. Soc.,* 141B:542–561 (1953); J. R. Lindsay, H. B. Perlman, H. G. Kobrak and W. D. Neff, "Physiology of the Ear," *Otolaryngology,* vol. 1, ch. 3 (1955).

(E. G. W.; W. D. N.)

EQUINOX, the term used to describe the two points where the projection of the sun's apparent annual path onto the celestial sphere (ecliptic) crosses the celestial equator (*q.v.*). When the sun is at either equinox, day and night are of equal length. The names of the two equinoxes are associated with the seasons of the northern hemisphere in which they occur. The vernal or spring equinox occurs about March 21, when the sun apparently crosses the celestial equator moving northward. The autumnal equinox occurs about Sept. 23 when the sun again crosses the celestial equator, but moving southward.

In locating the position of heavenly bodies, the vernal equinox serves as the initial reference point for the measurement of right ascensions along the celestial equator and for the measurement of celestial longitudes along the ecliptic.



CELESTIAL SPHERE SHOWING THE POSITIONS OF THE EQUINOXES

Approximately 2,000 years ago, the position of the vernal equinox coincided on the celestial sphere with the beginning of the constellation Aries. Thus the vernal equinox came to be re-

ferred to as the "first point of Aries." Because of the precession of the equinoxes, the vernal equinox is now in the constellation Pisces, about 30° removed from the constellation Aries. The traditional reference to the vernal equinox as the first point of Aries remains in use, however. (*See* ASTRONOMY; PRECESSION OF THE EQUINOXES.)

For centuries the vernal equinox has been associated with the rebirth of nature and has been celebrated by many cultures and religions. For example, in ancient Rome the death and resurrection of Attis, the god of vegetation, was celebrated on March 22–27. (*See* J. G. Frazer, *The Golden Bough*, 1922.) At the time of the equinoxes, it is commonly believed that strong gales may be expected. This popular idea has no foundation in fact, for continued observations have failed to show any unusual prevalence of gales on the dates of the equinoxes. (R. L. De.)

EQUITES (plural of Lat. *eques*, "horseman"), in ancient Rome the name given originally to part of the Roman army, the cavalry, but subsequently also to a political and an administrative class, the "knights" or the "equestrian order." According to tradition Romulus instituted a cavalry corps consisting of three "centuriae" called after the three primitive tribes. At some time before the reign of Servius Tullius they were doubled and gave rise to the *sex suffragia*, the six centuries which formed the most aristocratic portion of the 18 centuries of *equites* in the centuriate assembly attributed to him (*see* COMITIA). The insignia of the *equites*, the *phalerae* (metal disks worn on the breast), the gold ring and the purple stripe on the tunic, also point to an origin in the class of highest rank in the state. Up to the reform of the centuriate assembly, probably in 241 B.C., the important first century in the vote was chosen from the *equites*, but after that time the privilege was given to a century selected by lot from the first class.

Although the *equites* were drawn from the wealthiest families with senators and members of senatorial families among them (the property qualification in early time was probably that of the first class), they received financial aid from the state, *aes equestre* for the purchase of the horse and *aes hordearium* for its upkeep, a sum levied on the property of widows and orphans. Hence the name *equites equo publico*. Need for additional cavalry at the siege of the Etruscan city of Veii (406–396) led according to Livy to the enrolment of other citizens with the "equestrian census" who volunteered to provide their own horses and were called *equites equo privato*. Service with a horse, public or private, became practically an obligation on the wealthy. Liability to service before beginning a career lasted for ten years. Pay was three times that of infantry, but *aes hordearium* was later abolished.

However the *equites* were chosen earlier, by *curiae* (*see* CURIA), king or consuls, the choice was made in historical times by the censors, who reviewed them with their horses at each censorship. Those whose physique, character, horse and equipment were satisfactory were bidden to lead their horse on (*traducere equum*), the others were told to sell their horse (*equum vendere*) and expelled. M. Porcius Cato the censor in 184 made merry over an *eques* who was too fat and expelled so prominent a senator as Cornelius Scipio Asiaticus. Distinct from this review was the annual procession of the *equites* on July 15 from the temple of Mars to that of Castor and Pollux in celebration of their intervention in the battle of Lake Regillus. Both review and procession fell into abeyance in the late republic but were revived by the emperor Augustus. How far the *equites equo publico* provided the cavalry normally needed in the field remains uncertain. In the census taken in 225 the rolls revealed about 25,000 Romans and Campanians qualified to serve. The cavalry in general were prominent in the great wars and retained a privileged position as Polybius shows. But in the late republic foreign horse tended to replace them in the field and thus to restrict the equestrian service to posts as officers or members of the general's staff. Junior officers, appointed or elected, such as the *tribuni militum a populo* (*see* TRIBUNE), were regularly *equites*.

Rise of the Ordo Equester.—Livy's reference to an equestrian census rating as early as the siege of Veii when referring to the first *equites equo privato* may be anachronistic, while early references to the *ordo equester* describe the cavalry, in which senatorial and equestrian groups were not differentiated. Yet the existence of nonsenatorial members of the equestrian centuries, ex-members of these centuries and nonmembers whose wealth qualified them tended to create a group comparatively few of whom could rise into the senatorial class. Although they generally remained extensive landowners, they were distinguished from senators who were excluded from commerce by the *lex Claudia* of 218 by being able to enter freely into the fields of commerce, finance and public contracts, and to use the opportunities afforded by the public needs during the great wars, and by the collection of taxes and exploitation of public lands, mines and quarries in the provinces. They formed companies, controlled large interests, were closely associated with the *publicani* (*q.v.*) and were ready to be molded into a recognized political order. The date of a separate *census equester* remains uncertain.

The equestrian order was created by three measures of the Gracchan period (*see* GRACCHUS, GAIUS SEMPRONIUS): first, a law of uncertain date excluding senators from the equestrian centuries, though members of senatorial families remained; secondly, a law of Gaius Gracchus which made the juries in the extortion court wholly equestrian; and thirdly, another measure of Gaius Gracchus which provided that the contracts for the collection of taxes in Asia should be let by the censors in Rome. The reluctance of senatorial jurymen to convict guilty senatorial governors had produced serious scandals which Gracchus proposed to reform by limiting the jurors to non-senators 30 years old or over. The measure about the letting of contracts for tax-collecting gave the *publicani* a great advantage in that the Asian communities had no opportunity to bid against them, while the other meant in practice that the governor and his staff, whose duty was to see that collections were just, had to face juries closely related to the *publicani* upon their return. Either the provincials went unprotected because the governor or his staff were prevented from acting effectively on their behalf by fear of vindictive prosecution, like Rutilius Rufus who did carry out a policy of protecting the provincials, the governor or members of his staff faced false charges after his term. Both measures were designed to secure equestrian support for the Gracchan program; in effect the two questions created an equestrian class in politics and kept it at variance with the senate for two generations.

An attempt to restore the courts to the senate by Q. Servilius Caepio in 106 was only briefly successful, and a compromise carried by the younger M. Livius Drusus (*q.v.*) in 91 was annulled. Sulla (*q.v.*) restored them to the senate in 82, but added 300 *equites* to that body, an act the more surprising because the knights had given Marius and Cinna (*qq.v.*) much support and Sulla had devastated them in the proscriptions, breaking much of their political and economic influence. It is less surprising if, as some historians think, *equites* here meant not the equestrian order as defined by a census amount but the 18 centuries of *equites equo publico* which contained people of senatorial family. The Aurelian law of 70 sponsored by L. Aurelius Cotta as praetor and supported by Gnaeus Pompeius and M. Licinius Crassus (*see* POMPEIUS and CRASSUS) lessened senatorial influence by providing for juries consisting of one-third senators, one-third *equites* and one-third the group called *tribuni aerarii*, of lower census than the *equites*, or, as some think, of equal census with them. In 67 the Roscian law reserved the first 14 rows of seats in the theatre for the *equites*. The commercial interests of the equestrian order may have affected provincial policy in the late republic and certainly brought support to the laws against piracy and those which set up Pompey's great commands. The union of the equestrian and the senatorial orders which supported Cicero (*q.v.*) against Catiline and seemed to him to provide the best guarantee of stability in the state was soon shattered when in 60 the senate refused the request of the company which had obtained the contract for collecting the taxes in Asia to have the contract abated. Julius Caesar received their support in large measure, yet he abolished the farming of the taxes in Asia, gave the *equites* no favours in his financial reforms, and changed the personnel of the juries to one-half senators and one-half *equites*. Despite their losses in the proscriptions and the

economic measures of Octavian (later Augustus) thereafter in Italy, they were active in his support at Actium. With the advent of the principate they were completely changed.

The Equites Under the Empire.—Augustus reorganized the *equites* on a military basis. He revived the procession on July 15 and combined it with the censorial review and held reviews at other times, assisted by a board of senators. *Equites equo publico* and the *ordo equester* became one and the same. The right of bestowing the horse was vested in the emperor and the rank was given for life unless forfeited for offenses or loss of fortune, though the *eques* could give up his horse at the age of 35. Senators' sons were eligible by right of birth but ceased to be knights on entrance to the senate. Qualifications were free birth (under Tiberius for three generations), good health, good character and an equestrian fortune, but a few cases are on record of the grant of the gold ring and the rights of the freeborn to freedmen. Sometimes children were admitted. One of five years is known and Marcus Aurelius was an *eques* at six. Insignia were the gold ring, the narrow purple stripe on the tunic, and among their privileges were the right to sit in the first 14 rows in the theatre and the circus. The active *equites* were divided into at least six squadrons (*turmae*), each under the command of an annually appointed *sevir*, who was expected to give games during his term. Their total number is uncertain, but Dionysius of Halicarnassus saw a review of more than 5,000. They met, passed resolutions in honour of the imperial house, and acclaimed the young princes as *principes iuventutis* (heads of the younger generation). The corps remained as a social organization into the late empire. Holders of the rank became increasingly important in the administration and society of the empire.

In an equestrian career a number of subordinate military posts, the *militiae equestres*, were obligatory, though dependent on nomination by the emperor. They were the command of an auxiliary cohort, the military tribunate in a legion, and the command of an auxiliary cavalry squadron (*ala*), and were regularly held in that order. To these Septimius Severus added the centurionate. A purely civil career became possible in the 2nd century with the expansion of the knights in the imperial household. From the military posts the *equites* advanced in the imperial service, somewhat irregularly in the early empire as the service itself was created. They held various financial procuratorships in imperial provinces and in charge of imperial properties in senatorial ones, and, in the city, prefectures such as the grain supply, the watch and the praetorian guard. They also commanded the fleet, held the military command of Egypt, a post not given to senators, and held military authority in a number of smaller provinces such as Noricum and Judaea. The Egyptian command and the praetorian prefecture were the crown of an equestrian career.

Equites Singulares Augusti.—The personal bodyguard of the emperor, consisting of foreign soldiers (chiefly Germans) was abolished by Galba and revived with this title under the Flavian emperors. Like the auxiliary troops, they served for 25 or more years and received Roman citizenship upon discharge, but occasionally included Roman citizens. They were recruited from the provincial cavalry. They numbered at first 500 and later 1,000. They had their own camp near the Lateran, and from the time of Septimius Severus onward were organized in two divisions under two tribunes of the praetorians.

BIBLIOGRAPHY.—T. Mommsen, *Römisches Staatsrecht* vol. 3 pp. 476 *et seq.* (1887); Kübler in Pauly-Wissowa, *Real-Encyclopädie der classisches Altertumswissenschaft* vol. 5, 1, 272–312 (1907); Hirschfeld, *Die kaiserliche Verwaltungsbeamten bis auf Diokletian* (1905); A. Stein, *Der römische Ritterstand* (1927); H. Hill, *The Roman Middle Class* (1952); H.-G. Pflaum, *Les procurateurs équestres sous le Haut Empire Romain* (1950); A. Alföldi, *Der frührömische Adel und seine Ehrenabzeichen* (1952); On the *Equites singulares:* W. Liebenam in Pauly-Wissowa, *Real-Encyclopädie der classisches Altertumswissenschaft*, vol. 5, 1, 312–321 (1907); A. von Domaszewski, *Die Rangordnung des römischen Heeres* pp. 50–53 (1908). (T. R. S. B.)

EQUITY, in popular usage, is a term synonymous with fairness and natural justice and, in this sense, plays a vital role in the work of judges, jurors, legislators and government administrators. In its more specialized meaning, equity refers to the principles and remedies originally developed by the old English court of chancery and administered in modern times in England and the United States by courts with comparable powers. Traditionally characterized by a discretionary treatment of the individual case in accordance with notions of natural justice, thus supplementing the common law, equity has been a creative, reforming force in Anglo-American law.

England.—The historical origins of the court of chancery were significant. By the end of the 13th century the king's common-law courts had largely limited the relief available in civil cases to the payment of damages and to the recovery of the possession of land and of chattels. They had refused to extend and diversify their types of relief so as to meet the needs of new and more complex situations. Disappointed litigants had turned to the king and council with petitions for justice because the courts had afforded either no remedy or one that was ineffective. These petitions were referred to the lord chancellor, who was the king's prime minister, a powerful member of the council, a bishop learned in the canon law and the chief secretary of the government. By the early years of the 14th century the petitions were going directly to the chancellor, and by the middle of that century the chancery was recognized as a new and distinct court.

The developments thus initiated resulted in the fashioning by the chancellor of new equitable remedies, together with policies governing their operation and collateral effects. The following are representative: specific performance of contract, whereby the victim of a breach might compel the exact performance promised if damages would be a poor substitute, as in contracts to sell land and unique chattels; the enforcement of trusts, where one who had been given title to property in order to manage it for another was required to fulfill his fiduciary obligations (*see* TRUST); injunction to prevent threatened or continuing wrong, such as destruction of the plaintiff's invaluable shade trees (*see* INJUNCTION); restitution of benefits wrongfully acquired, by compulsory surrender of the ill-gotten gains, in order to prevent unjust enrichment; the correction and cancelation of written instruments for mistake and misrepresentation; and the equity of redemption, which enabled a defaulting mortgagor to reclaim his land if he tendered principal and interest within a reasonable time after forfeiture and before foreclosure (*see* MORTGAGE). Such new equitable remedies, in contrast to the narrow rigidity of the common-law remedies, had far-reaching consequences.

This growth was facilitated by the unusual procedures of the court of chancery. At a time when the parties in a law case were not permitted to testify because of their prejudicial interest, the court of chancery, on a subpoena borrowed from the heresy procedures of the church, interrogated the defendant in an equity case regarding every detail of the complaint. As compared with the tendency in the courts of law to leave most issues of fact to the verdict of a jury, the chancellor himself found the facts in equity cases, sitting alone as a professional trier of the facts, without a jury. Where the law court's judgment spoke only in terms of a successful plaintiff's right to have and to recover his damages or property, to be enforced if practicable by the sheriff's execution, the decree of the chancellor, in the name of the king, imposed a personal duty on the defendant to comply with specific orders, violations to be punished as contempt, not of the chancellor, but of the king. In contrast, the refusal of a losing defendant in a law case to comply with the judgment was not contempt of court.

The full growth of the equitable remedies was retarded, however, by political pressures from the judges and parliament not to trespass upon the province of the separate law courts. As a result, the chancellor was forced to agree not to hear any case where there was an adequate remedy (usually damages) at law. In some situations, *e.g.*, the trust, the courts of law afforded no remedy. Where they did, as in cases of breach of contract and tort (wrongs other than breach of contract and crime), specific relief could not be obtained unless the remedy at law was inadequate or the threatened injury would be irreparable.

Another restrictive influence was the development in the chancery of a system of precedent. For generations the chancellors had not considered themselves bound by precedents or rules of law; emphasis had been put mainly upon the discretionary treatment of

needs of the individual case. Indeed, reports of cases decided in the chancery did not begin until the middle of the 16th century. The courts of law, on the other hand, had long adhered to precedents and rules; their reports went back to the late 13th century. Cardinal Wolsey was the last of the great ecclesiastical chancellors. His successors were usually common lawyers and they, utilizing the new reports, began shaping equity into an established set of rules. By the middle of the 17th century the equity administered by the court of chancery had become a recognized part of the law of the land: equity gave justice according to law rather than executive justice. By 1818 Chancellor Eldon was able to say: "The doctrines of this court ought to be as well settled and made as uniform almost as those of the common law, laying down fixed principles, but taking care that they are to be applied according to the circumstances of each case. I cannot agree that the doctrines of this Court are to be changed with every succeeding judge. Nothing would inflict on me greater pain, in quitting this place, than the recollection that I had done any thing to justify the reproach that the equity of this court varies like the Chancellor's foot." (*Gee* v. *Pritchard*, 2 Swanst. 402, 414; 36 Eng. Reprint 670, 674.)

By the Judicature act, 1873, the competitive, separate law and equity courts, with their attendant delays, expense and injustices, were abolished and their work assigned to a single, departmentalized supreme court of judicature. This consists of the court of appeal and the high court of justice, which for many years has sat in three divisions: the chancery division, of which the lord chancellor is the head; the queen's bench division, the successor to the old law courts; and the probate, divorce and admiralty division. Most equity matters are dealt with in the chancery division, sitting in London. Some, such as specific performance of contract and injunction against tort, may also be dealt with by the judges of the queen's bench division, in London and in the provinces. And the county courts have equity jurisdiction limited (except by consent) to cases where the amount involved does not exceed £500. (*See* also Court; English Law; Common Law.)

United States.—In the early history of the United States there was strong opposition to the development of equity. To a large extent this was attributable to the historic relation between the court of chancery and the king in England, to abuses by colonial governors while serving as chancellors in some of the colonies, to distrust of the broad discretion exercised by the chancellors, and to the lack of trial by jury in equity cases. Thus, in 1789, when congress enacted the first federal Judiciary act, courts of chancery existed in only five states; in three states there were only common-law courts; in three others the law courts had a few equity powers; and in two states the legislature to a limited extent served as a court of equity.

Nevertheless, congress gave to the new federal courts jurisdiction at law and in equity, to be exercised by the same judge in the same court but by distinct procedures. The procedure in equity cases was to be that of the English court of chancery, as modified by rules of the supreme court of the United States. And the adequacy test was emphasized by a restrictive statute that "suits in equity shall not be sustained in any court of the United States in any case where a plain, adequate and complete remedy may be had at law" (1 Stat. 82, sec. 16).

Eventually all of the states established courts of equity, either as separate tribunals or as in the federal courts. This development came slowly, however, in Pennsylvania and the New England states; Massachusetts did not confer full equity powers upon its courts until 1877.

During the last hundred years most U.S. courts have abolished the distinctions between actions at law and suits in equity and have fused their administration in one procedural system, with but one civil action, in the same court. The movement began with the Field Code of Civil Procedure, enacted in New York in 1848 and later in over half of the states. The federal courts followed the same pattern in the Federal Rules of Civil Procedure, promulgated by the supreme court in 1938. (*See* Practice and Procedure: *United States.*) New Jersey in 1947 followed the English pattern, with a chancery division in the new superior court. In the 1960s separate law and equity courts and procedures were still maintained, however, in Delaware, Maryland (in Baltimore), Mississippi, Tennessee and Vermont.

An important by-product of the fusion of law and equity in the United States has been a decreasing emphasis upon the inadequacy of other remedies as a condition to equitable relief. It is significant that congress in enacting the Revised Judicial code of 1948 omitted the statutory adequacy test of 1789 as obsolete, it being superseded by the fusion provisions of the federal rules of 1938.

Modern equity, often assisted by legislation, has greatly expanded the equitable remedies. More of this has taken place in the U.S. than in England, where precedents and practices have sometimes limited the development of equity. Thus, specific performance of contract has been extended beyond contracts for land and unique chattels and has been made available where the contract called for the entire output of a factory, an orchard or an oil well; for the construction of a building or the joint use of a railroad terminal (where the plaintiff's dire need or the public interest justified the court's assumption of the burden of supervision); or for an employee to refrain from competing with his employer after discharge or resignation. And statutes have facilitated specific performance of co-operative-marketing contracts and agreements to arbitrate future commercial or labour disputes.

Injunctions have been extended beyond merely negative restraint to the so-called mandatory injunction, compelling affirmative action (*e.g.*, the removal or restoration of a structure). And where the emergency need is urgent, either negative or affirmative injunctive relief may be obtained promptly, in advance of trial. The old notion that equity only protects property rights has been virtually abandoned; an injunction may now be had, where other factors of appropriateness permit, against threatened injury to interests of personality, such as civil liberties, privacy, reputation and the domestic relations. Enabling legislation has immensely increased the resort to injunction by government agencies, especially to prevent violation of regulatory statutes, notwithstanding criminal penalties. Conversely, legislation has curbed the issuance of injunctions in some areas. Thus, congress has severely restricted federal court injunctions in labour disputes and has provided that only a federal trial court of three judges may enjoin the enforcement of a state statute as unconstitutional.

Legislation has strengthened the measures for enforcing equity decrees. Punishment for contempt of court long ago lost much of its efficacy for this purpose, and enforcement was further weakened in relation to decrees for the payment of money by the abolition of imprisonment for debt. Even before this occurred, however, some legislatures had extended to such decrees the common-law method of enforcing law judgments for damages, namely, the sheriff's sale of the defendant's assets, including interests in land. Such legislation has become almost universal. Similarly, legislation has everywhere overcome the reluctance of the court of equity, for historical reasons, to transfer interests in land directly and has enabled the court to effect such a transfer by a provision in the decree or the action of an officer of the court.

Bibliography.—E. H. Snell, *Principles of Equity*, 25th ed. (1960); Frederic W. Maitland, *Equity*, ed. by Brunyate, rev. ed. (1947); Zechariah Chafee, Jr., *Some Problems of Equity* (1950); Henry L. McClintock, *Equity*, 2d ed. (1948); Sidney P. Simpson, "Fifty Years of American Equity," 50 *Harvard Law Review* 171 (1936).

(M. T. V. H.)

EQUIVALENT, in chemistry, the quantity of a substance which exactly reacts with, or is equal to, the combining value of an arbitrarily fixed quantity of another substance in a particular reaction. It follows that substances react with each other in stoicheiometric (*see* Stoicheiometry) or chemically equivalent proportions and that a common standard must be adopted. For an element the equivalent weight is that quantity which combines with or replaces 1.00797 g. of hydrogen or 7.9997 g. of oxygen or their equivalent; or, the weight of an element that is liberated in an electrolysis by the passage of 96,500 coulombs of electricity. The gram atomic weight of an element divided by its valence gives the equivalent weight. For example, some equivalent weights are: silver (Ag), 107.870 g.; magnesium (Mg), 24.312/2 g.; aluminum (Al), 26.9815/3 g.; sulfur (S, in forming a sulfide), 32.064/2 g.

For compounds that function as oxidizing or reducing agents (*see* OXIDATION AND REDUCTION), the equivalent weight is the gram molecular weight divided by the number of electrons (*see* ELECTRON) lost or gained by each molecule; *e.g.*, potassium permanganate ($KMnO_4$) in acid solution, 158.038/5 g.; potassium dichromate ($K_2Cr_2O_7$), 294.192/6 g.; sodium thiosulfate ($Na_2S_2O_3.5H_2O$), 248.1825/1 g. For all oxidizing and reducing agents, elements or compounds, the equivalent weight is the weight of the substance that is associated with the loss or gain of N (6.02486×10^{23}) electrons. The equivalent weight of an acid or base for neutralization reactions, or of any other compound that acts by double decomposition, is the quantity of the compound that will furnish or react with or be equivalent to 1.00797 g. of hydrogen ion or 17.0074 g. of hydroxyl (OH^-) ion; *e.g.*, hydrochloric acid (HCl), 36.461 g.; sulfuric acid (H_2SO_4), 98.078/2 g.; sodium hydroxide (NaOH), 39.997 g.; calcium hydroxide $Ca(OH)_2$, 74.09/2 g.; sodium carbonate (Na_2CO_3), 105.989/2 g.

The equivalent weight of a substance may vary with the type of reaction it undergoes. Thus potassium permanganate reacting by double decomposition has an equivalent weight equal to its gram molecular weight, 158.038/1 g.; as an oxidizing agent under different circumstances it may be reduced to the manganate (MnO_4^{2-}), to manganese dioxide (MnO_2) or to the manganous ion (Mn^{2+}), with the equivalent weights of 158.038/1 g., 158.038/3 g. and 158.038/5 g., respectively. The number of equivalents of any substance dissolved in one litre of solution is the normality of that solution. (J. B. Ps.)

ÉRARD, SÉBASTIEN (1752–1831), French manufacturer of musical instruments, distinguished especially for the improvements that he effected in the harp and the pianoforte, was born at Strasbourg, France, on April 5, 1752. Before he was 25 he had set up in business for himself, his first workshop being in the hotel of the duchesse de Villeroi. There he constructed in 1777 his first pianoforte, which was also one of the first manufactured in France. Soon overwhelmed with commissions, and finding assistance necessary, he sent for his brother, Jean Baptiste, in conjunction with whom he established in the rue de Bourbon, in the Faubourg St. Germain, his famous piano manufactory.

On the outbreak of the Revolution he went to London where he continued to make pianos. Returning to Paris in 1796, he began to build grand pianofortes, made in the English fashion, with improvements of his own. In 1808 he again visited London, where, two years later, he produced his first double-movement harp. In 1812 he returned to Paris where, in 1823, he crowned his work by producing his model grand pianoforte with the double escapement. He died at Passy on Aug. 5, 1831. *See* also HARP and PIANOFORTE.

ERASISTRATUS (*fl.* 300 B.C.), Alexandrian anatomist, was born at Ceos (Keos) near Attica, and lived first at the court of Seleucus Nicator and then at Alexandria, in Egypt, where he formed a school of anatomy. Like Herophilus (*q.v.*) he traced the origin of the nerve trunks to the brain and distinguished sensory and motor nerves. But unlike Herophilus, he rejected the old humoral pathology and supposed that the arteries contained pneuma or "vital spirits" which, when it escapes, allows the blood to flow from the veins into the arteries, thus making it possible for the arteries to bleed. Any hindrance in the action of the pneuma, or any overflow of the blood into the arteries, resulted in disease; yet, curiously enough, he did not advocate bloodletting. Erasistratus is to be credited with a detailed knowledge of the convolutions of the brain, the investigation of the bile, spleen and liver and of the anatomy of the heart, the naming of the trachea and the invention of a catheter.

ERASMUS, DESIDERIUS (*c.* 1466–1536), humanist and the greatest figure in the northern Renaissance, was born at Rotterdam or possibly Gouda in the Low Countries, Oct. 26–27, in about 1466. He was the second child of Margaret, a physician's daughter, and Roger Gerard, priest. The romantic story woven by Erasmus round these facts is suspect, but the legal impediments of illegitimacy haunted his later career. He went first to school at Gouda where Peter Winckel (later one of his three guardians) was his rather obtuse usher, and spent the years 1475–83 till his father's death at the school of St. Lebwin's, Deventer (with a short spell as choir boy at Utrecht). There, though he could have had little contact with them, such teachers as John Sintheim (Sinthius) and Alexander Hegius, members of the Brethren of the Common Life (*q.v.*), were opening the school to humanist ideas. In this part of Holland the influence of the *devotio moderna* was strong, and Erasmus owed more than he knew to its practical Christocentric piety, as he was repelled by the obscurantism and austerities of its latter end. When the guardians sent Erasmus and his elder brother Peter to the school at 's Hertogenbosch, Erasmus found the shades of a coming claustral discipline irksome. But they were in no position to withstand their guardians' plans, and Peter entered a monastery at Sion near Delft, while Erasmus became an Augustinian canon at Steyn. There he made or renewed friendships with William Herman, Cornelius Aurelius and Servatius Roger. He was ordained priest on April 25, 1492.

Though he now had an opportunity to study the classics and the poets, he felt the constraint of monastic life, which bred in him a passion for personal freedom and an irritation with restraint which became a trait of his complex character. He grasped eagerly the opportunity to leave, probably in 1494, and became Latin secretary to Henry of Bergen, bishop of Cambrai. But he did not take kindly to the sort of diplomatic career which suited many scholars of the age, and was glad when his friend James Batt secured permission for him to study theology at the University of Paris. There he entered the college of Montaigu at the time when the rector, the formidable Jan Standonck, was attempting to graft onto Paris the piety of the *devotio moderna* in its most austere form. It was too much for Erasmus, whose health gave way, nor did his heart lie in the scholastic theology or in preaching, in which he dabbled at this time. Like other humanists, he was affronted by the dogmatic theologians, especially of the religious orders, with their party cries and intolerance and their violence and wooden hostility toward new methods. In Paris, however, he made acquaintance with the humanists, and a commendatory epistle in Robert Gaguin's *De Origine et gestis Francorum compendium* (1495) was his first writing to attract public attention. He returned to Paris from a convalescence in Holland and took pupils, first Christian and Henry Northoff from Lübeck, and then the Englishmen Thomas Grey and Robert Fisher. He wrote manuals for their instruction, *De conscribendis epistolis, De copia verborum, De ratione studii* and another *Familiarum colloquorum formulae* ("Familiar colloquies"), which underwent many revisions and became one of his most famous works.

First Visit to England.—What Erasmus really wanted was to be allowed to study in reasonable comfort, and his employments were of value only to that end. He had gradually to build up a precarious income, which always lagged behind his reputation, from pensions, gifts and dedications which in these early years involved fulsome touting and which caused William Tyndale to say that Erasmus was one "whose tongue maketh of little gnats great elephants, and laudeth up above the skies, whosoever giveth him a little exhibition." His tutorial adventures introduced him to William Blount, Lord Mountjoy, who proved a steadier patron than most, and whose invitation to England Erasmus accepted in 1499. Despite his often acid criticisms of England and the English, their barbarism, their weather, their beer, the country fascinated him and drew him back again and again. One reason no doubt was the patronage of Archbishop William Warham and John Fisher, and even more the congenial friendship of such scholars as William Grocyn, Thomas Linacre, William Latimer and, above all, John Colet and Thomas More. Of these last he has left verbal portraits which have the vivid detail of a Holbein painting. On him Colet made an early vivid impression, while his more intimate friendship with More ripened through the years. Though Colet was not the first to turn his thoughts from "good letters" to "sacred letters," he did challenge Erasmus to join him at Oxford in the battle against the obscurantists and in the exposition of a more directly biblical theology. Erasmus rightly saw that he had not the technical equipment for such intervention, but he was now evidently more deeply concerned for "sacred letters." "I wished," he wrote, "that good letters should find that Christian character

which they have lacked in Italy, and which, as you know, ended in glorifying pagan morality." Erasmus' new concern did not diminish his interest in the classical authors, for he stood for a combination of "good" with "sacred letters." He had long been acquainted with such poetic models as Virgil, Horace, Ovid, Juvenal, Statius, Martial, Claudian, Persius, Tibullus and Propertius, and in prose with Cicero, Quintilian, Sallust and Terence. Throughout his life he edited works by Latin authors (*e.g.*, Pliny, Seneca), and produced translations of Greek classical writings (*e.g.*, Lucian).

The Wandering Scholar.—In 1500 he left England, horrified at having to relinquish his earnings to the customs at Dover, and, in June, worked off some of his disgust by producing a collection of proverbs, the germ of his great *Adagia*. He spent the next months between France and Holland, and in this unexciting period began arduously to gain that mastery of Greek which was the foundation of his later pre-eminence. His thoughts began to turn toward a great edition of Jerome, the archetype of his own ideal combination of good and sacred letters. At St. Omer he learned to know and admire the saintly Franciscan Jean Vitrier, who may have deepened in him a genuine strain of Pauline piety not to be overlooked. One result was the little *Enchiridion militis Christiani* (1501), which reflects some of the most attractive elements in his religion, at the close of which he wrote,

> Make the prophets, Christ and the Apostles your friends. Above all, choose Paul . . . it has long been my cherished wish to cleanse the Lord's temple of barbarous ignorance and to adorn it with treasures from afar, such as may kindle in generous hearts a warm love for the scriptures.

He continued to produce edifying discourses at intervals, though none of them attained the quality of the *Enchiridion*. Thus followed his *De interdicto esu carnium* (1522), *Exomologesis* (1524), a *Modus orandi Deum* (1533), and *Institutio Christiani matrimonii* dedicated to Catherine of Aragon. He commented on certain Psalms in 1528 and at the end of his life discussed preaching in a comment on the book of Ecclesiastes.

In the summer of 1504 Erasmus found in the library of the Premonstratensians at Parc, near Louvain, a manuscript of Lorenzo Valla's *Annotationes* on the New Testament which criticized the Vulgate and advocated a critical method giving priority to exact, philological exegesis. The encounter was momentous for Erasmus, who published his own edition of this manuscript through Josse Badius (1505). He returned to England (1505–06), where he made his own copy of the Greek New Testament from such manuscripts as he could borrow. But a half-promised benefice did not materialize, and he seized an opportunity to visit Italy as tutor of the sons of Giovanni Battista Boerio, physician of Henry VII. He took his D.D. at Turin (Sept. 4, 1506) and in the next months saw at first hand the splendour and secularization of the church in Italy, and those bloody consequences of papal diplomacy which much offended his temper and his conscience. He moved to Bologna and to Venice, where he collaborated with the great publisher Aldus Manutius. There, in the house of Andrea Asolani, he shared lodgings with the phenomenal young scholar Jerome Aleander (later, turned diplomat, to be what A. Renaudet calls "his most intimate enemy"). At Venice in Sept. 1508 was published a great edition of his *Adagia* swollen to over 3,000 proverbs collected from the classical authors, a work which established his reputation as the foremost scholar in northern Europe. He became tutor to the young archbishop of St. Andrews, Alexander Stewart, and went with him to Padua and Siena. He visited Rome, but though he found friends at the papal court England drew him back. He returned there to stay with Sir Thomas More, and there he wrote (1509) the *Encomium moriae* ("Praise of Folly") in that Lucianic vein which was congenial to both him and More.

Erasmus and More had exercised themselves in turning Lucian into Latin, so that the *Encomium moriae* drew upon ancient and medieval inspiration but most deeply from Erasmus' own wit; it justifies J. Huizinga's comment that "only when humour illuminated that mind did it become truly profound," for the sharp satire became one of the most popular and enduring of his writings. He moved to Cambridge university where he lectured on Greek and on the epistles of Jerome. Archbishop Warham gave him the benefice of Aldington in Kent, which he commuted into a pension of £20. But Erasmus could not settle anywhere for very long and was off again in 1514 to Basel, to a new fruitful partnership with the publishing house of Johann Froben. About this time there appeared a scurrilous satire against the late pope—the *Julius exclusus*. The balance of probability is slightly in favour of Erasmian authorship, tipped by the curiously evasive character of Erasmus' denials.

A shock came in June 1514: his old friend Servatius Roger, now prior of Steyn, wrote recalling him to the monastic life. His reply was a passionate plea for freedom, a defense of his own evident vocation for letters. It now became urgent for him to find authoritative support for his way of life, and he drew up an appeal to the papal chancery in which he put his case under the name of an imaginary Florentius. (This appeal, addressed to "Lambertus Grunnius," together with the letter to Roger and his own *Compendium Vitae*, compiled in 1524, are primary sources of the life of Erasmus.) In Jan. 1517 he received two dispensations from Pope Leo X, the one allowing him to live in the world, and the other dispensing him from the dress of his order. On his return to England he received these dispensations formally on April 9, 1517, at the hands of his friend Ammonius Andreas, the papal legate.

The New Testament.—In 1516 he had published his most immediately influential work, the edition of the New Testament which gave priority to the Greek text. He compared such manuscripts as were available to him, and various readings of the Fathers, while other scholars, N. Gerbelius, J. Oecolampadius and W. F. Capito, assisted with the philological apparatus, especially with Hebrew. Although he did not have access to the earliest manuscripts, he provided a much more accurate text than that of the Vulgate. For instance, he omitted the reference to the Trinity interpolated into I John v, 7–8 by most Latin manuscripts (though he was forced to put it back in later editions). Alongside the Greek text he placed his own elegant Latin version, which greatly attracted scholars, and he appended critical notes, some of which were far from academic, many being as biting as the glosses of later Reformers. His prefatory essay was a noble plea for the study of the Scriptures, including the famous lines:

> I could wish that every woman might read the Gospel and the Epistles of St. Paul. Would that these were translated into each and every language so that they might be read and understood not only by Scots and Irishmen, but also by Turks and Saracens . . . Would that the farmer might sing snatches of Scripture at his plough, that the weaver might hum phrases of Scripture to the tune of his shuttle, that the traveller might lighten with stories from Scripture the weariness of his journey.

The work was dedicated to Pope Leo X, who accepted the dedication. Nonetheless, it was strongly attacked, not least by the Englishman Edward Lee, by the Fleming J. Latomus, and by the Spaniard S. Zuñiga. There were indeed obvious inaccuracies in the work, many of which were corrected in later editions. But this, the work of one man, could vie with the great Spanish team of Alcala and had a deeper influence on contemporaries than did their Complutensian Polyglot. To forward-looking scholars (like Vadianus of St. Gall, Luther and A. Karlstadt in Wittenberg, Thomas Bilney and Robert Barnes in Cambridge) the new version offered a more attractive, truer text, while its prefaces amounted to a manifesto of the new learning, expounding the new critical method, and at the same time transmitting important gleanings from Erasmus' reading of the Fathers. Important too were his paraphrases, first of the Epistles, then of the Gospels, the English translation of which had an important and official vogue in the reign of Edward VI.

At Louvain.—Erasmus spent the years 1517–21 at Louvain, where he took great interest in the new college for the study of the sacred languages founded by Jerome Busleiden. During this period his correspondence greatly increased and he became one of the most prolific letter writers among the humanists for whom semiprivate correspondence of this kind afforded a prime means of learned conversation. A high proportion of the letters of Erasmus are distinguished from the bulk of humanist correspondence only by their more polished Latin, but every now and again, writing

to some intimate, B. Rhenanus or Ammonius Andreas, something new appears, pictures of contemporary life wittily observed, that Brueghel-like quality which in his *Encomium morias* and in his *Colloquia familiaria* anticipated the writing of a new age.

At Basel.—In Nov. 1521 he settled at Basel, perhaps the most satisfying to him of his many abiding places. From there he published a great series of works, and became a central figure among those scholars for whom the return to the Fathers of the Church was an integral part of the renewal of sacred letters. His edition of Jerome in nine volumes (1516) was a great achievement, but others followed: Cyprian (1520), Pseudo-Arnobius (1522), Hilary (1523), Irenaeus (Latin, 1526), Ambrose (1527), Augustine (1528), Chrysostom (Latin, 1530), Basil (1532), Origen (Latin, 1536), though in some of the later editions Erasmus took little direct personal share. In Basel, too, he revised and extended the range of the *Colloquia familiaria* into which he contrived to pack an amazing amount of comment on contemporary manners, some of it "replete with malice spiteful."

The outbreak of the Reformation ended the Erasmian dream of a golden age of letters. His own criticism of the church had been formidable enough for Aleander to denounce him as the real author of schism, and worse than Luther. To Erasmus, however, Luther must have seemed a typical dogmatic theologian, violent, aggressive. Luther in his first literary contacts (they never met) was polite, though he soon diagnosed a deep divergence of which the Erasmian preference for Jerome against Augustine was symptomatic.

As the church struggle widened, both sides appealed to Erasmus, who was at one time hailed as the author both of Luther's *De Babylonica captivitate* (1520) and of Henry VIII's reply. Yet it counted for Luther that in critical months during 1519–21 Erasmus did not come out against him, and indeed offered his services and mediation, while in an interview with Frederick the Wise of Saxony Erasmus himself may have stiffened the resolve of the prince that Luther should not be refused a hearing at the Diet of Worms. In the next years Erasmus was more and more alienated from the Reformers, though some of them, including Huldreich Zwingli, John Oecolampadius, Martin Bucer and Philip Melanchthon, had been among his fervent admirers. A long wrangle with J. Lefèvre d'Étaples had shown him to be hypersensitive to criticism, and in 1523 a bitter polemic with Ulrich von Hutter resulted in *Spongia Erasmi . . . adversus aspergines Hutteni* ("Erasmus' sponge against Hutten's mire"), the least creditable of all his writings. Then, in 1524, Erasmus yielded at last to the prompting of his patrons and in great haste attacked Luther in *De libero arbitrio*. The work incensed his opponents without much satisfying his friends, and, though an able enough essay in Erasmian moderation, it exposed too his theological limitations. Luther's reply, the *De servo arbitrio* (1525), violent and perverse, is in a more profound dimension. Erasmus worked off his resentment in the tiresome *Hyperaspistes* (1526), but he much improved a second edition which contains some of his finest mature utterances.

Last Years.—Though the triumph of the Reformation in Basel under Oecolampadius in 1529 was less violent than elsewhere, Erasmus with other humanists left and went to Freiburg in Breisgau. There he stayed until 1536 when, aging and unwell, he returned to Basel. Although an imperial councilor, who from the time of writing his *Institutio principis Christiani* (1516) had sought to advise the secular powers, he did not attend the Diet of Augsburg. But his interest in the peace of the empire and of the church brought from him in 1533 *De amabili ecclesiae concordia*, and among those who were trying to heal the breach there were moderate men who were his disciples. A new pope, Paul III, nominated him as dean of Deventer, and at last there was talk of a cardinal's hat. But it was all too late in the day, for Erasmus had outlived his preoccupation with dignities, and turned instead to paraphrase a psalm for a humbler but loyal friend, Christopher Eschenfelder, the customs officer at Boppard, to set his affairs in order (he had sold his library in 1534 to the Polish nobleman Johannes a Lasco) and to prepare for an end which could not be far off. He died in Basel on July 12, 1536. That he died without the sacraments shocked Luther, but Boniface Amerbach, his heir, who was with him to the end, wrote of his death, "As was his life, so was the death of this most upright of men. Most holy was his living, most holy his dying." Poignantly the last words of the great cosmopolitan, *Lieve God* ("Dear God"), were in Dutch.

Character and Teachings.—It is unlikely that Erasmus had any integrated, doctrinaire program of a Christian renaissance, and it is to read too much into the famous "I will put up with this Church until I shall see a better" to suppose that he had serious thoughts of a third church, neither Roman nor Protestant. Of his practical obedience to church authority there is no doubt, as he expressed it to Luther: "I always freely submit my judgment to the decisions of the Church whether I grasp or not the reasons which she prescribes." Nor was he much preoccupied with Renaissance notions of human dignity or with the individual, but his deep concerns for freedom and peace, personal and temperamentally conditioned as they were, became ruling convictions about how human affairs should be ordered, not less significant because they ran counter to the stormy temper of the new age.

His satire, like that of other Catholic humanists, was directed especially at the religious orders and their theologians, at an element of legalism and superstition in contemporary religion which Gilbert Burnet later called "superannuated Judaism." But these witty commentaries were perhaps a too-acid ferment which bit corrosively, deepening the intense anticlericalism of the age.

His humour, easily mistaken for levity, and his love of Lucianic satire, gave his enemies a ground for asserting him to be more skeptical than he dared appear, and certainly some of the new radical notions put forward by C. Hoen, M. Servetus and L. Hetzer came from men who had studied his writings. His own stress on practical piety, on the "philosophy of Christ," was an oversimplification which failed to comprehend the dogmatic theology of Augustine, of the schoolmen, even of Luther. Yet his distinction between matters central to the Christian religion and things peripheral, which influenced the emperor Charles V and his advisers, was fruitful in a later age, not least in the tradition of Richard Hooker and of William Chillingworth. It was not perhaps lack of energy alone which prevented him from mastering Hebrew, but that he was as instinctively repelled by it as he was drawn by Greek, and his general distaste for the Old Testament led him nearer than he knew to a gnosticism in which the moralists of the classical world supplied the place of the Old Testament.

Beatus Rhenanus has left a vivid portrait of him: ". . . his figure compact and elegant. His constitution was extremely delicate and easily affected by trifling changes, as of wine or food or climate . . . his complexion was fair, with hair that in his younger days had a touch of red, bluish grey eyes and a lively expression of face." The paintings by Metsys, Holbein and Dürer bring vividly to mind the little Dutchman with the frosty twinkle in the clear eyes, the sensitive hands and fingers.

The range of Erasmus' knowledge, his wit and style and his use of Greek brought him fame as a scholar. As an editor and expositor of classical texts and the writings of the Fathers he was surpassed perhaps by other humanists, by Budaeus, Lefèvre d'Étapes and possibly Rhenanus, but none of them could have produced the later editions of Erasmus' *Adagia* with its immense flair for wide-ranging, apposite allusions. Nor could they have achieved the attractive satire of the *Praise of Folly* and the *Colloquies*. In his combination of textual criticism and a theology based directly on the New Testament rather than on the categories of the systematic theologians, he has some claim to be regarded as the first modern New Testament scholar, and in this lies his importance as a theologian, scholar and divine.

As a scholar he had obvious limitations. He was bred in the rhetorical tradition of literary humanism, which had little care for the artistic achievements or the scientific premonitions of the age. His imagination could run away with him, well out of sight of facts, and the editions of his letters with more than half the dates misplaced show how far removed he was from modern views of exact scholarship.

But if he could be hasty and slipshod, if like Luther he wrote too much and too fast, it was because he, also like Luther, wrote to the new, speedier rhythm of the age of printing. The older

humanism was always in peril of dilettantism, a study for study's sake which handed round manuscripts and gossiped learnedly about them within a closed circle. Erasmus knew it was important to get the books out into the world, to make the humanist tools available for the new widening audience of lettered laymen.

His limitations of character, too, are evident. Finicky, peevish, spiteful, in his early years sycophantic to a nauseating degree, almost pathologically thin-skinned, he was saved by his humour from pomposity and from undue preoccupation with the importance of being Erasmus. But the judgment of P. S. Allen remains true: Erasmus was truly great, and the clue to his greatness lay in the combination of brilliant intellectual gifts with absolute sincerity and enduring purpose.

The many-sided character of his influence is the best testimony to the reality of Erasmian moderation. In England Robert Barnes and Thomas Bilney among the first Reformers, and their opponents Stephen Gardiner and Sir Thomas More, were among his disciples. He was the friend of Bucer and Melanchthon, and the teacher of the mediating Catholic theologians A. Pighius and J. Gropper in their endeavours to bridge the gulf, before the Council of Trent. He had enemies among the Protestants, and among them must be counted Luther. But he had always Catholic opponents in the Low Countries, Spain and Rome, who after his death placed his works on the *Index Librorum Prohibitorum*. Yet he had real influence in Catholic piety, especially in Italy and Spain.

See also references under "Erasmus, Desiderius," in the Index volume.

BIBLIOGRAPHY.—The following complete editions of Erasmus' works have appeared: ed. by Beatus Rhenanus, with a sketch of his life, 9 vol. (1540–42); by J. LeClerc, 10 vol. (1703–06). His complete correspondence, ed. by P. S. Allen, H. M. Allen and H. W. Garrod, appeared in 11 vol. between 1906 and 1947 and an English translation of his letters up to 1517 was made by F. M. Nichols, 3 vol. (1901–18). Important editions of separate works include the *Colloquies* by N. Bailey (1725) and H. de Vocht (1928), the *Inquisitio de fide* by C. R. Thompson (1950), the early *Opuscula* by W. K. Ferguson (1933) and vol. i of selected works ed. by H. Holburn.

See also J. A. Froude, *Life and Letters of Erasmus* (1894); P. S. Allen, *Erasmus* (1934) and *The Age of Erasmus* (1914); A. Renaudet, *Études Érasmiennes* (1939), *Érasme et l'Italie* (1955), *Humanisme et Renaissance* (1958); A. Hyma, *The Youth of Erasmus* (1930) and *Renaissance to Reformation* (1951); P. Mestwerdt, *Die Anfänge des Erasmus* (1917); K. A. Meissinger, *Erasmus* (1948); J. Huizinga, *Erasmus* (1924; Eng. trans. by F. Hopman, 1924), perhaps the finest single essay; M. M. Philips, *Erasmus and the Northern Renaissance* (1949); J. B. Pineau, *Érasme, sa pensée religieuse* (1924); E. V. Telle, *Érasme de Rotterdam et le septième sacrement* (1954); A. Flittner, *Erasmus im Urteil seiner Nachwelt* (1952); M. Battaillon, *Érasme et l'Espagne* (1937). For bibliography see *Bibliotheca Erasmiana* published by Ghent university library since 1897. (E. G. Ru.)

ERASTUS, THOMAS (1524–1583), Swiss physician and religious controversialist, whose name is preserved in the word "Erastianism," a theory of the relationship between church and state which he himself never taught. Born on Sept. 7, 1524, at Baden, Aargau, his original surname being Lüber, he studied philosophy and medicine at Basel, Bologna and Padua. In 1557 he was invited by the elector Otto Heinrich of the Palatinate to become professor of therapeutics in the new faculty of medicine at Heidelberg university, where he rapidly acquired a considerable reputation as a physician and teacher.

Erastus owes his posthumous reputation to the part which he played in the ecclesiastical affairs of the Palatinate during the electorate of Frederick III (1559–76). As a Zwinglian, he was closely associated with the introduction of Reformed Protestantism at the beginning of the new reign. He was made a member of the church council (*Kirchenrath*) of the electorate in 1559 and attended the sacramentarian conferences of Heidelberg (1560) and Maulbronn (1564) where he supported the Zwinglian doctrine of the Eucharist. When the Calvinist party in the Palatinate, led by Caspar Olevianus, attempted to secure the adoption of the Genevan system of church discipline, Erastus was its leading opponent. The controversy came to a head in 1568 over the theses presented for the doctorate of theology at Heidelberg by an English Puritan, George Withers, who affirmed the necessity of excommunication and of presbyterian church order. To refute Withers, Erastus drew up a series of 100 theses (afterward reduced to 75) against excommunication. The Calvinist party, however, had the support of the elector and in 1570 the presbyterian system was established by electoral decree. Erastus continued to oppose the new order and was consequently excommunicated for two years though allowed to retain his professorship. After the restoration of Lutheranism under Louis VI (1576–83), Erastus was forced to leave Heidelberg. He returned to Basel in 1580 and was appointed to a chair of ethics (1582). He died there on Dec. 31, 1583.

Erastus' most famous work is his treatise on excommunication, *Explicatio gravissimae quaestionis*, etc., published posthumously in 1589. In addition to the 75 theses, the volume contains a defense of the theses in six books replying to the criticisms made by Theodore Beza, Zacharias Ursinus and Hieronimus Zanchius, and an appendix of letters to Erastus from the Zürich theologians Heinrich Bullinger and Rudolph Gualther, written during the controversy. Erastus, who is mainly concerned in this work with the problem of excommunication, and not with the question of church and state, argues that the practice of excommunication is unscriptural; that the sacraments should not be withheld from anyone who genuinely wishes to receive them; and that in a Christian society the punishment of sin is a matter for the magistrate alone. Erastus was not an "Erastian" in the modern sense of the term, for, though he considered that the magistrate was responsible for the external government of the church, he explicitly limited his argument to Christian rulers and nowhere suggested that the interests of religion are subordinate to those of the state. The term "Erastian" first seems to have come into use in England during the debates in the Westminster assembly (1643), when it was used as a term of abuse by the Presbyterians for the party of John Selden, John Lightfoot and Bulstrode Whitelocke. The theses were twice translated into English in the 17th century; in 1659, as *The Nullity of Church Censures;* and in 1682, as *A Treatise of Excommunication*. A new translation by Robert Lee was published in 1844. An answer to the *Explicatio* by Theodore Beza was published in 1590 under the title *Tractatus . . . de vera excommunicatione*.

Erastus also wrote a number of medical and scientific treatises in which he attacked such popular superstitions as the belief in astrology and the transmutation of metals; yet he shared the contemporary belief in witchcraft and his *Repetitio disputationis de lamiis seu strigibus* (1578) is a defense of the use of the death penalty against witches and sorcerers.

BIBLIOGRAPHY.—A. Bonnard, *Thomas Éraste et la discipline ecclesiastique* (1894); R. Wesel-Roth, *Thomas Erastus* (1954); J. N. Figgis, "Erastus and Erastianism" in *The Divine Right of Kings*, 2nd ed., pp. 293–342 (1914). (W. D. J. C. T.)

ERATOSTHENES OF ALEXANDRIA (*c.* 276–*c.* 194 or 192 B.C.), Greek scientific writer who measured the circumference of the earth, was born at Cyrene. He studied grammar at Alexandria, philosophy at Athens and settled in Alexandria as librarian of the museum where he wrote on astronomy, geography, mathematics and other subjects. Being informed that at noon at the summer solstice the sun lit up a well at Syene (now Aswan) throughout its depth (*i.e.*, was directly overhead), Eratosthenes measured the sun's zenith distance at noon at the summer solstice at Alexandria and also the distance between Alexandria and Syene. This gave the earth's circumference. For a fuller explanation of the method *see* GEODESY: *Historical*. The exact size of the units (stadia) he used is doubtful but it is supposed that 10 stadia equaled 1 mi., giving the size of the earth within 50 mi. of the present estimate. He also measured the obliquity of the ecliptic by observing the difference in the altitude of the sun at the summer and winter solstices. It is believed that Eratosthenes compiled a catalogue of 675 stars. In his *Geographica* he laid the foundations of mathematical geography.

For Eratosthenes' mathematical work we have mainly to rely on the reports of Pappus. It appears that he wrote on the problem of the duplication of the cube. This dealt with the problem of finding the mean proportionals between two given lines, and Eratosthenes invented an ingenious mechanical device for solving this. His other work concerned prime numbers. His method was

to write down the odd numbers from three upward, to strike out every third number (multiples of three), every fifth number (multiples of five) and so on, the remaining numbers being, of course, primes. This system is often referred to as the "sieve of Eratosthenes."

Eratosthenes also worked out a calendar using "leap" years and endeavoured to fix the dates of literary and political events since the battle of Troy. He wrote a poem inspired by astronomy and also works on the theatre and ethics. Toward the end of his life he suffered from ophthalmia, became blind and is said to have committed suicide by voluntary starvation.

See also references under "Eratosthenes of Alexandria" in the Index volume.

See J. Delambre, *Histoire de l'Astronomie Ancienne*, 2 vol. (1817); G. Bernhardy, *Eratosthenica* (1822); E. H. Berger, *Die geographischen Fragmente des Eratosthenes* (1880). (CN. A. R.)

ERBIUM, a metallic element of the rare-earth group, has very few commercial uses. It has been employed as a constituent of infrared-absorbing glass and as an activator in certain phosphorescent materials.

The element was discovered in 1843 by C. G. Mosander, who originally called it terbia. Due to the confusion arising from the similarity in the properties of the rare-earth elements, the name was changed to erbium about 1860. J. C. G. Marignac prepared the first reasonably pure fraction in 1878. The element occurs in many rare-earth minerals; the more important are gadolinite, euxenite, xenotime, and samarskite. The name erbium is derived from the name of the town of Ytterby in Sweden where gadolinite was found. Erbium occurs also in the products of atomic fission. Before the chemical developments brought on by World War II, the element was purified by fractional crystallization, etc., but subsequently was produced commercially by ion-exchange techniques.

Erbium has the chemical symbol Er; its atomic number is 68 and its atomic weight is 167.27. The naturally occurring stable isotopes and their percent abundance are: Er^{162} (0.136%), Er^{164} (1.56%), Er^{166} (33.41%), Er^{167} (22.94%), Er^{168} (27.07%), Er^{170} (14.88%). The thermal neutron absorption cross section is 166 barns.

Erbium behaves as a typical trivalent rare-earth element and forms a series of pink salts. The ionic radius is 0.881 Å. The trivalent ion in solution is pink and has a number of discrete sharp absorption bands in the ultraviolet and visible spectral regions which are useful for quantitative analysis. The Er^{3+} ion is strongly paramagnetic due to the presence of unpaired electrons.

The metal has been prepared by thermoreduction of the anhydrous fluoride with calcium. It is grayish-silver in colour and reactive; it is oxidized slowly by oxygen or water. The metal melts in the range 1.475°–1.525° C.; its density is 9.058 g. per cubic centimetre for the hexagonal close-packed form. At low temperatures, the element is antiferromagnetic and, at very low temperatures, ferromagnetic. Erbium is also a superconductor at extremely low temperatures. *See* RARE EARTHS. (LD. B. A.)

ERCILLA Y ZÚÑIGA, ALONSO DE (1533–1594), Spanish soldier and poet, was the author of *La Araucana*, the most celebrated of Spanish Renaissance epic poems. He was born in Madrid, Aug. 7, 1533, went to Chile (1555) where he distinguished himself in the wars against the Araucanian Indians and began the poem based on his experiences, returned to Spain (1563), and died in Madrid, Nov. 29, 1594. *La Araucana* appeared in three instalments in 1569, 1578 and 1589 and consisted of 37 cantos; two further cantos were added in 1590. The poem is weak in respect of form, unity, relevance and versification, but it excels in graphic descriptions, particularly of fighting.

See editions, 5 vol. (1910–18), 2 vol. (1933) and Eng. trans. by W. Owen (1945–); also J. T. Medina, *Vida de Ercilla* (1948).

ERCKMANN-CHATRIAN, the combined name under which two authors, the first French writers of the "regional novel," published their work. ÉMILE ERCKMANN (1822–1899) was born at Phalsbourg, May 20, 1822, and died at Lunéville, March 14, 1899; and ALEXANDRE CHATRIAN (1826–1890) was born at Soldatenthal (Meurthe), Dec. 18, 1826, and died at Villemomble, near Paris, Sept. 3, 1890. Close friends, they decided to collaborate in writing novels, essentially patriotic and popular in character. They chose as their heroes the people of their native province, Alsace-Lorraine, and based their plots on events in its history. They began their joint career in 1847 with stories later published as *Contes fantastiques* and *Contes de la Montagne*, but it was with *L'Illustre Docteur Mathéus* (1859) and *Le Fou Yégof* (1862) that they achieved fame, increased by *Madame Thérèse* (1863), the moving story of a vivandière left for dead and later saved by a doctor, and finally confirmed by the success of *L'Ami Fritz* (1864). Later works included *L'Histoire d'un Conscrit de 1813* (1864; Eng. trans. 1909) and *Waterloo* (1865; Eng. trans. 1909). They adapted some of their novels for the stage, and *Le Juif polonais* (produced in 1869) was a study in remorse resembling Dostoevski's *Crime and Punishment* which had appeared three years earlier but which was unknown to them. They quarreled in 1889, abandoned their partnership and took their dispute, allegedly the result of a slander, to the law courts. Anti-imperialist and pacifist, they depicted the horrors of war with the same vividness and truth to life that they used in describing the province where they were born.

BIBLIOGRAPHY.—J. A. Barbey d'Aurevilly, *Les Oeuvres et les hommes* (1860–90), vol. 4; Émile Zola, *Les Romanciers naturalistes* (1881); A. Lamartine, *Cours familier de littérature* (1856–69), vol. 23; L. Schoumacker, *Erckmann-Chatrian* (1933). (R. DL.)

EREBUS, in Greek literature, the "darkness" of the western region where lie Hades and the sunless Cimmerian land. In Hesiod's *Theogony* Erebus is personified as the son of Chaos, brother and husband of Night, father of Aether ("upper air") and Day.

ERECH (Sumerian URUK; Greek ORCHOË; modern WARKA), an ancient Mesopotamian city about 35 mi. N.W. of Ur (Tall al Muqayyar) in the Nasiriyah district of modern Iraq. Warka, which lies within the Euphrates valley, has been excavated from 1912 onward by a series of expeditions under the auspices of the Deutsche Orient Gesellschaft and the German Archaeological institute. Uruk, which was one of the greatest of the cities of Sumer, was enclosed by defensive walls of plano-convex brickwork about 6 mi. in circumference, which according to legend were built by the mythological hero Gilgamesh. Within this circuit there are abundant traces of successive cities which date from the prehistoric 'Ubaid period, perhaps before 4000 B.C., through four millenniums down to Parthian times. Urban life in what is known as the Uruk-Jamdat Nasr period, about 3500–3000 B.C., is more fully illustrated here than in any other Mesopotamian city. Writing was then invented and developed from a pictographic into a cuneiform script. Some hundreds of primitive clay tablets, mostly economic texts, were found and show that at this period business activity was controlled by the temples and that agriculture, stock farming and metallurgy were being actively pursued.

One of the oldest sacred areas of the city is that known as Bit Resh, of which the chief landmark is the "Anu" ziggurat crowned by the "white temple" of the Jamdat Nasr period, probably built shortly before 3000 B.C. A series of much earlier temples now buried within the core of the ziggurat goes back to the preceding Uruk period. The façades of these temples were decorated with a polychrome mosaic composed of thousands of terra-cotta cones or pegs stuck into the wall plaster. The two principal Sumerian divinities worshiped in Uruk appear to have been An (Anu), a sky-god, and the goddess Inanna, predecessor of Ishtar. Many early temples and Sumerian remains were also found in and around a second ziggurat known as Eanna.

An unusual feature of this and the subsequent Early Dynastic period is a series of ritual deposits containing animal bones. One of them, the so-called *Riemchengebäude*, was a sunken mud-brick enclosure with central chamber and sealed doorway surrounded by passages which contained hundreds of objects, pottery, stone vessels and wooden furniture with mosaic incrustation; the entire inner chamber with its contents had been ritually burned.

The Uruk-Jamdat Nasr period was one of great prosperity. Gold, silver and copper were skilfully worked; seals and amulets, especially those representing animals, reflected a brilliant miniature craftsmanship. The life-size head of a goddess, a realistic carving

in white marble, was far in advance of the later Early Dynastic sculpture. Another remarkable monument is an alabaster vase over three feet high which illustrates in relief the spring festival, a scene at which the ruler presents the first fruits and all manner of agricultural produce to the presiding goddess.

The temenos (sacred enclosure) of Eanna, rebuilt and repaired continuously until Parthian times, bore witness to the attention of many powerful kings, including Lugalzaggesi (*c.* 2370 B.C.) and Ur-Nammu, first king of the 3rd dynasty of Ur, who was a scion of Uruk. Ur-Nammu did much for the layout of the city which then benefited from a neo-Sumerian revival. Various architectural developments were associated with the Isin-Larsa period, about 2030–*c.* 1770 B.C. Subsequently the Kassites established themselves in the city and one of their kings, Kara-Indash, 15th century B.C., built a unique temple, the façade of which was ornamented with burnt-brick reliefs depicting divinities carrying vases overflowing with the waters of the Tigris and Euphrates; the style reflects Elamite influence. After the Kassites, Assyrian, neo-Babylonian and Achaemenian rulers including Cyrus the Great and Darius the Great left the imprint of their architectural activity, especially in the district of Eanna.

In Seleucid times the city continued to prosper. To this period belong *bit-akitu* (spring-festival house) and a gigantic temple to the sky-god An and his consort Antum. A temple dedicated to the god Gareus, houses and graves prove that the city was well populated in Parthian times when the last of a long-founded school of learned scribes was still editing documents (*c.* 70 B.C.) in the cuneiform script. *See* also BABYLONIA AND ASSYRIA.

BIBLIOGRAPHY.—*Uruk Vorläufiger Bericht* (1930 *et seq.*), of which 17 vol. had appeared by 1961; A. Parrot, *Archéologie Mésopotamienne* (1946); E. Heinrich, *Kleinfunde aus den archaischen Tempelschichten in Uruk* (1936); H. J. Lenzen, *Die Entwicklung der Zikurrat* (1941); A. Falkenstein, *Topographie von Uruk* (1941) and *Archaische Texte aus Uruk* (1936). (M. E. L. M.)

ERECHTHEUM, an Ionic temple of Athena on the Acropolis at Athens. Even when new it was officially known as the archaic or venerable temple because, like its immediate predecessor the Doric Peisistratid temple, it inherited the sanctity of the primitive temple of Athena mentioned in the *Iliad* and *Odyssey*. It was also known as the temple of Athena Polias. The name Erechtheum was doubtless of later popular origin, derived from a shrine dedicated to the legendary king Erechtheus in its rear portion. The name was employed to distinguish it from the neighbouring Doric temple of Athena, popularly known from the time of Demosthenes as the Parthenon (*q.v.*).

History.—The date of the beginning of the temple is not recorded, but was undoubtedly just after the Peace of Nicias, 421 B.C. It was left incomplete at the time of the expedition to Sicily in 415, and in the summer of 409 B.C. a commission was appointed to report on its condition and carry it to completion. The work progressed steadily until 405 B.C. according to fragmentary records inscribed on marble. These records, consisting of the report of 409 and expense accounts of 409–405 B.C., are of immense value not only for the history of the temple, the names and social status of the workmen, and the prices of labour and materials, but also for information regarding the restoration of missing parts. The name of the original architect is lost, though we may, for reasons given below, infer that it was Mnesicles, the architect of the Propylaea (*q.v.*). The later architects were appointed annually, Philocles in 409 and Archilochus in 408.

After a fire in the Roman period, during the reign of Augustus, the upper part of the west end was reconstructed with the addition of the present windows, and the great north doorway was repaired. During the middle ages it served as a church of the Virgin, and after the Turkish occupation in 1458 as a private house (not the commandant's harem, as often stated). Lord Elgin carried away to London, 1801–03, the northeast corner column and the anta capital behind it, some blocks of the wall crown, architrave and cornice, and one of the caryatids. The caryatid was replaced by a terra-cotta cast in 1846, renewed in portland cement in 1912. The building was much injured during the Turkish siege of the Acropolis in 1827 and again, after a partial reconstruction between 1837 and 1847, when the upper part of the west wall fell during a storm in 1852. Between 1902 and 1909 the temple was scientifically restored, so far as the original blocks permitted; and at the same time Gorham P. Stevens made the stone-by-stone examination which eventually resulted in the definitive publication by the American School of Classical Studies at Athens (1927).

Architectural Features.—The Erechtheum is remarkable for its complexity. It has three porches and four inner rooms, all at different levels, and four different orders: three types of Ionic columns and one of caryatid figures. As originally built, it included, at the highest level, an eastern portico of six columns and a cella with two windows flanking the doorway, dedicated to Athena. This part was separated by a marble cross-wall from three western rooms at lower levels and did not communicate with them. One of these rooms, the western vestibule, was entered from the north flank by a doorway of great richness under a portico (also of six columns, but arranged with four on the north front and an extra column on either side). A smaller door and an L-shaped flight of steps gave access toward the south to the famous caryatid porch, at a level as high as that of the east portico. A third door, in the west wall, gave access to the sacred precinct at the west known as the Pandroseum (approached also by a lateral door in the north portico). Two other doorways, in the east wall of the vestibule, increasing the number to five in this narrow room, gave access to two central rooms, side by side, with floors at slightly different levels; these were closed by pairs of white marble doors with black stone let into the stiles and rails (unique except in tombs), according to the report of the commission. The T-shaped marble partition separating these three rooms rose only 12 ft. 9 in. above the vestibule pavement (*i.e.*, to the bases of the engaged columns decorating the west wall). This height permitted the wooden ceiling high above, supported by a heavy cross girder with diagonal struts resting on marble consoles in the flank walls, to be seen from all three rooms. In these western rooms, according to Pausanias, were an altar dedicated to Poseidon and Erechtheus, another of the hero Butes, and a third of Hephaestus, as well as paintings of the family of the Butadae.

Among the peculiarities of the temple may be cited the huge lintel which carries the entire southwest corner over what was obviously the subterranean tomb chamber of the legendary king Cecrops, founder of Athens, and the recessed niche which diminishes the weight of the wall above it, with a projecting marble

ALISON FRANTZ

NORTH PORCH OF THE ERECHTHEUM

floor precariously balanced on two adjacent edges. Also in the western vestibule was evidently the salt spring and the mark of Poseidon's trident in the rock, now obliterated by a medieval rock-cut cistern. In the north portico a square hole in the floor opened upon a crypt with deep depressions which may be identified as marks of the thunderbolt of Zeus; and immediately above is a similar hole through the marble ceiling and roof. This was probably the altar of Zeus Hypatos described by Pausanias outside the (north) entrance; the stones of the hollow altar are described in the report of 409 B.C.

The fame of the Erechtheum depends largely upon the exquisite perfection of its details. The Ionic capitals are the most beautiful that Greece produced, with extra moldings subdividing the volutes, a torus carved with the guilloche inserted above the egg-and-dart of the ovolo, and a band of anthemion decoration below (*see* MOLDING). In the north portico the capitals were further enriched by glass beads, red, blue, yellow and purple, inserted in the guilloche. The tops of the flutes were enframed by a delicate astragal, probably gilded, like a pendant fringe; a gilded bronze stem ending in a palmette followed the convolutions of each volute. The eyes of the volutes were also gilded, according to the expense accounts. Besides the anthemion bands on the capitals there were others on the walls and antae, nine varieties in all. The upper torus of the bases, fluted elsewhere in the temple, is carved with a guilloche in the north portico, different in the corner columns (and antae) from the intermediate columns. Traces of painted patterns appear on the smooth moldings, and the carving also must have been picked out in paint, though the colours are now indeterminable. The frieze, around the main building as well as the north portico, was polychromatic in another sense; the background was of blue-black Eleusinian stone, against which were fixed white marble figures in relief, many of them preserved and others described in the expense accounts.

The Original Plan.—The irregular plan caused Vitruvius (iv, 8) to include the Erechtheum in his list of temples either unsymmetrical or with "what is regularly found on the front transferred to the sides (added at right and left against the shoulders)." This irregularity, as well as the unstudied manner in which the west wall awkwardly meets projecting portions of the north portico and the caryatid porch, suggest that the original architect was prevented from completing a more perfectly organized plan. That changes occurred during erection is shown by structural evidence: the outer west wall was thrust inward 26 inches, the inner wall of the western vestibule by half as much, and at the southwest corner occurred an unforeseen collision with the tomb of Cecrops. An attempt to divine the architect's intention was made by W. Dörpfeld, but with less success than in the case of the Propylaea (*q.v.*). Preferably, by working backward from the existing remains, the evidence suggests that the temple was first planned symmetrically, to stand on the site of the ruined Peisistratid temple just to the south, with four rooms all at one level and with porticoes, each of four Ionic columns, at east and west. Then, in view of a decision to retain the western half of the old temple as a treasury (the Opisthodomus), the Erechtheum was thrust northward to a site encumbered by religious tokens on different terrace levels. Consequently the pavement level of the west portico had to be located 9 ft. 10½ in. lower than that on the east, which remained at the originally proposed level; and, to avoid such a difference in the roof levels, the east columns were made shorter and consequently more closely spaced, six instead of four. The larger west columns were likewise increased to six, the two extra columns being placed on the sides of the portico. But the west portico, as thus located, would have covered the sacred olive tree, and so was rotated to the north flank, placed against the shoulder of the main building. To counterbalance it, the caryatid porch was added as an afterthought against the shoulder on the south flank, likewise with six supports, four in front and one more on each side. But since this was at the higher level, overlapping the platform of the old temple, and since the flat roof had to be kept lower than the wall of the main building, a much smaller order was devised, with caryatids standing on a high parapet and supporting a compressed entablature of Asiatic Ionic form. While the north portico, its doorway and the small door to the caryatid porch, were aligned with respect to a transverse axis, the eastward shifting of the two west walls by different amounts dislocated the axis of the western vestibule by 19½ inches, and required a narrowing of the caryatid porch. The upper part of the denuded west wall then received a decorative series of smaller Ionic engaged columns, with grilled openings between them (replaced in Roman times by windows). The frank adaptation of the revised design to the varying ground levels, the unorthodox solutions of new problems in a manner paralleled only in the Propylaea, and the similarity of some details suggest that the original architect may have been Mnesicles, whose task of landscaping the Acropolis may well have continued until this time. *See* also references under "Erechtheum" in the Index volume.

BIBLIOGRAPHY.—W. Dörpfeld, "Der ursprüngliche Plan des Erechtheion" in *Mitteil. Athen.* (1904); G. P. Stevens and J. M. Paton, *The Erechtheum, Measured, Drawn, and Restored* (1927); W. B. Dinsmoor, "The Burning of the Opisthodomos at Athens" in *Am. Journ. Archaeology* (1932); W. Dörpfeld and H. Schleif, *Erechtheion* (1942).

(W. B. D.)

ERECHTHEUS, legendary king (and probably also a divinity) of Athens, "whom once Athena reared, after the fertile corn-land bore him; she established him in Athens in her rich temple. There the sons of the Athenians propitiate him with bulls and rams, as the years come round" (*Iliad*, 2, 546 ff.). In later times no one was thought to share the temple with Athena except a great snake, and there is good evidence that Erechtheus was, or became, a snake; that is, an earth or ancestor spirit.

The earliest Athenian kings tended to have similar names suggesting a connection with the earth (*chthon;* Erechtheus, Erichthonius, Erysichthon), to be born of the earth (Erechtheus, Erichthonius, Cecrops), to be raised by Athena (Erechtheus, Erichthonius) and to have something serpentine about them (Erichthonius, Cecrops). Erichthonius, for example, reportedly was born when Hephaestus, trying to make love to Athena, spilled his seed on the ground. Athena gave the child to the daughters of Cecrops in a box; they, opening it against orders, were driven mad by the sight of its snakelike contents and leaped from the Acropolis. Snakes are often earth or ancestor spirits, so that Athena's sharing her temple with Erechtheus, whom she has herself nurtured, may well be the mythical way of expressing her guardianship of the ancient royal house of Athens, and of the land itself and its fertility, with which ancient kingship was intimately connected. A line of the *Odyssey* (7, 81) seems to recall a time when Athena lived in Erechtheus' house rather than the other way round. This Erechtheum may be the Mycenaean palace which once occupied the Acropolis; if so, the present Erechtheum commemorates it.

In his lost play *Erechtheus* Euripides gave this king three daughters, one of whom was appropriately named Chthonia. At war with neighbouring Eleusis and its ally King Eumolpus, Erechtheus learned from Apollo that Athens would win if he sacrificed his daughter. He sacrificed Chthonia and won, but apparently got little good by it, for he was destroyed by Poseidon, or some say by a thunderbolt from Zeus.

BIBLIOGRAPHY.—H. J. Rose, *A Handbook of Greek Mythology* (1929); Edith Hamilton, *Mythology* (1940); *Oxford Classical Dictionary* (1949).

(G. E. DK.)

ERESHKIGAL, a goddess in the Sumero-Akkadian pantheon. She is "lady of the great place" (*i.e.*, the abode of the dead) and, in texts of the 3rd millennium B.C., wife of the god Ninazu (elsewhere accounted her son), in later texts, of Nergal (*q.v.*). Her name occurs in pre-Sargonic documents. During the 3rd dynasty of Ur, early in the 2nd millennium B.C., she appears with the name Allatum on tablets which record offerings to deities at Nippur and Ur.

Ereshkigal's sister was Inanna (Ishtar), and between the two there was great enmity. In the Sumerian version of the latter's descent to the underworld, Ereshkigal is seated on her throne while the seven judges give judgment against Inanna. In the rendezvous of the dead, Ereshkigal reigns in her palace, ever on the watch for lawbreakers and on guard over the fount of life lest any of her subjects take of it and so escape her rule. Her offspring and servant is Namtar, the evil demon, Death. Her power extends to earth

where, in magical ceremony, she liberates the sick possessed of evil spirits. A ritual text relating to her worship at Uruk (Erech) lays down that the flesh neither of bulls nor of fowl may be offered to her.

Two Hittite treaty texts, a damaged text found in Egypt at Tell el Amarna, evidence from the kingdom of Elam east of Babylonia and the existence of the south Arabian deity Allat (*i.e.*, Allatum) show the wide extent of Ereshkigal's cult. In Mesopotamia the chief temple known to be dedicated to her was at Cuthah. From Egypt has come a magical Greek text of the 3rd century B.C., in which the goddess is named between Persephone and Adonis. *See* also Babylonia and Assyria: *Religion.* (T. Fh.)

ERETRIA, an ancient coast town of the island of Euboea, Greece, about 15 mi. S.E. of Chalcis by road, opposite to Oropus, now Nea Psara. Eretria, like its neighbour Chalcis (*q.v.*), founded colonies in the west and north of Greece. It also acquired dependencies among the Cyclades and had close alliance with Miletus. Defeated by Chalcis in the Lelantine War (*c.* 700 B.C.), it lost influence in the west but remained important through the archaic period. Its interference in the Ionian revolt (499 or 498) brought upon it the vengeance of the Persians, who destroyed it shortly before the battle of Marathon (490). The city was soon rebuilt, and was a member of both Athenian leagues. The Athenians rescued it from the tyrants suborned by Philip II of Macedonia in 341. Under Macedonian and Roman rule it fell into insignificance.

The modern village, which is called Nea Psara because the inhabitants of Psara, off Chios, were transferred there in 1821, is on unhealthy low-lying ground near the sea. The ancient site was excavated by the American School of Classical Studies at Athens (1890–95). At the foot of the acropolis lies the theatre, where only seven rows of seats remain, but much of the stage is preserved. Nearby are a temple of Dionysus, a large altar and a gymnasium. In 1900 a temple of Apollo Daphnephoros was found. The wall of the acropolis remains to the height of eight courses; other walls to north and south connect it with the harbour and mole.

Bibliography.—Strabo, *Geography,* x, 447; *Papers of the American School of Classical Studies at Athens,* vol. vi (1897).

EREVAN: *see* Yerevan.

ERFURT, a city of Germany, which after partition of the nation following World War II was included in the *Land* (state) of Thuringia, and after 1952 was capital of the *Bezirk* (district) of Erfurt, one of three into which Thuringia was divided, in the German Democratic Republic. It lies on the Gera river, on the railway and the Autobahn from Berlin to Frankfurt am Main. Pop. (1958 est.) 186,066.

Round the old part of the town is a wide ring of residential and industrial districts. In the west Erfurt is dominated by the Petersberg hill and the Cyriaksburg castle. The cathedral and the church of St. Severus stand side by side on top of an eminence called the Domberg. The cathedral, begun in the 12th century and rebuilt in Gothic style in the 13th, contains 15th-century glass and numerous works of art. Among the city's many interesting buildings are the Scottish church (*c.* 1200); a Dominican monastery (13th century); the Augustinian monastery where Martin Luther was a monk for five years, now used as an orphanage; the Krämer bridge with houses and shops built on both sides (1325); the Anger museum, a former merchant's house (1705–11), containing a picture gallery, glass, faïence and exhibitions of local history and prehistory; the house of the governor of the electorate of Mainz (1711–20); the Teaching institute (1953–59); and the Medical academy (1954–59). There is a scientific library which contains the Amplonian collection of 1400 and earlier; a library of municipal archives; a natural history museum; the Thuringian Social History museum and an aquarium. Among the many open spaces are the municipal park and the Roter Berg with a zoological garden.

Erfurt is an important road and railway junction and a commercial centre. There is an airport 3 mi. to the north. The chief industries are metalworking and the manufacture of electrical apparatus, typewriters, shoes, clothing and food. Erfurt is known for the cultivation of a wide variety of flowers and vegetables—an industry that originated in the large gardens attached to the monasteries. Seeds and manufactured foods are exported; metals, leather and fabrics are imported.

The town was first mentioned in documents as Erpesfurt. St. Boniface (*q.v.*) founded a bishopric there in 742. In 805 Erfurt was a military strong point and an administrative and commercial centre on the eastern border of the Frankish empire. About 1000 the town came into the possession of the archbishop of Mainz who, about 1250, granted it municipal rights. In the middle ages almost 100 villages and country houses belonged to Erfurt. Early in the 15th century it joined the Hanseatic league and until 1600 it was a great commercial centre for woad, a plant used in the middle ages for its blue dye extract. In 1392 the university, the first in Europe to embrace the four faculties, was opened; it was suppressed in 1816. During the Thirty Years' War the town was for years a Swedish garrison and in 1664 it was occupied by the troops of the archbishop of Mainz, after which it became part of the electorate of Mainz. In 1802 Erfurt came into the possession of Prussia and from 1806 to 1813 it was under French domination. In 1814 it was returned to Prussia. In 1808 Napoleon and the emperor Alexander I of Russia, with the kings of Bavaria, Saxony, Westphalia and Württemberg met at the congress of Erfurt, and there in 1850 a conference took place between members of the short-lived Prussian union (*see* Germany: *History*). (F. Wd.)

ERGOT, a fungal disease of grasses, especially rye, induces the transformation of the grain into enlarged, hard, beanlike structures (sclerotia) that constitute the drug ergot.

The Plant Disease.—The causal agent, *Claviceps purpurea,* is an ascomycete or sac fungus. In the ear of rye infected with ergot there is exuded a sweet, yellowish mucus, which after a time disappears. The ear loses its starch and ceases to grow, and its ovaries become permeated by the mycelium of the fungus which forms the sclerotium in autumn. (*See* also Fungi: *Ascomycetes* [*Sac Fungi*]; Plant Diseases: *Historical Background.*)

The Drug.—The drug, almost exclusively obtained from infected rye, consists of curved, dark-purple grains ½ to 1½ in. long and 1 to 4 lines broad, having two lateral furrows, a rancid taste and a fishy odour, intensified by potash solution.

The active principles of ergot are a number of alkaloids that have lysergic acid as a basic nucleus. The most important of these alkaloids are ergonovine and ergotamine. Other ingredients include a fixed oil, a glucoside, and trimethylamine, which gives the drug its unpleasant odour.

Ergot has no external action. When it is given internally as the crude drug or its isolated alkaloids, its most important pharmacologic action is limited to the unstriped muscle of the pregnant uterus. The ergot alkaloids are the most powerful stimulants of the pregnant uterus, inducing muscular contractions and increased muscle tone. It is this tonic contraction of the uterus that limits its therapeutic usefulness as an oxytocic drug to the placental stage of labour and the period immediately after delivery. Ergonovine is used widely in obstetrical practice in the management of the placental stage in the prevention of excessive blood loss. This alkaloid has been used for the relief of migraine headaches with fairly good results, but its pharmacologic action in this regard is not known.

Chronic poisoning or ergotism occurred frequently among peoples fed on rye infected with *Claviceps.* There were two types of poisoning. In the gangrenous form various parts of the body developed cessation of circulation with resultant gangrene. In the spasmodic form the symptoms were nervous. The disease began with itching and tingling of the skin and a loss of cutaneous sensations, which began in the extremities. Amblyopia, deafness and psychotic disorders also occurred. Death usually resulted from involvement of respiratory muscles. The last-known epidemic of ergotism occurred in Lorraine and Burgundy in 1816, but sporadic cases still occur.

Ergot has been produced in radioactive form. The constituents of ergot, broken down by chemical separation, retain their radioactivity. These tagged alkaloids and other substances promise to provide new knowledge concerning the pharmacology and usefulness of an old drug.

See also Obstetrics. (M. E. Ds.; X.)

ERIC V GLIPPING (*c.* 1249–1286), Danish king, son of Christopher I (d. 1259) and Margaret Sambiria of Pomerania (d. 1282). Eric ascended the throne in May 1259, at the height of an ecclesiastical dispute in which his father had imprisoned Archbishop Jacob Erlandsen and Denmark had been laid under a formal but ineffective interdict. Queen Margaret released Erlandsen, but he would not crown Eric; an assembly of bishops eventually did so, at Viborg on Dec. 25, 1259. In July 1261 the royal army was defeated by a coalition, organized by Erlandsen from Schleswig and Holstein, in favour of Eric's cousin, Eric, duke of Schleswig. Both the king and queen mother were captured. The papacy and the German princes intervened, however, and the royal prisoners were restored. Declared of age in 1266, Eric soon began to rule autocratically. After conflicting decisions and lengthy negotiations the king, in 1272, achieved a favourable settlement with the papacy and the archbishop; before it could be implemented the archbishop died, but in the same year (1274) the interdict was lifted.

With the death of his cousin in 1272 Eric had gained control of Schleswig and in 1276, despite opposition from the magnates, he had his two-year-old son, later Eric VI Menved, recognized as his successor. In 1282, however, he was forced to grant Denmark's first royal charter. This protected the lords from arbitrary imprisonment and instituted a yearly meeting (*parlamentum*) of the magnates. There was a complete change of government and of foreign policy, and extensive reforms. Discontented members of the royal family received crown lands. All this stopped suddenly on Nov. 22, 1286, when Eric was found murdered at Finderup, Jutland, in mysterious circumstances (*see* ERIC VI).

See A. E. Christensen, *Kongemagt og Aristokrati* (1945); H. U. Yrwing, *Kungamordet i Finderup* (1954). (A. E. CN.)

ERIC VI MENVED (1274–1319), Danish king, was the son of Eric V Glipping and Agnes of Brandenburg. After his father's murder in Nov. 1286 Eric was recognized as king. At a parliament at Nyborg at Whitsun 1287, some members of the late king's government accused of the murder, including Count Jacob of Halland and the famous Stig Andersen (Marsk Stig), were outlawed. This judgment, invalidated by modern research, was undoubtedly political and led to prolonged turmoil. The outlaws, abetted by the Norwegian king, harried the Danish coasts and were soon joined by Duke Valdemar of Schleswig and the new archbishop, Jens Grand. Eric reached an agreement with Norway and Schleswig in 1295, but the conflict with Grand developed into a long and bitter ecclesiastical struggle. Grand was imprisoned, but escaped and in 1297 obtained a severe papal pronouncement against the king; in 1302, however, Eric persuaded the pope to mitigate this and remove Grand. From 1301 he resumed Denmark's former imperialistic policies in northern Germany and won formal recognition of his sovereignty from some princes, which the German King Albert I sanctioned in 1304. However there was continual trouble with Norway and Sweden and growing opposition within Denmark from the church, the magnates and the peasants. When Eric died, childless, on Nov. 13, 1319, the country was greatly weakened; estates and provinces were mortgaged to the counts of Holstein and other North German princes, and his own brother and successor Christopher (II) was in rebellion.

See E. Arup, *Danmarks Historie*, vol. ii (1932); I. M. Andersson, *Erik Menved och Venden* (1954). (A. E. CN.)

ERIC XIV (1533–1577), king of Sweden from 1561 to 1569, was born at Stockholm on Dec. 13, 1533, the only son of Gustavus I Vasa by his first wife Catherine of Saxe-Lauenburg. He was crowned king at Uppsala on June 29, 1561. In his will Gustavus I had constituted the sons of his second marriage dukes, with extensive powers, which Eric sought to curb by the articles of Arboga (1561). In the same year he set up a royal court of appeal and a bench of justices. His constitution of Uppsala (1562), strictly enforced, defined the military obligations of the nobility.

To free Swedish trade from Danish domination Eric sought contacts with western Europe (he was a suitor to Elizabeth I of England). In the east he asserted his sovereignty over Reval and parts of Estonia. In July 1561 his half-brother John, duke of Finland, signed a marriage treaty with Poland. In 1562 Eric forbade the marriage and imprisoned the recalcitrant John. In Sept. 1563, Frederick II of Denmark, now allied with Poland, declared war on Sweden and thus began the Seven Years' War of the North. Eric's conduct of the war was ineffective. Always suspecting treason, he moved against the great Sture family in May 1567: its leading members, and other nobles, were killed. Eric personally assisting in the murder of Nils Sture. For this his low-born adviser Jöran Persson was temporarily imprisoned. Eric having lapsed into madness. Soon after his recovery he restored Jöran Persson to favour early in 1568 and in July had his mistress Karin Månsdotter crowned queen. In Sept. 1568 Stockholm yielded to John (released earlier that year), and in Jan. 1569 Eric was deposed. He died in prison at Orbyhus on Feb. 26, 1577, poisoned, it is said, by order of his half-brother, now King John III.

See I. Andersson, *Erik XIV* (1935), *A History of Sweden* (1956).

ERIC OF POMERANIA (*c.* 1381–1459), king of the united realms of Denmark, Norway and Sweden, was son of Duke Vratislav VII of Pomerania and great-nephew of Margaret, queen of the united realms, who adopted him after the death of her son in 1387. On June 17, 1397, he was crowned king at Kalmar. After Margaret's death in 1412 Eric became effective ruler. He aimed to build a powerful Baltic empire but never reconciled the opposing interests of his extensive domains. He pursued Margaret's attempt to recover Schleswig from the Holsteiners, but with dire results: his wars (1416–22 and 1426–35) to enforce judgments in his favour by the Danish diet (1413) and by the German king Sigismund (1424) lost him much of his predecessor's gains. In 1426 the Holsteiners were joined by Lübeck and the Wendish towns, who saw their trading privileges endangered by Eric's encouragement of Scandinavian merchants; *e.g.*, in the anti-Hanseatic trading ordinance of 1422. Worsted in Schleswig, Eric nevertheless fought off Hanseatic fleets in the Sound and, about 1428, imposed the first Sound tolls.

This specifically Danish war policy with its heavy taxation and restrictions on trade made Eric increasingly unpopular, especially in Sweden. When the miners, hit by trade disruption, rose in 1434, the Swedish magnates exploited the strife. At Kalmar in 1436 Eric had to accept a plan for constitutional union put up by the Swedes with Danish support. His attempts to reverse this settlement led to his deposition in 1439 (in Denmark and Sweden; 1442 in Norway). He continued to intrigue from Visborg, Gothland, until 1449. He then withdrew to Pomerania, where he died on May 3, 1459.

See K. Erslev, *Erik af Pommern* (1901); E. Lönnroth, *Sverige och Kalmarunionen* (1934). (A. E. CN.)

ERIC THE RED (ERIC THORVALDSSON) (fl. late 10th century), founder of the earliest Scandinavian settlement in Greenland (985) and father of Leif Ericsson, the first known European discoverer of North America. He left Norway as a child in the family of his exiled father Thorvald. By the Scandinavian style of the time he was known as Eric Thorvaldsson and in his youth was nicknamed Eric the Red. The exiles settled in western Iceland whence, from the mountain tops at the extreme northwest, across 175 mi. of water, could be seen another very high land (Greenland). Later, when Eric was himself exiled, though for a term of only three years, he decided to spend his outlawry period exploring this western land. He headed west in the spring of 981, with about 30 of his family and neighbours and their livestock. They probably traveled in a typical ship of the time, less than 100 ft. long and clinker-built. Eric tried to reach the east coast but was fenced from the plainly visible land by southward-drifting ice. It proved to be impossible to get ashore until the ship had rounded the southern tip and proceeded up the west coast to the modern district of Julianehaab. There the party landed and began to prepare for an Iceland-style winter. During their three years of outlawry they saw no people, but found signs of them along the shores. The settlers explored to the northwest to what they called Bear Island (modern Disko). Judging conditions in this new land to be better for stock raising than in Iceland, they named the entire country Greenland.

Returning to Iceland in 985, Eric publicized Greenland so well

that in 986 a contingent of 25 ships followed him southwesterly. Some were lost and some turned back, so that only 14 won through; they are believed to have landed about 350 colonists, with their livestock and gear, at various points in what was later called the Eastern Colony. To the year 1002 the settlement had no mishap; but then came an epidemic, doubtless brought by one of the immigrant ships, and many died in the community which by that time probably numbered 1,000.

In 999 the middle of Eric's three sons, Leif (pronounced Lave, as in slave), who was also called Leif Ericsson and Leif the Lucky, sailed from Greenland to Norway by way of the Hebrides, rather than on the usual course by way of Iceland. Returning in the summer of 1000 he did not touch the Hebrides, nor any land, but attempted to hit the southern tip of Greenland. He missed it, probably in thick weather, and struck the North American mainland, perhaps near Hamilton Inlet, Labrador, or somewhat farther south. Realizing he was not in Greenland, Leif promptly turned back, reaching his father's home in the southern part of west Greenland before autumn.

Christianity was now becoming a fashionable religion among Scandinavian men; it already had been popular for some decades with women. Leif brought to Greenland the news that the king of Norway, Olaf I Tryggvessön, had turned Christian; and he presented two missionary priests, announcing that the king had sent them. Leif's mother accepted the priests eagerly, saw to it that a church was built for them, and divorced Eric because he refused baptism. The church was soon established in Greenland and in time a bishopric was organized there.

The colony gradually declined and ultimately disappeared. In 1347, however, the Roman Catholic medieval culture of Greenland was still flourishing to the extent that a ship built there was able to reach Iceland on a return voyage from Labrador with a cargo of American timber. This is documented, for the crew was arrested in Iceland for illegal trading and taken to Norway for trial, and there are both Icelandic and Norwegian records. Through monopoly practices of the Norwegian king, brought out by Fridtjof Nansen's book *In Northern Mists* (1911), commerce with Europe now declined rapidly, and all regular contacts with Iceland and Norway were broken off around 1412. *See* also GREENLAND: *History;* VINLAND; ICELANDIC LITERATURE.

BIBLIOGRAPHY.—The best analysis of the sources is found in three books by Halldór Hermannsson, *The Northmen in America* (1909), *The Problem of Wineland* (1936) and *The Vinland Sagas* (1944). On the historicity of the Greenland-Vinland sagas *see* Sir William A. Craigie, *The Icelandic Sagas* (1913). For history *see* also Vilhjalmur Stefansson, *Greenland* (1942). (V. S.)

ERICA, the true heaths, a large genus of shrubs, sometimes treelike, of the family Ericaceae, native in South Africa and around the Mediterranean. *See* HEATH.

ERICACEAE, the heath family, an assemblage of woody dicotyledonous flowering plants of considerable horticultural importance, containing such ornamentals as rhododendrons and azaleas and fruits such as blueberries and cranberries. The family is world-wide in distribution and of tropical origin, with about one-half of its 1,800 species in frost-free regions. Of the 70 genera, the tropical American species of *Befaria* exhibit primitive characters, among these, that of having as many as ten free petals. Most of the remaining genera have the petals reduced in number and progressively more fused into a tubular corolla often composed of five or sometimes only four segments. The sepals and carpels usually equal the number of petal segments, with the stamens twice the number.

J. HORACE MCFARLAND CO.

AZALEA (RHODODENDRON [AZALEA] ARBORESCENS)

The family is primarily one of shrubs, although it contains various trees, such as the madroña (*Arbutus*) of Mexico and the Pacific coast of the U.S. which grows to 90 ft.; Mediterranean species of *Arbutus* (*e.g.,* the strawberry tree) are less robust. The group also contains small shrublets, such as trailing arbutus or Mayflower (*Epigaea*) of eastern North America, and high-climbing vines occur in tropical America. Coarse leathery evergreen leaves are common, deciduous leaved forms being found frequently in temperate regions. In seasonally arid areas, in highly acid bogs and on cold soil where the water is physiologically unavailable, the leaves tend to be small and evergreen with inrolled margins, thereby considerably reducing water loss.

Most of the species occur on acid soils which favour the development of mycorrhizal fungi (*see* MYCORHIZA). The strands of these nonpathogenic fungi are associated with the root tissues of the plants of this group and also ramify widely through the soil. In this symbiotic association the fungus apparently receives growth hormones as well as carbohydrates from the green plant and furnishes its associate with water, minerals and organic substances. Failure to supply proper growing conditions for the fungal associate leads to failure in the cultivation of members of the family. In general the group has a high light requirement, although some will tolerate shade; this accounts for their abundance in abandoned pastures and open bogs, or at high elevations and latitudes where the forest cover is sparse or lacking.

Subfamilies.—The family may be divided into two major segments, the Rhododendroideae and the Vaccinioideae.

Rhododendroideae.—This subfamily is characterized by a superior ovary, exceptions being found in a few species of *Gaultheria.* It has seemed expedient to bring certain members of this subfamily, such as rhododendrons, azaleas, rhodoras and tsutsutsis, under the generic name *Rhododendron* because of the complex situation in Asia, where confusing combinations of the variant leaf and flower types are present. Under this concept *Rhododendron,* with its more than 600 species, extends from Australia through the mountains of New Guinea and Malaysia into Asia and on into Europe and North America. The development of garden rhododendrons, in the ordinary sense, began in England, a later centre being the Puget sound area of western North America. The main centres for the culture and development of azaleas are Japan, Belgium, the Netherlands and the Atlantic and Gulf coastal regions of the U.S.

The nearly 500 species of *Erica,* the true heaths, are concentrated in South Africa, extending northward into Europe. Briar (or bruyère) tobacco pipes are made from the large root burls of several Mediterranean species. *Calluna,* the heather or ling, with over 70 garden varieties, consists of a single species of Europe and Asia Minor. *Gaultheria,* with four species in North America, among them the eastern checkerberry and west coast salal, circles the Pacific basin with its more than 100 species. A number of genera—*Kalmia,* containing the eastern North America mountain laurel; *Enkianthus* of Asia; *Pieris* of eastern North America and eastern Asia; *Leucothoë,* found in North and South America, Japan, the Himalayan region and Madagascar; and others—also contribute beauty to gardens.

Vaccinioideae.—This subfamily is characterized by an inferior ovary, or one rarely only half submerged in the floral receptacle. The group contains about 35 genera and is world-wide in distribution, being most highly developed in Andean South America. It appears not to have evolved from the inferior ovaried gaultherias and may be a parallel development; for this reason it sometimes is treated as the Vacciniaceae, a family apart from the Ericaceae. The flamboyant crimson, salmon and yellow flower clusters of many of the tropical genera are but little evident in the northern blueberries, cranberries, whortleberries and bilberries, all in the genus *Vaccinium.* This genus of 140 species, widespread in the temperate and subarctic northern hemisphere, also is represented on tropical mountains. In eastern North America are found the horticulturally most important species of the blueberry and cranberry subgenera. *Gaylussacia,* the true huckleberry, extends from Brazil into eastern Canada. The term huckleberry often is applied to various low-growing species of blueberry, but it should be reserved for species of *Gaylussacia.* Other technical differences

aside, the two genera are easily differentiated while eating the fruit, for the seeds of blueberries are numerous, small and insignificant, whereas true huckleberries have no more than ten large, hard seeds which can be painful when wedged between the teeth. *See* also separate articles on the members of this family; *see* also references under "Ericaceae" in the Index volume. (W. H. C.)

ERICE, a town in Trapani province, Sicily, called Monte S. Giuliano until 1934, lies 14½ km. (9 mi.) E. by N. of Trapani, 2,464 ft. above sea level. Pop. (1957 est.) 13,540 (commune). Area of commune 119 sq.mi.

Situated on the top of Monte S. Giuliano, it affords an immense view. The streets of the town, narrow and twisting, retain their medieval character. The important monuments are the remains of Cyclopean walls with Phoenician inscriptions; the Romanesque church (Chiesa Matrice); and the medieval castle, which contains the remains of an ancient temple of a local nature deity. Under the Romans this was famous throughout Italy and Africa as the centre of the cult of Venus Erycina (from Eryx, the ancient name of the town). The railway station is at Trapani, but the two towns have good communications. Local occupations are stock-breeding and vine and olive cultivation. (M. T. A. N.)

ERICSON (ERICSSON), **LEIF:** *see* ERIC the Red.

ERICSSON, JOHN (1803–1889), Swedish-U.S. naval engineer, the builder of the first armoured turret ship, was born at Långbanshyttan, Värmland, Swed., on July 31, 1803. He served for a time in the Swedish army and in 1826 he went to London. There, in partnership with John Braithwaite, he constructed the "Novelty," a locomotive engine for the Liverpool and Manchester railway competition at Rainhill in 1829. The prize, however, was won by Stephenson's "Rocket." Among other things he worked out a plan for marine engines placed entirely below the water line. In 1833 his caloric engine was made public and in 1836 he took out a patent for a screw propeller, which was first tried in 1837 on the "Francis B. Ogden," built in London. At this time Captain Stockton, of the U.S. navy, gave an order for a small iron vessel to be built by Laird of Birkenhead and to be fitted by Ericsson with engines and screw. This vessel reached New York in May 1839.

A few months later Ericsson followed his steamer to New York, and there he resided for the rest of his life. In 1848 he was naturalized as a citizen of the United States. During the American Civil War, the navy department invited proposals for the construction of armoured ships. Ericsson's design was accepted and the first armoured turret ship, the "Monitor," was launched on Jan. 30, 1862. On March 9, she fought the celebrated action with the Confederate ram "Merrimack" (*see* "MONITOR" AND "MERRIMACK," BATTLE OF). In later years Ericsson devoted himself to the study of torpedoes and sun motors. In 1876 he published *Solar Investigations* and *Contributions to the Centennial Exhibition.* Ericsson died in New York on March 8, 1889, and in the following year, at the request of the Swedish government, his body was sent to Värmland, where it was buried at Filipstad.

See W. C. Church, *The Life of John Ericsson,* 2 vol. (1890); H. P. Spratt, *Marine Engineering* (1953). (H. P. ST.)

ERIDANUS, in astronomy, a large constellation of the southern hemisphere, stretching from near Rigel at the heel of Orion (*q.v.*) to Achernar (α Eridani), not far from the south polar circle. Achernar, a blue star of the first magnitude, is located at a distance of the order of 100 light years from the sun. The multiple system o_2 Eridani is of especial interest. It consists of three stars: a yellow dwarf—not very unlike our sun—of naked-eye visibility, together with two faint telescopic companions. One of the latter is a red dwarf, while the other belongs to the remarkable group of stars known as white dwarfs (*see* SIRIUS).

Eridanus fluvius is the name under which the constellation appears in the catalogue of Tycho Brahe, Eridanus being a Greek river-god and also an ancient name for the Po river in Italy. In the older catalogues of Ulugh Beg and of Ptolemy (who catalogued 34 stars in it) it appeared simply as "the river" stars.

ERIDU, an ancient city 14 mi. S. of Ur (Tall al Muqayyar) in the Nasiriyah district of modern Iraq, now situated in the desert and marked by the mound known as Abu Shahrain. The city, which was once washed by an arm of the Euphrates, was revered as the most ancient in Sumer according to the king lists. Cuneiform texts also state that it once stood on the shore of the sea, probably on the margin of what is now known as the Hammar lake which once reached as far north as the city. The great tower (ziggurat) built by Amar-Sin about 2050 B.C. is still visible many miles away in the flat desert that surrounds the site.

The site, covering about 60 ac., was excavated principally between 1946 and 1949 by the Iraq antiquities department under the direction of Fuad Safar and Seton Lloyd. Eridu proved to be one of the most important of the prehistoric urban centres in southern Babylonia. Founded on sand dunes probably in the 5th millennium B.C., it fully illustrates the sequence of the preliterate 'Ubaid civilization, the hallmark of which is a painted pottery artistically decorated with geometric and sometimes floral and animal designs in a dark paint on a buff or drab clay. A long succession of superimposed temples has displayed the growth and development of an elaborate mud-brick architecture from the simplest beginnings. At a very early stage there appears the characteristic Sumerian temple plan with central nave and chambers opening off it, podium at the short end, façades decorated with niches and buttresses. Associated with the floor of the latest of the 'Ubaid temples there was a large quantity of fishbones probably dedicated to the god Ea or Enki, the Sumerian patron of Eridu, "lord of the sweet waters that flow under the earth." The inhabitants were engaged both in agriculture, for which they employed primitive flint sickles, and in fishing, and irrigation is attested by the remains of a prehistoric canal bed.

The city continued to be occupied down to about 600 B.C., but was less important in historic periods.

BIBLIOGRAPHY.—*Sumer,* vol. iii–vii, published by the Directorate General of Antiquities, Government of Iraq (1947–51); A. Perkins, *The Comparative Archaeology of Early Mesopotamia* (1949). (M. E. L. M.)

ERIE, the seat of Erie county, Pa., U.S., is an important manufacturing, commercial and resort city and port of entry situated in the northwest corner of the state on Lake Erie. Pop. (1960) city, 138,440; standard metropolitan statistical area (which includes all of Erie county), 250,682. Suburban residential communities are Lawrence Park, Mill Creek, Harborcreek, Fairview and Wesleyville.

Named for the Erie Indians, the city was the site of the French fort Presque Isle, built in 1753. Abandoned to the British in 1759, the fort was destroyed by the Indians during the Pontiac conspiracy of 1763. In 1788 the wilderness area was purchased by Pennsylvania from the federal government. A U.S. fort Presque Isle was built at the site in 1795 and a city was laid out by Gen. Andrew Ellicott, first surveyor general of the United States.

Erie county was established in 1800 and the city was named county seat in 1803. Two years later it became a borough and in 1851 a city. Most of Commodore Oliver H. Perry's vessels used in defeating the British in the battle of Lake Erie, Sept. 10, 1813, were built at Erie.

In 1913 the flagship "Niagara" was partially rebuilt and placed on display near the public dock. In 1844 Erie was connected with Pittsburgh by a canal used until 1871. Early manufacturing industries were largely dependent upon agriculture. The first iron foundries used bog ore from the bay swamps.

Industries are diversified, numbering over 400. Metal products are most important in value, with food products and paper next. Manufactures include electric locomotives, castings and patterns, power shovels, forgings, meters, engines, boilers, plastic products, toys, asbestos products, sterilizing equipment, wringers and radio parts.

Erie's Presque Isle bay is Pennsylvania's only St. Lawrence seaway port and one of the finest harbours on the Great Lakes. It is protected by Presque Isle peninsula, 11 mi. long, and has 4 mi. of harbour frontage. Incoming ships carry pulpwood, iron ore, limestone and grain. Outgoing vessels carry Pennsylvania coal, oil and manufactured products. Sand and gravel are brought in from the lake bottom.

The surrounding countryside, a plain of rich sandy soil extend-

ing south from the lake, yields rich harvests of grapes, pears and cherries. Both the bay and lake are abundant with fresh-water fish.

Within the city are three Roman Catholic colleges: Gannon, for men, founded in 1944; Mercyhurst, for women, founded in 1926; and Villa Maria, for women, founded in 1925. Nearby are the Behrend centre of Pennsylvania State university and the State college at Edinboro. Roman Catholic and Protestant Episcopal diocesan headquarters are located in the city.

A Greek revival period customhouse, a community playhouse, a philharmonic orchestra, a public museum, Glenwood park and zoo, the Wayne blockhouse, a memorial stadium and more than 40 parks and playgrounds are of interest in the city.

Presque Isle, across the bay, is a state park of 3,000 ac., including 7 mi. of beaches. (W. A. C.)

ERIE, LAKE, named from the Erie Indians, is the fourth largest of the Great Lakes of North America. Bordering on it are the states of Michigan, Ohio, Pennsylvania and New York and the province of Ontario, Can. It has a long dimension of 241 mi., trending east-northeast, and a maximum width of 57 mi. The surface area of the lake is 9,930 sq.mi. and the area of its drainage basin exclusive of lake surface is 22,750 sq.mi. The average annual precipitation of the area is 34 in. The mean height of lake surface above sea level is 572 ft. With an average water depth of about 58 ft. and a maximum depth of 210 ft., Erie is the shallowest of the Great Lakes; short-period fluctuations in lake level, caused by storms, are of relatively frequent occurrence—at the ends of the lake these amount to several feet—and strong winds produce unusually short, choppy seas. (For comparisons with the other Great Lakes, a discussion of their origins, connections, utilization, etc., *see* GREAT LAKES, THE.)

History.—The Lake Erie country was held by Iroquois Indians, who were unfriendly toward the French and their Indian allies, long after the French had developed an extensive trade on the other Great Lakes. The first known white man to see Lake Erie was Louis Jolliet, who traversed its northern shore in 1669 while returning eastward from Lake Superior. The British, with whom the Iroquois were allied, began serious development of trade along Lake Erie in the late 1600s. Antoine de la Mothe Cadillac established a fort at Detroit in 1701, to protect French interests, and this, with periodically strengthened French fortifications at the mouth of the Niagara river, kept both ends of Lake Erie under French control. British pressure continued to develop, however. Ft. Niagara was taken in 1759, and then Maj. Robert Rogers with 200 rangers took Detroit in 1760. During the American Revolution Indian allies of the British made extensive raids on American colonists in the Ohio and Mohawk valleys, but no major engagements occurred in the region. At the close of the war the Canadian-American boundary was drawn through the Niagara river, Lake Erie and the Detroit river. Many British loyalists moved from American territory and settled in southern Ontario north of Lake Erie immediately after the war, but American settlement of the "western reserve" and other lands south of Lake Erie was delayed until after 1796. Buffalo was founded at the eastern tip in 1803.

One of the important engagements of the War of 1812 was the battle of Lake Erie, in which U.S. Commodore Oliver H. Perry, with a hastily built fleet, defeated the British squadron, under Capt. Robert H. Barclay, and thus secured the northwest for the United States.

Physiography.—The principal rivers tributary to Lake Erie are the Detroit (which carries the discharge of Lake Huron, 178,000 cu.ft. per second, and of Lake St. Clair watershed), the Huron and Raisin rivers of Michigan, the Maumee, Portage, Sandusky, Cuyahoga and Grand rivers of Ohio, the Cattaraugus creek of New York, and the Grand river of Ontario. The west end of the lake contains all of the islands, the largest of which is Pelee Island, Ont. (8.6 mi. long). The discharge of the lake, through the Niagara river which flows northward from its east end, averages 195,800 cu.ft. per second.

The greater part of Lake Erie lies in a belt underlain by soft Devonian shales, with older Devonian limestones under the north shore. The islands in the west end are composed of limestones and dolomites of Silurian and Devonian age. Sandy beaches are nearly continuous around the lake, and the adjacent land generally is gently rolling. Wave-cut cliffs in glacial deposits and lake clays are common features, particularly along the south shore. Former beaches abandoned by the lake and lying inland are utilized as routes for railways and roads. One of these beaches indicates that the lake level once was 220 ft. above its present position.

Harbours and Commerce.—The important steel industry of the Lake Erie region, including the Pittsburgh district, results from the movement of iron ore and limestone to Lake Erie ports and the transportation of coal by rail from the mines of Pennsylvania and West Virginia. Additional coal from the same sources furnishes part of the return cargoes. Wheat and other grain from the prairie provinces of Canada and the great grain belt of the United States furnish enormous cargoes to eastern markets and for export through the St. Lawrence seaway (*q.v.*).

Navigation from Lake Erie to Lake Ontario passes through the Welland ship canal, with a channel depth of 27 ft., constructed and operated by Canada (*see* WELLAND SHIP CANAL). The New York state barge canal has one of its Great Lakes connections at Tonawanda, N.Y., on the Niagara river, but also maintains terminals at Buffalo. This canal, which connects with the Hudson river and has branches to Lake Ontario and Lake Champlain, has a minimum depth of 12 ft.

Originally a few harbours were formed by natural bays, but most of them are at the mouths of streams which were improved by protective piers and breakwaters and by dredging to accommodate the large lake vessels.

The most important ports of the United States on Lake Erie are Sandusky, Toledo, Huron, Lorain, Cleveland, Fairport, Ashtabula and Conneaut, Ohio; Erie, Pa.; and Buffalo, N.Y.; and of Canada, Port Colborne, Ont., at the entrance to the Welland ship canal. Cleveland, Ashtabula and Conneaut are the most important iron ore ports. Toledo leads all the rest in its water shipments of soft coal and Buffalo is by far the largest grain port.

See also references under "Erie, Lake" in the Index volume.

BIBLIOGRAPHY.—*Great Lakes Pilot*, vol. i and charts, Canadian Hydrographic Service (annual); *Great Lakes Pilot*, and charts, U.S. Lake Survey (annual); Harlan Hatcher, *Lake Erie* (1945). (J. L. HE.)

ERIE CANAL, an inland waterway connecting Lake Erie at Buffalo, N.Y., with Albany and Troy on the Hudson river and, by way of the Hudson, with the port of New York. Construction was begun in 1817. The 15-mi. section between Utica and Rome was opened in 1819 and was immediately successful. Traffic began to move through the entire length of the canal in 1825.

History.—At the close of the 18th century the population of the United States was confined largely to the Atlantic coastal region. Movement of settlers westward and of produce eastward was sharply limited because of poor roads, scarcely more than forest and mountain trails, which made transportation costs almost prohibitive. By 1800 in Massachusetts, New York, Pennsylvania, Maryland and Virginia serious consideration was given to proposals for the development of land or of water transportation from the Atlantic seaboard to the west including construction of canals or of combined canal and land routes which would lower costs and increase the volume of shipments to the extent that the great central valley west of the Appalachian mountains could be settled rapidly.

In 1810 DeWitt Clinton, later governor of New York, supported a water route across the state to open up transportation from Albany to Lake Erie at Buffalo. In spite of bitter rivalries, jealousy and political opposition Clinton continued to lead the movement for the Erie canal. In 1811, accompanied by Gouverneur Morris, he traveled to Washington, D.C., to seek aid for this project but was unsuccessful. However, the citizens of New York refused to abandon plans for a canal, and in 1817 the legislature authorized the expenditure of $7,000,000 on the credit of the state for the construction of the canal.

Construction.—The canal was 363-mi. long, 40 ft. wide and 4 ft. deep. The route west from Troy and Albany followed the Mohawk valley through a wild and sparsely settled region. Although

it was called the water-level route because it did not cross high elevations and followed the Mohawk river grade much of its length, enjoying the most favourable operating conditions for a canal of any route from the Atlantic to the Ohio-Mississippi river region, it carried boats over a more than 500-ft. rise in elevation from Troy to the Lake Erie water level by the use of 82 locks. The canal was built solely by horsepower and man power. No roads existed over which supplies could be hauled. The canal was built through Montezuma swamp and was carried across several streams in watertight aqueducts which were large enough to permit the passage of barges. In several places it was necessary to cut through rock aided only by the use of black blasting powder. In spite of many difficulties the canal was completed at a cost of approximately $20,000 per mile.

On Oct. 25, 1825 the Erie canal was opened. The canal boat "Seneca Chief" left Buffalo for Albany with Governor Clinton and other state officials as well as the canal commissioners on board, and carrying two kegs of Lake Erie water which was later poured into the Atlantic ocean in New York harbour.

Influence.—The Erie canal contributed directly not only to the development of the entire region through which it passed but also to the growth of New York city, Albany and Buffalo. It served as the main artery for both freight and passenger traffic between New England and New York and the rapidly growing states of Ohio, Indiana, Illinois and Michigan. A substantial majority of the settlers coming from the east into these states traveled through the canal. From the western terminus at Buffalo lake freighters hauled supplies for settlers, machinery and other manufactured goods and brought increasing amounts of grain to Buffalo for shipment through the canal to the port of New York and thence to European markets. The number of lake boats entering and leaving Buffalo harbour increased 150% during the first two years of the operation of the canal.

During the first decades of the 19th century a fierce struggle was waged over the marketing of the products of the middle west. Rival ports such as New Orleans, La., Quebec, Que., and the Atlantic seaboard ports of Baltimore, Md., New York city and Philadelphia sought to secure this trade. More than any other single factor, the Erie canal turned this trade eastward to the port of New York. Cost of transportation and time taken to market goods exerted a strong influence. Freight rates between New York and Buffalo dropped from about $100 per ton, by land, to $10 a ton through the canal. Transportation time from Albany to Buffalo was cut from 15 days by wagon to 6 days by canal. Fast passenger packets, drawn by relays of horses driven along the towpath at a trot, could travel the length of the canal in about 3½ days.

Finances.—The Erie canal was a financial success. Tolls collected during the first two years exceeded operating costs and interest on the debt. The first nine years of operation produced income from tolls in excess of the initial cost of construction. From 1825 to 1882, the year canal tolls were abolished, the Erie canal had paid for its construction and operation, had furnished money for the building of several lateral or feeder canals, and had provided some revenue for the general government funds of the state. The gross income from 1825 to 1882 was $121,461,871. After all construction, operation and maintenance costs had been paid, the canal showed a profit of $42,599,717. Unfortunately little money was spent improving the canal. In 1835 New York began a program of canal enlargement. By 1862 it had been deepened to 7 ft. and widened to 70 ft. The locks were lengthened to accommodate larger barges. In spite of the almost universal use of steam power on railroads and ships, as late as 1900 canal barges continued to be drawn through by horses driven along the towpath. Completion of railroad lines closely parallel to the canal brought keen competition and bitter rivalry. Railroads opposed any improvements for the canal and gradually it became neglected. In spite of competition, opposition and neglect, freight tonnage increased slowly until 1889 in which year approximately 4,500,000 tons of freight was moved. After this peak year, freight tonnage declined to about 2,000,000 tons in 1900.

N.Y. State Barge Canal System.—The continued industrial expansion and increasing population in the middle western states brought a substantial increase in freight shipments. Railroads found some difficulty in handling this growing volume of traffic. However, New York state found that shipments moving through it were declining. As a result a demand grew for improvement of its canals. A bond issue furnished $100,000,000 for the construction of the New York State Barge Canal system. Work was begun in 1905 and was completed in 1918 at a cost of more than $150,000,000. The barge canal system consists of four divisions or canals: the Erie, from Troy to Buffalo; the Champlain, north from Troy to Lake Champlain; the Oswego, from Three Rivers (near Syracuse) to Lake Ontario at Oswego; and the Cayuga-Seneca, connecting these two lakes with the Erie division. There are 56 locks on the improved canals, 34 of them on the Erie, exclusive of the U.S. government lock at Troy. The system was planned to accommodate barges about 300 ft. long and 42 ft. wide with a draft of 10 ft. and a cargo capacity of 2,200 tons. A slow but steady increase in tonnage followed the completion of the system and by the second half of the 20th century 4,750,000 tons were being shipped annually. About 25% of the tonnage moved eastward; about 75% moved westward. Petroleum products accounted for about 70% of the tonnage passing through the barge canal system.

Bibliography.—Dorothie Bobbé, *De Witt Clinton* (1933); Seymour Dunbar, *History of Travel in America* (1915); Archer B. Hulbert, *Paths of Inland Commerce* (1921); Francis P. Kimball, *New York—The Canal State.* (1937). (R. E. At.)

ERIGENA, JOHANNES SCOTUS (*c.* 810–877), Irish philosopher and theologian, one of the most notable thinkers of the Carolingian middle ages. Little is known about his life. He was born in Ireland, the ancient Erin (whence Erigena or Eriugena) or Greater Scotia (whence Scotus; in the 9th century the Irish were called Scots). About 851 he was at the court of the West Frankish king Charles the Bald as master of the palace school. He took part in the theological disputes stirred up by Gottschalk's doctrines, and at the request of Archbishop Hincmar (*q.v.*) wrote *De praedestinatione.* This work, far from settling the controversy, exacerbated it, and was condemned at the councils of Valence (855) and Langres (859). At Charles the Bald's behest he translated the works of the Pseudo-Dionysius (*see* Dionysius [The Areopagite]) and Maximus the Confessor's *De ambigua* as well as Gregory of Nyssa's *De imagine.* His own principal work is the *De divisione naturae;* he also wrote commentaries on the Pseudo-Dionysius' *Celestial Hierarchy,* on Boethius' *Opuscula sacra* and on the prologue of St. John's gospel.

Unusually learned for his time (he quotes many Latin and Greek Fathers), Erigena stands out by his own strong personality. His version of the Pseudo-Dionysius, who as a supposed disciple of St. Paul commanded great authority, acquainted the west with many Neoplatonic tenets and raised the problem of their links with Christian thought. His own views as set out mainly in *De divisione naturae* are of especial interest.

Erigena was not a rationalist or free thinker. He discusses human authority, but accepts that of the church and divine revelation. Reason comes before human authority: the latter is merely truth discovered by reason and set down in writing for the benefit of posterity. The authority of the Fathers is not denied, but reason is entitled to examine their thought respectfully in order to select what seems the likeliest solution. Divine authority remains unchallenged, but we must give rational meaning to certain biblical expressions which taken literally would be false. In looking for the spiritual meaning even the Fathers could err, witness the many disagreements between them. Belief must not be blind, but faith comes before reason; faith illumines life's mystery and the whole of reality: unless we believe we cannot understand St. Augustine had reached the same conclusion and St. Anselm was to do so.

In his theory of knowledge, Erigena explains that, as the objects of knowledge are diverse, so are our cognitive powers. Thus (1) "sense" or perception receives distinct impressions of things —of their nature by its "internal" function, of their appearances by its external; (2) "reason," on the other hand, is exercised on the examination of the natures and causes of created things; and

(3) the "intellect," mind or spirit, is concerned with God. Reason is a faculty or power generated in us by the intellect and inferior to it; it rises to God through his theophanies or manifestations of himself and is called "right" or "true" reason in so far as it concords with the intellect. It is not an independent source of knowledge or an autonomous criterion of truth, but draws on the divine light and finds the rules of its right judgment in the divine truth which is mysteriously with us. Knowledge is always illumination; reason, the mind's eye, is a true instrument of religious knowledge.

Erigena's theory of knowledge presupposes Catholic teaching on man's development; creation, full redemption and finally the vision of God. It also presupposes a conception of God as transcendent and infinite. Negative theology considers God as substance, which in our ignorance of the divine essence we know by what he is not, rather than by what he is. Positive theology excludes agnosticism by considering God as cause, so that the soul can continually attribute to him the perfections of creation: God is the beginning, middle and end; the Creator omnipotent, one, simple and unique, infinite, immutable; the exemplary cause, without form yet form of all things, their essence yet transcending them; farthest and nearest, abysmal darkness and light of the mind; hidden, unfathomable but showing himself to us in all things, which yet do not exhaust his nature. He is formless, as not having any determined, finite form; but form of things, because creator of all forms. He is the essence of things not because they are part of him or because, pantheistically, he is the soul of all things, but because things are what he has known and wanted them to be. He creates himself in things in the sense that he shows himself through them. Creatures are metaphorically parts of God. Pantheism is consciously rejected.

The doctrine of creation is central to *De divisione naturae*. By "nature" Erigena means the totality of things distinguished under four heads: (1) nature uncreated and creating; (2) nature created creating; (3) nature created but not creating; and (4) nature neither created nor creating. The first and the fourth are God as beginning and end of things, the second and the third show creatures in their double mode of existence, the intelligible and the sensible. The intelligible creatures, nature created and creating, are the primordial causes. The creative act being eternal, these primordial causes are eternal in so far as they do not exist in time, but since they have a beginning Erigena prefers to call them co-eternal with the Word, receiving their being from God. Primordial causes, which we see as many because we refer them to many things, are one and indivisible in God. They are the reasons of creation; and the exemplary ideas are based on them and subsist in them, just as they reside in the Word. The doctrine of exemplary ideas explains the double existence of things, in the Word as eternal and in themselves as temporal. The world is an unending series of theophanies.

Man is a theophany of an exceptional kind. As he is made in God's image his innermost being shares in God's unknowability. He is the notion that God has of him, and he is also the *copula mundi* or "link of the universe," as everything has been created in him and in his idea (all ideas are therefore innate). Sin, however, has dimmed man's vision and brought his animal nature to the fore, producing the division into sexes. After the fall and the ensuing loss of man's status as mediator, the divine Word has returned incarnate, adding to the given creation the gift of grace, thus leading all creatures back to the Creator. The workings of grace are a manifestation of the workings of divine providence: creation and redemption are the two poles of the great human and cosmic drama; the fall does not change the purpose of creation, but only the means of realizing it. The return of creation begins with release from sin, physical death and entry into the life hereafter. In the return to God, or conversion, there will be a progressive spiritualization of the created, a deification of creatures without however any loss of their individuality or particular being as creatures. As air permeated with light, or iron glowing in fire, retains its nature, so creatures in union with God retain theirs.

The main collection of Erigena's writings is that by H. J. Floss, printed in J. P. Migne's *Patrologia latina*, vol. cxxii (1853); there is a modern edition of the commentary on Boethius' *Opuscula sacra* by E. K. Rand, *Quellen und Untersuchungen zur lateinischen Philologie des Mittelalters*, i, 2 (1906).

BIBLIOGRAPHY.—A. Gardner, *Studies in John the Scot* (1900); H. Bett, *Johannes Scotus Erigena* (1925); C. Albanese, *Il pensiero di Giovanni Erigena* (1929); M. Cappuyns, *Jean Scot Erigène* (1933); G. Bonafede, *Saggi sul pensiero di Scoto Eriugena* (1950); M. Dal Pra, *Scoto Eriugena* (1951); F. Vernet, "Erigène," in *Dictionnaire de Théologie Catholique*, vol. v (1913). *See* also E. Gilson, *La Philosophie au moyen âge* (1944); G. Bonafede, *Storia della filosofia medievale* (1957). (GI. B.)

ERIGONE, in Greek mythology, daughter of Icarius, the hero of the Attic deme (township) of Icaria. Her father, who had been taught by Dionysus to make wine, gave some to some shepherds, who became intoxicated. Their companions, thinking they had been poisoned, killed Icarius and buried him under a tree on Mount Hymettus (or threw his body into a well). Erigone, guided by her faithful dog Maera, found his grave and hanged herself on the tree. Dionysus sent a plague on the land, and all the maidens of Athens, in a fit of madness, hanged themselves like Erigone. Icarius, Erigone and Maera were set among the stars as Boötes (or Arcturus), Virgo and Procyon. The festival called Aeora (the "Swing") was subsequently instituted to propitiate Icarius and Erigone. Various small images (Lat. *oscilla*) were suspended on trees and swung backward and forward and offerings of fruit were made. The story was probably intended to explain the origin of these *oscilla*.

ERIN, an ancient name of Ireland. *See* IRELAND.

ERINNA (date unknown), Greek poetess of the Dorian Island of Télos (near Rhodes), was famous in antiquity for "The Distaff" described in the lexicon by Suidas as a poem of 300 lines written in a mixture of the Dorian and Aeolic dialects. Her surviving works (three short quotations, three epigrams in the *Palatine Anthology* and a papyrus fragment of a lament for her friend Baucis) can be ascribed to the 4th or 3rd century B.C. (Eusebius puts her *c.* 350). She is said to have died at the age of 19.

See M. Bowra, *Problems in Greek Poetry* (1953). (JN. A. D.)

ERINYES, in Greek mythology, the name given to the deities of vengeance; probably personified curses, but possibly in their origin ghosts of the murdered; in Roman literature they are called Furies. According to Hesiod they were the daughters of Earth (Gaea), and sprang from the blood of her mutilated spouse Uranus; in Aeschylus they are the daughters of Night, in Sophocles of Darkness and of Earth. Sometimes one Erinys is mentioned, sometimes several. Euripides was the first to speak of them as three in number. Later Alexandrian writers gave them the names Alecto (unceasing in anger), Tisiphone (avenger of murder) and Megaera (jealous).

Their home is the world below, whence they ascend to earth to pursue the wicked. They punish all offenses against the laws of human society, such as perjury, violation of the rites of hospitality and, above all, the murder of relatives. Though just, they are merciless and take no account of mitigating circumstances. Being deities of the underworld, they are often identified with spirits of the fertility of the earth, such as the Semnai or Eumenides (*q.v.*) at Athens (*see* also DEMETER). In Aeschylus, the Erinyes are represented as awesome, Gorgonlike women wearing long black robes, with snaky locks, bloodshot eyes and clawlike nails. Later, they are winged maidens of serious aspect, in the garb of huntresses, with snakes or torches in their hair, carrying scourges, torches or sickles.

See H. J. Rose, *A Handbook of Greek Mythology* (1933).

ERIS, in Greek mythology, the personification of Strife, daughter of Night, sister and companion of Ares (*q.v.*). Homer describes her initial insignificance and ultimately formidable growth (*Iliad* 4,440 ff.), also her role in inciting men to insane battle (*Iliad* 11, 3 ff.). Eris is best known for her part in starting the Trojan War. When she alone of the gods was not invited to the marriage of Peleus and Thetis, she threw among the guests a golden apple inscribed "For the most beautiful." Hera, Minerva and Aphrodite each claimed it, and Zeus, refusing to make the difficult decision, assigned the matter to Paris (*q.v.*), whose judgment was to be final. Paris awarded the apple to Aphrodite and received Helen, but, in the war that resulted, Hera and Minerva

remained implacable enemies of Troy. The Latin equivalent of Eris is Discordia, known in Ennius' *Annals* and of great symbolic importance (representing the disorder which Augustus ultimately ends) in *Aeneid* 8.702 ff. (WM. S. A.)

ERITH, a municipal borough (1938) of Kent, Eng., which under the London Government Act 1963 was to be amalgamated on April 1, 1965, with Bexley, Crayford and the northern part of Chislehurst and Sidcup to form the new London borough of Bexley (*see* LONDON). Erith municipal borough, 14 mi. E. of London bridge, has a frontage of more than four miles on the Thames. Pop. (1961) 45,043. Area 6 sq.mi. The parliamentary constituency is Erith and Crayford. The industrial section in the north has heavy and electrical engineering works and factories for making cables, chemicals, radio and television equipment, paints, plastics and toys. There are also warehouses, wharves and facilities for shipment of automobiles. The southern outfall works of the London main drainage system are at Crossness in the Erith marshes south of which is the residential section where the land reaches 200 ft. above sea level. In this area are Abbey Wood and Belvedere (the ancient district of Lessness [or Lesnes] Heath with remains of an abbey in the woods which cover 215 ac.) and Northumberland Heath. Erith parish church is partly Norman (restored). There are boys' and girls' grammar schools, a technical college and technical school and, in Belvedere park, the Royal Alfred home (1856) for aged merchant seamen. Franks park covers about 40 ac. of natural woodland. Erith is mentioned in a Latin charter, possibly of the 7th century, in connection with a gift of land made by the bishop of the East Saxons. It was anciently a borough and was granted a market and fairs in 1313. Down to the end of the 17th century it was of some importance as a naval station, and in its royal dockyard the "Great Harry" was fitted out for Henry VIII. An annual regatta is held in August, and there are many sports fields including a stadium opened in 1955.

ERITREA, a province on the Red sea, in federation with Ethiopia (*q.v.*) under the sovereignty of the Ethiopian crown since 1952, and formerly an Italian colony, extends about 600 mi. from Ras Kasar (Cape Kasar), lat. 18° 2′ N., to Ras Dumeira, lat. 12° 42′ N., in the Strait of Bab el Mandeb. It is bounded northwest by the Republic of the Sudan, southeast by French Somaliland and south by Ethiopia proper. Total area (including the islands of the Dahlak archipelago), 47,876 sq.mi.

Physical Geography.—The northern part of Eritrea extends across the northern stretches of the Ethiopian high plateau from the plains of the Sudan in the west to the coastal plain of the Red sea in the east. The coastal plain varies in width from 10 to 50 mi. South of the Gulf of Zula, the lowlands widen to include the plain of the Danakil region, which contains the depression known as the Kobar sink (nearly 400 ft. below sea level), the northern end of which extends into Eritrea. The coastal plain and the Danakil plain are part of the East African rift and are sharply limited on the west by the fault line of the eastern escarpment of the plateau which, although deeply eroded, presents a precipitous obstacle to travelers from the coast. The plateau is dissected by the valleys of westward-flowing rivers, in particular by the tributaries of the Barka and by the Mareb (Gash), which forms part of the southern boundary of Eritrea. The high plateau in Eritrea has been reduced to a comparatively narrow strip of country above 6,500 ft., extending from Amba Alagi to about 50 mi. N. of Keren, but containing higher points such as Mt. Soira (9,885 ft.) near Senafé. The land slopes gradually westward toward the Sudan from this high country and the western third of Eritrea consists of broken and undulating country at an average height of about 1,500 ft. Geologically Eritrea consists of a base of ancient crystalline rocks (granites, gneiss, micaschists) which is widely exposed as the overlying sandstones have been subjected to extensive erosion. Relatively small caps of sandstone do, however, remain, giving rise to the characteristic steep-sided *amba* formations.

Eritrea has a wide variety of climatic conditions. The differences in altitude produce a significant temperature range. Massawa on the coast has one of the highest averages in the world, and while the plateau is neither high enough nor extensive enough in Eritrea to produce really cool conditions, temperatures at Asmara are more than 20° F. lower. The rainfall on the plateau and the western borderlands is of the same type as that of the Ethiopian plateau to the south, although the amount is much smaller. This summer rainfall comes in a southwesterly air stream and decreases in quantity and length of rainy season toward the northeasterly extremes of the plateau. The eastern edges of the plateau and, to a smaller extent, the coastal fringes receive much smaller quantities of rainfall from the northeasterly air stream in winter and early spring. The interior regions of the Danakil plain are, however, practically rainless. The basic climatic variations in Eritrea can be seen in the following table:

Place	Annual rainfall (in.)	Annual average temperature (°F)	Hottest month	Coolest month	Altitude region (ft.)
Massawa	6¼	86	93	85	Sea level (*)
Ghinda	32	76	86	65	3,200 (†)
Asmára	20	62	71	60	7,500 (‡)
Kerén	18½	70	78	65	4,600 (§)
Sabderat (near Kassala)	12	80	90	70	1,700 (‖)

*Red sea coast. †Eastern edge of escarpment (summer *plus* winter rain). ‡Eritrean plateau. §Northern plateau. ‖Western lowlands, Eritrea-Sudan border.

Population.—The largest groups are the Semiticized peoples of the plateau. The most significant is the great Tigrinya-speaking block of the northern Ethiopian highlands which is centred on the old kingdom of Aksum but extends northward to the Asmara region. This area is largely agricultural and the people mostly adhere to the Coptic Church (*q.v.*). In the northern half of Eritrea, occupying the highlands, the eastern plains and parts of the western plains, is a largely pastoral and nomadic population, almost entirely Muslim and speaking the Semitic language of Tigré. The spread of Tigré has produced dialectal differences between the eastern and western lowlands and within the area are islands of people belonging to the Hamitic (Cushitic) language group speaking one of the Agaw languages. Several of the tribes are bilingual (Tigré-Bilen; Tigré-Beja), but the Agaw languages, generally used in the family circle at home, are disappearing. The desert tribesmen south of Massawa speak a Cushitic language, Saho, akin to the language of their southern neighbours, the Danakil or Afar, who speak the Danakili language and who live to the south of the Gulf of Zula. The Saho and the Afar are Muslim nomads. Recent migrants to the coastal settlements and the inhabitants of the regions bordering Sudan to the west and north include many who are Arabic speaking.

The total population was estimated at 1,100,000 in 1956; no census has ever been taken. The European population, mostly Italian, was then about 20,000, but before 1935 had seldom exceeded 5,000.

The capital city, Asmara (*q.v.*), is largely of Italian construction and is situated on the extreme edge of the escarpment at an altitude of 7,500 ft., overlooking the port of Massawa (*q.v.*). This is the chief port, but an increasing amount of trade is shared by the

BY COURTESY OF UNITED NATIONS

MOSQUE OF KEREN, WESTERN PROVINCE, ERITREA

developing port of Assab (*q.v.*) to the southeast. On the plateau the main Eritrean settlements are Keren, Agordat, and Barentu on the road to Tessenei in the western lowlands: Adi Ugri and Adi Quala on the road southwest to Gondar; Decamere, Saganeti, Adi Caieh and Senafe on the southern highway to Addis Ababa; and Ghinda in the foothills on the road between Massawa and Asmara. None of these plateau towns has a population exceeding 4,000. Rural population densities may reach 250 per square mile in the most closely settled areas of the plateau, but the practically uninhabited regions of the eastern lowlands reduce the average for the whole territory to approximately 23 per square mile.

History.—The migrations of Semitic peoples from the south Arabian kingdom of Saba (Sheba) which began about 1000 B.C., resulted in the supersession and absorption of the Hamitic inhabitants of the territory, who were probably not indigenous. These Semitic invaders with their well developed culture established the kingdom of Aksum (*q.v.*), which by the end of the 4th century A.D. had become rich and powerful and had gained effective control over the northern stretches of the Ethiopian plateau (now Eritrea) and the Ethiopian province of Tigré. Aksum controlled the eastern lowlands and an important trade route led from the Aksumite port of Adulis (near modern Zula on the Gulf of Zula) to the city of Aksum. This route is now marked by the remains of Aksumite towns and buildings at Coloe, Matara and Yeha.

Aksum, after extending its power at times as far afield as Egypt and Yemen, began to decline into obscurity in the 6th century A.D., and not much is known of the developments during the following three or four centuries. By A.D. 950, however, a kingdom of Ethiopia seems to have been established and to have held sway to a fluctuating extent over the whole plateau country and the Red sea coast line. The centre of control of the kingdom moved at various times to different locations on the plateau and the town of Aksum retained only a religious and historic significance. The territory which is now Eritrea was usually administered by a governor holding the title of *bahr nagash*, appointed by the emperor and controlling the territory on the Eritrean side of the Mareb river. His writ nominally ran over the highlands and the Red sea coast, but off the plateau it appears to have been only partially effective. The Red sea coast, moreover, was the battleground of the Portuguese and the Turks. In 1557, the latter, after having first occupied the Dahlak islands for several years, occupied Massawa and maintained, interruptedly, a garrison there for more than three centuries. In the 16th century, both Ethiopia and Eritrea were affected by the invasions of the Muslim imam of Harar, Mohammed ibn Ibrahim (the Gran). After the expulsion of Gran's forces, the coastal areas of Eritrea were occupied more extensively by the Turks.

The gradual growth in Ethiopia of a kingdom of Tigré, nominally subject to but frequently warring with the titular emperor, brought about the virtual supersession and disappearance of the *bahr nagash*. From then the history of Eritrea, or that part of it subject to Ethiopian rule, became merged in that of Tigré. The first powerful ruler of Tigré was Mikael Suhul, who in 1746 rebelled against the emperor Yasu II. He and his successors ruled a united Tigré, with intermissions, for more than a century. In 1820 the Egyptians invaded the Sudan and raided deep into the Eritrean lowlands. The Turks also penetrated inland from Massawa.

By this time, the project of the Suez canal had drawn the attention of the great powers toward the Red sea area. The Napier expedition to Magdala in 1868 (*see* NAPIER OF MAGDALA, ROBERT CORNELIS) and the suicide of the emperor Theodore II of Ethiopia are part of the history of that country. After Theodore's death, a struggle ensued for the crown. Finally, Kassa, a descendant of Mikael Suhul, was crowned on Jan. 21, 1872, as the emperor of Ethiopia with the title of John (Johannes) IV. He was the first king of Tigré to ascend the imperial throne. His inheritance was troublesome, and the portents were ominous. The activities of Wolde Mikael, who had been governor of what is now the Eritrean plateau for 12 years, caused anxiety; he was flagrantly disobedient and he ravaged the country. There was also the growing menace of Egypt, which had acquired, by *firman* from the Sublime Porte (ignoring protests from the Ethiopians), the port of Massawa, in 1865. In 1867 Werner Munzinger, a Swiss merchant and scholar who had officiated as both a French consular agent and a British agent, was appointed governor of Massawa by the Egyptians. The Egyptians advanced to Keren, to the Bogos country, occupying country inland from Massawa, and fought on toward the high plateau and Danakalia. Allied to them in the highlands was Wolde Mikael. Munzinger was killed by tribesmen when he landed at Danakil (1875). On the same day a force of Egyptians which had reached the Mareb was annihilated by Ethiopian forces. Shortly afterward the Egyptians were defeated in a decisive battle at Gura (1876), but they continued to hold out in Keren. Attempts at mediation between Egypt and Ethiopia by Col. (later Gen.) Charles Gordon, governor of the Sudan, failed, and the territory was practically an Egyptian province for a few years. In 1879 Wolde Mikael was imprisoned. Ras Alula, a faithful general of the emperor, assumed the governorship making his headquarters in Asmara, then a tiny village.

Italian Domination.—The Sudan was in upheaval, and the Italians had entered the scene. Between 1869 and 1880 the Italian Rubattino Navigation company purchased from the local sultan stretches of Red sea coast adjoining the village of Assab, and in 1882, these acquisitions were transferred to the Italian state. In 1885 Italian troops were landed at Massawa, Assab and other places. There was no Egyptian resistance at Massawa and protests made by the Turks and by the emperor were ignored. Meanwhile the mahdi had expelled the Egyptians from the Sudan. A British mission under Adm. Sir William Hewitt was sent to the emperor in 1884 to secure his co-operation in evacuating the Egyptian garrisons in the eastern Sudan through Eritrea. Consequently a treaty was signed and the Keren area was restored to Ethiopia. The mahdists surged over the Eritrean lowlands, and numerous chiefs appealed to the Italians for aid. The Italian forces systematically spread out from Massawa toward the highlands, ignoring fierce protests from Ras Alula, who massacred a body of Italian troops at Dogali (1887), and again the war clouds gathered. Reinforcements came from Italy, and the emperor John himself descended to the coast to meet the challengers. By now he was harassed on all sides and he turned away from the Italian positions to attack the dervish armies massed on the Sudan border, where he was killed in 1889.

Meanwhile in Ethiopian Shoa, the star of Menelik had risen and he was now acknowledged as emperor by all the prominent chiefs of the empire except Mangasha, the son of John and the faithful Ras Alula. He was crowned Menelik II (*q.v.*) of Ethiopia on Nov. 6, 1889. He had reached agreement with the Italians and in return for the promise of help he ceded lands in the Eritrean region and the Italian forces now spread out to occupy Keren and the highlands up to the Mareb. By the treaty of Uccialli which was signed on May 2, 1889, Menelik recognized the sovereignty of the king of Italy over the "Italian possessions in the Red sea." On Jan. 1, 1890, the Italian colony of Eritrea, named after the Mare Erythraeum of the Romans, was officially proclaimed, and Eritrea became a geographical entity. Controversy, however, soon arose over the meaning of an article in the treaty. The Italians claimed that it meant the emperor had accepted the status of a protectorate for Ethiopia. This interpretation was hotly and justly rejected by Menelik. The dispute, which was exacerbated by a quarrel over the Ethiopian frontier, resentment of Italian aid to Mangasha and their frequent incursions over the Mareb into Tigré, resulted in high tension. The Italians occupied most of Tigré at one time, but Menelik fought back. A succession of incidents culminated in the crushing defeat sustained by the Italian forces at the battle of Aduwa on March 1, 1896. The treaty of Addis Ababa was concluded in Oct., 1896, abrogating the Uccialli treaty, establishing the Mareb as the frontier between Ethiopia and Eritrea and recognizing the full independence of Ethiopia south of the Mareb-Belesa line. The boundaries of Eritrea with the Sudan were later demarcated by mutual agreement, and a period of comparative peace ensued.

In 1935 Eritrea was made the main Italian base for the invasion and conquest of Ethiopia, and the aspect of the colony was transformed for this purpose. Ports, roads, bridges and airfields were

built or enlarged. Men and materials poured in. Overwhelming Italian forces crossed the Mareb on Oct. 3, 1935, and entered Addis Ababa on May 5, 1936, after fierce resistance from the Ethiopian forces. Eritrea, to which was added Tigré, then became one of the six provinces of the new Italian empire, although the Italians can never be considered to have secured complete control over the territory in the face of continued Ethiopian and Eritrean resistance.

World War II and Federation.—After Italy entered World War II British forces advanced from the Sudan early in 1941 and drove the Italians out of western Eritrea. A battle ensued for the strong point of Keren, which controls the main route to the high plateau, and ended in the defeat of the Italians. Asmara was occupied on April 1, 1941; the Italians surrendered in Ethiopia as the result of combined actions by British and Ethiopian forces, and Eritrea came under British military administration. The old boundaries were restored. Until 1944, Eritrea was a centre in Africa for the Allied war effort, but as the war receded these activities died down. This temporary British administration lasted for 11½ years. In the peace treaty of 1947 Italy renounced all right to Eritrea and other colonies in Africa and the question of the future of Eritrea was referred to a four-power commission of investigation (the U.S.S.R., France, Great Britain and the United States). The commission failed to agree and the question went to the United Nations at the end of 1948, where prolonged debates produced no agreed solution. A commission was sent to Eritrea in 1950, and its report revealed a wide divergence of opinion. Eventually the UN general assembly recommended that Eritrea should be federated, as an autonomous unit, with Ethiopia, under the sovereignty of the Ethiopian crown. The United Nations appointed a commissioner to draw up a constitution for Eritrea, in consultation with Ethiopia, to give effect to these recommendations, which were to be made effective by Sept. 15, 1952. The constitution was adopted by an elected assembly and a wholly Eritrean administration took over the reins of government from the British administration which handed over powers and responsibilities on the due date. The Federal act and the Eritrean constitution were signed by Haile Selassie, emperor of Ethiopia, on Sept. 11, 1952, the beginning of the Ethiopian year. (*See* also ETHIOPIA: *History.*)

Administration.—Under the 1952 constitution foreign affairs, foreign and interstate commerce, defense and finance are administered by the Ethiopian government, whose relations with Eritrea are maintained through the emperor's representative in Asmara. Eritrea is represented on the Imperial Federal council in Addis Ababa and sends representatives to the Ethiopian parliament. The Eritrean government has full control over internal affairs through the chief executive elected by the Eritrean assembly. Secretaries nominated by the chief executive are in charge of the departments of government (interior, finance, economic affairs, social service, law, justice and state property). Thence responsibility descends through nominated district and tribal chiefs to subordinate chiefs, headmen and heads of tribal sections. There is a comprehensive judiciary under a president of the supreme court.

The legislature is a single-chamber assembly to which members are elected every four years on the basis of one for every 15,000 persons; for this the country is divided into 66 electoral constituencies. Representatives from Asmara and Massawa are elected by direct ballot but in other parts of Eritrea election is through an electoral college of representatives chosen by the various population groups within each constituency. All male Eritrean citizens with a residence qualification of one year in the constituency are eligible to vote. Asmara, Massawa and Assab are municipalities with mayors at the head of the city administrations.

By the early 1960s, in addition to Koranic schools, there were more than 160 elementary schools, about 30 middle schools, 2 secondary schools, 1 vocational training school, 1 nurses' training school and 1 teachers' training college with a total school population of nearly 26,000 students, including nearly 7,000 girls.

The Economy.—Eritrea is inherently a poor agricultural country with a shortage of water and a shortage of land where there is water. In the highlands, where the agriculture is dry farming, there is a tendency to overstocking. The crop yield is subject to considerable variations due to unreliable rainfall and locust swarms. Domestic animals are the principal source of wealth of the indigenous population. In some areas, however, there are concessions farmed chiefly by Italians, irrigated and operated on modern lines, which produce milk, butter, vegetables, fruit, cotton, sisal, bananas, tobacco and coffee. Examples are at Tessenei, Agordat and Keren, in the valleys between Keren and Asmara and between Asmara and Massawa. A large scale project for an irrigation plan at Zula, south of Massawa, has been financed by the federal government.

Industry is largely based on the processing of agricultural products. Asmara is the main industrial centre concerned with food products, tobacco or drink; leather and leather-products industries are also important. A salt industry is based on the deposits in the Kobar sink, the salt being transported through Mersa Fatma on the coast. There is also an annual production of about 250,000 tons from the salt pans at Massawa and Assab.

Economic relations with Ethiopia are based largely on the export of manufactured goods from Asmara, while Ethiopia supplies in return much of Eritrea's food requirements. Before 1952 Eritrea had always to be subsidized but in the early 1960s it showed slight budget surpluses despite increased expenditure on public and social services. The Ethiopian dollar of 100 cents has formed the currency since federation and banking operations are almost entirely in the hands of the State Bank of Ethiopia.

Communications in Eritrea are well developed. An Italian-built railroad, of gauge 95 cm. (37.4 in.) and of length 191 mi., links Massawa with Agordat through Asmara. The excellent road system, originally of Italian construction, comprises nearly 2,000 mi. of main and secondary roads passable at all seasons. The system links up with Ethiopia to the south through Senafé to Adigrat and Addis Ababa, and to the southwest through Adi Ugri to Aksum, Gondar and Gorgora on Lake Tana. Westward the road system leads to Kassala in Sudan through Keren, Agordat and Tessenei, and to the eastward to Massawa through Ghinda. International and local air services call at Asmara and Assab. Telephone, telegraph and radio communications link the main towns with Asmara and with Ethiopia.

BIBLIOGRAPHY.—N. Pearce, *Life and Adventures, 1810–1819* (1831); P. V. A. Ferrat and J. G. Gallinier, *Voyage en Abyssinie* (1847); T. J. Holland and H. M. Hozier, *Record of the Expedition to Abyssinia* (1870); A. B. Wylde, *The Padua Campaign and After* (1900); G. F. H. Berkeley, *The Campaign of Adowa and the Rise of Menelik* (1935); M. Rava, *L'Eritrea* (1927); A. Pollera, *Le popolazioni indigene dell' Eritrea* (1937); S. H. Longrigg, *A Short History of Eritrea* (1945); *Report on the Disposal of the Former Italian Colonies,* Cmd. 8819 (H.M.S.O., 1953); United Nations, *Report of the U.N. Commission for Eritrea* (1950); G. K. N. Trevaskis, *Eritrea, a Colony in Transition, 1941–52* (1960); Y. A. Haggag, *A Contribution to the Physiography of Northern Ethiopia* (1961). (G. C. L.)

ERLANGEN, a town of Germany, which after partition of the nation following World War II was included in the *Land* of Bavaria in the Federal Republic of Germany. It lies on a fertile plain at the confluence of the Schwabach and the Regnitz 16 km. (10 mi.) N.N.W. of Nürnberg by road. Pop. (1959 est.) 66,319. It is divided into an old town (Altstadt), mostly baroque in style, and a new town (Neustadt). By the market place stand the town hall and the former palace of the margraves of Kulmbach-Bayreuth, now the main building of Friedrich-Alexander university. Founded by the margrave Frederick (d. 1763) in 1742 at Bayreuth, this Protestant university was moved to Erlangen in 1743. The chief industries are the manufacture of electrical apparatus, gloves, hats and drapery. There are cotton mills and breweries—the beer of Erlangen is famous throughout Germany, large quantities being exported.

In 1017 Erlangen was transferred from the bishopric of Würzburg to that of Bamberg; in 1361 it was sold to the king of Bohemia. It was chartered in 1398 and passed to the Hohenzollerns, burgraves of Nürnberg, in 1416. It owes the foundation of its prosperity chiefly to the French Protestant refugees who settled there on the revocation of the edict of Nantes and introduced various manufactures. In 1810 it came into the possession of Bavaria.

Erlangen was for many years the residence of the philosophers Johann Gottlieb Fichte and Friedrich Wilhelm Joseph von Schelling. In World War II it was captured by U.S. forces in April 1945.

ERLANGER, JOSEPH (1874–), U.S. physiologist who with Herbert Spencer Gasser (*q.v.*) received the 1944 Nobel prize in physiology and medicine for their work on functional differentiation of nerves. He was born in San Francisco, Calif., Jan. 5, 1874, and after graduation from the Johns Hopkins university medical school in 1899 became associated with the physiologist William H. Howell. Through his invention of a graphic method for measuring blood pressure and his work on conduction in the heart, Erlanger soon revealed his ability as a physiological investigator. His first appointment as a professor was at the University of Wisconsin, Madison. From 1910 until his retirement in 1946 he was professor of physiology at Washington university, St. Louis, Mo. His interest in circulation of the blood continued for a time and included a study of the mechanism of production of the sounds that are used in measuring blood pressure by the auscultatory method. Then, in response to the new outlook afforded by the rise of electronic methods, he turned to neurophysiology and the work for which he received the Nobel prize. In his later work he described the phenomena taking place in nerve fibres when stimuli have less than threshold strength. He also performed pioneering experiments leading to the modern view that conduction in myelinated nerve fibres takes place in jumps; *i.e.*, that conduction is saltatory rather than continuous. (H. S. G.)

ERLKÖNIG (Erl-King), a mythical character in modern German literature who has no existence in early mythology. The name, which means "alder-king," is linguistically the result of a mistake by J. G. Herder, who in his *Stimmen der Völker in Liedern* (1778) used it to translate the Danish *ellerkonge, ellekonge* or *elverkonge,* which means "king of the elves." The correct German would have been *Elbkönig* or *Elbenkönig.* Herder was probably misled by the Danish word *elle,* which means both elf and alder tree. His mistake has been perpetuated by both English and French translators, and the *Erlkönig* has been personified as an eldritch figure who "rides through the night, in the woodland wild," carrying away children to death. Those unaware of the error have explained the myth as having its origin in ancient tree worship, or in the vapours which rise around alder trees at night. The legend inspired one of Goethe's most famous ballads, which was translated by Sir Walter Scott and set to music by Schubert.

ERMAK TIMOFEEVICH (d. 1584), Cossack leader who played an important part in the Russian annexation of western Siberia and who later became a hero of Russian folklore. In 1579, the merchant and factory-owning Stroganov family enlisted Ermak's help, together with a band of Cossacks, to defend their possessions from inroads by Siberian tribesmen. In 1581, the Stroganovs financed an expeditionary force of 840 men led by Ermak. It set off on Sept. 1 along the river Chusovaya to the Ural mountains. In spring 1582, traveling along the rivers, the force reached the central regions of the Tatar khanate of Sibir whose head, Kuchum, ruled over the local tribes. Khan Kuchum tried to halt and destroy the Cossacks. Though Ermak's forces were far weaker in numbers they were victorious as they carried firearms. On Oct. 26, 1582, Ermak occupied the capital, Kashlyk (or Sibir), on the Irtysh river. Kuchum went off into the steppes, whence he continued the fight. Ermak sent the news of his victory to the Stroganovs and at the same time dispatched his lieutenant, Ivan Koltso, to the tsar, Ivan IV, announcing the proclamation of the latter's sovereignty over Siberia. The tsar sent Ermak another 500 men, led by Prince S. D. Bolkhovski. After the first successes, however, revolts flared on all sides. In Aug. 1584, Kuchum attacked a small party of Cossacks led by Ermak and destroyed it. Ermak, fighting his way to the boats was drowned in the Irtysh, apparently by the weight of the coat of chain mail that the tsar had sent him. In 1585, the remainder of the first expedition was forced to retire beyond the Ural mountains. A small detachment under Ivan Mansurov, which had been sent by the Russian government, was also obliged to retreat. The systematic annexation of Siberia was begun in 1586, when the Russian government learned that the remains of Ermak's and Bolkhovski's forces had had to leave Siberian territory. The town of Tyumen was founded in 1586, that of Tobolsk, near the former capital Kashlyk, in 1587. Tobolsk long remained the military and administrative centre of Siberia. (N. An.)

ERMANARIC (d. between 370 and 376), king of the Ostrogoths, the ruler of a vast empire in the Ukraine (*see* Goths). A list of the chief peoples subject to him has survived, but most of them cannot be identified, so that the exact limits of his country are obscure, though the Dniester is known to have been the boundary between the Ostrogoths and the Visigoths and the Don the boundary between the Ostrogoths and the Alans. It is said on questionable authority that the king belonged to the family of the Amals and that he was the son of Achiulf. All that is certain is that his great deeds made him feared by neighbouring peoples and that when the Huns invaded his territories in the 370s he despaired of resisting them successfully and committed suicide. His kingdom was thereupon destroyed, and his people were subject to the Huns for three-quarters of a century. At an early date he became the centre of popular tradition. According to the 6th-century Goth Jordanes (*Getica,* xxiv), Ermanaric put to death a woman named Sunilda by tying her to two wild horses and driving them apart, because her husband had treacherously deserted him; thereupon her two brothers Sarus and Ammius stabbed him in the side and so caused his death. This story, which is unhistorical, had a profound influence on medieval Germanic literature, including that of England and that of Scandinavia. The form of Ermanaric's name differs in different authors and dialects: "Ermenrichus" in Ammianus Marcellinus (whose book xxxi is the chief source for his career), "Hermanaricus" in Jordanes, "Iormunrekr" in the Norse writers, "Eormenric" in the Anglo-Saxon, etc.

Bibliography.—L. Schmidt, *Geschichte der deutschen Stämme: die Ostgermanen* (1934). For summaries of some of the stories in which Ermanaric figures, *see* H. M. Chadwick, *The Heroic Age* (1912); C. Brady, *The Legends of Ermanaric* (1943). (E. A. T.)

ERMELO, a municipality in northwestern Gelderland, Neth., comprises the three villages of Ermelo, Nunspeet and Elspeet picturesquely situated near the Ijsselmeer. Pop. (1957 est.) 25,774. Ermelo was first mentioned in 855. The church has a Romanesque tuff steeple (11th century) and a 14th-century choir. There are the ruins of the monastery of the knights of St. John (12th to 13th century), two hospitals and a home for the blind. The main industries are farming, chiefly poultry and dairy, dye and metal industries. (H. M. Ms.)

ERMINE, the name applied to several small northern weasels belonging to the genus *Mustela,* especially when they display their white or winter coats, the ermine of fur trade. The original ermine, which furnished the ermine for robes of royalty in Europe, is *M. erminea,* also called the stoat. *See* Stoat; Weasel.

ERMINE STREET, the Saxon name (sometimes spelled Irmine) for at least two Roman roads in Britain. Long unexplained, it is now recognized as being derived from a Saxon group, the *Earningas* (Earn's people), occupying an area on each side of the Cam river in Cambridgeshire, through which the road passed. The ford by which the road crossed the Cam was named Earningaford. As with other Saxon road names, the name, formerly applied to part of the road, has now been given to the road as a whole. The name relates primarily to the Roman road from London (Londinium) to Lincoln (Lindum) and thence to York (Eburacum), but it is used also for the road from Silchester (Calleva Atrebatum) to Cirencester (Corinium Dobunnorum) and Gloucester (Glevum). No record has survived of the Roman name for these roads.

The London–York road left London by Bishopsgate. As a through road it would have crossed the city from the Roman bridge, passing probably along the east side of the basilica and forum. Its course is preserved through London in modern streets (Shoreditch High street, Stamford Hill) to Upper Edmonton. It then goes by Ware, Braughing (an important Roman intersection), Royston, Arrington (Erningetone, in the Earningas territory), Godmanchester (Durovigutum) and Chesterton (Durobrivae), a small town and centre of the Castor pottery industry. North of

Chesterton it crossed the Nen river by a bridge, then forked northwestward from King street, by way of Stamford, Great Casterton (site of an early fort and town), and through the high ground east of Grantham to Ancaster (Causennae) and Lincoln. Thence it continued a very direct course northward, crossing the river Humber at Brough (Petuaria) and after a few miles dividing for York and the northwest on the one hand and Malton and the north on the other.

The Silchester-Gloucester Ermine Street was so called as early as the mid-14th century by the chronicler Ranulf Higden (*Polychronicon* i, Rolls Ed. 41 [1869] pp. 42 *et seq*). The name Ermin street is actually applied to the western portion beyond Speen. From the west gate of Silchester it ran westward to Thatcham, where it crossed the Kennet river, then north of Newbury to Speen (Spinae, a Roman site not certainly located) and northwestward by way of Wickham and Baydon to Wanborough (Durocornovium). It then crossed the river Thames just below Cricklade to Cirencester (Corinium Dobunnorum). Most of the Speen-Cirencester stretch coincides with the modern road, as does the final piece over the Cotswolds to Gloucester. Descending the scarp steeply at Birdlip, the road is aligned not on the Roman city but on the original fortress at Kingsholme to the north, the later city being linked to it by a diversion.

Ermine street (*i.e.*, the London–York road) was one of the four great roads of England which figured in the laws of William the Conqueror as being subject to a special peace under heavy penalties for the protection of travelers using them.

BIBLIOGRAPHY.—A. Mawer and F. M. Stenton, *Place-names of Bedfordshire and Huntingdonshire*, pp. 2–4 (1926); I. D. Margary, *Roman Roads in Britain*, vol. i (1955) and vol. ii (1956). *See* also, Ordnance Survey Map of Roman Britain (1956). (WI. F. G.)

ERMOLDUS NIGELLUS (fl. *c.* 830), author of an adulatory poem in Latin of scant literary merit on the emperor Louis I the Pious. Probably Aquitanian by birth, he was a scholar, possibly a monk or chaplain, at the court of the emperor's son Pepin I, king of Aquitaine. He bore arms in an expedition against Brittany (823) and was ordered into exile at Strasbourg by Louis the Pious after 824. At Strasbourg he addressed two elegies to Pepin and wrote the biographical poem on Louis, which ran to four books. This eulogy apparently bought his pardon and he may have returned to Pepin's court some time after 830. He died before 835.

See E. Faral (ed.), *Ermold le Noir—poème sur Louis le Pieux et épitres au roi Pepin* (1932). (J. DE.)

ERNAKULAM, a town in Kerala, India, is centred on a backwater opposite to, and 2 mi. N.E. of, Cochin (*q.v.*) harbour with which it is connected by an excellent road-bridge and by frequent ferry services across the lagoon. Pop. (1961 est.) 120,000. Cochin airport is 4 mi. distant. Formerly the town was the capital of Cochin state, and since the integration of the states in 1949 it has been the seat of the Kerala High court. It is an educational centre with more than 2,000 undergraduates in five colleges, three of which are Roman Catholic, all affiliated to Kerala university; of these one is exclusively for women. It is also an industrial centre with large mills manufacturing oil, cocogem, soap, glycerine and other toilet preparations, and many subsidiary industries connected with coastal fishing and coconut farming. (G. KN.)

ERNE, the name of a river and two lakes in northwest Ireland. The area of the drainage basin is 1,689 sq.mi. and the length of the river Erne is 64 mi. The river rises in Lough Gowna, County Longford (Republic of Ireland), flows north through Lough Oughter and then enters County Fermanagh (Northern Ireland) which contains the two lakes, Upper and Lower Lough Erne. The river re-enters the Irish republic beyond Belleek, reaching Donegal bay below Ballyshannon. Upper Lough Erne, a shallow sheet of water 12 mi. long, is thickly strewn with wooded islands. The lower lough is larger, 18 mi. long and up to 5½ mi. wide, and has fewer islands. Boa Island, in the north, is 4 mi. long and carries a picturesque road which elsewhere follows the lake shore. Two small islands, Devenish Island and White Island, are famous for remains of the early Christian period. Between the lakes the river winds past Enniskillen, the county town of Fermanagh.

The lakes lie in an almost level carboniferous limestone plain which is an extension of the central plain of Ireland. Although formed by the solution of the limestone, their basins are so blanketed with heavy boulder clay (whose drumlin forms account for the tortuous channels and lake shores) that heavy rains quickly fill the lakes. The level of the upper lake in particular has been liable to fluctuations because of the constricted outflow through the winding interlough channel. Flood-relief schemes have been carried out at various times; one of these, completed in 1959, was designed to control the water supply to the two hydroelectric power stations at Ballyshannon and Cliff. The cost was shared by the Northern Ireland government and the Electricity Supply board of the Republic of Ireland. The lake levels are controlled by barrages constructed at Portora (Enniskillen) and at Cliff. As a result of the project, 11,000 ac. around the upper lough no longer suffer from summer flooding and a further 18,000 ac. have been freed from excessive waterlogging. The lakes and waterways provide a fascinating variety of scenic attraction and good fishing. There is a small commercial eel fishery. (E. E. E.)

ERNEST I (1784–1844), duke of Saxe-Coburg and Gotha, was born at Coburg on Jan. 2, 1784, the son of Francis, duke of Saxe-Coburg-Saalfeld. On his father's death (1806) he succeeded him as Ernest III of Saxe-Coburg-Saalfeld, but the duchy was occupied by Napoleon and Ernest did not recover it until the peace of Tilsit (July 1807). He commanded the Saxon V army corps in 1813–14 and reduced Mainz by blockade; he also took part in the campaign of 1815. At the congress of Vienna he received the principality of Lichtenberg, which he sold to Prussia in 1834.

In 1826, after the death of his brother-in-law Frederick, the last duke of Saxe-Gotha (1825), Ernest gave up Saalfeld and received Gotha, becoming Ernest I of Saxe-Coburg and Gotha. In 1821 he had given a constitution to Coburg, but he did not change the traditional system of estates in Gotha. He died on Jan. 29, 1844. He married twice: (1) in 1817, Louise of Saxe-Gotha, whom he divorced in 1826; (2) in 1832, Mary of Württemberg. Of his sons, Ernest (b. 1818) succeeded him and Albert (b. 1819) married Queen Victoria of Great Britain. His brother Leopold (1790–1865) became king of the Belgians and his sister Victoria (1786–1861) was Queen Victoria's mother.

See G. Hirschfeld, *Die Errichtung des Herzogtums Sachsen-Coburg und Gotha* (1927); *Fürst Metternich und Herzog Ernst von Sachsen-Coburg und Gotha* (1929).

ERNEST II (1818–1893), duke of Saxe-Coburg and Gotha, was born at Gotha on June 21, 1818, the eldest son of Duke Ernest I (d. 1844) and his first wife Louise of Saxe-Gotha. During the years of reaction after 1849 Ernest was almost the only German prince to remain faithful to the liberal and national ideal, offering asylum to political exiles from Prussia and Saxony. In 1852 Ernest gave Gotha a new constitution which in part co-ordinated the administration of his two duchies. After 1860 he became patron and protector of the *Nationalverein* and allowed his court to become the centre of the rising nationalist agitation. In 1861 he concluded a military agreement with Prussia, subordinating his troops to Prussian command, in case of war. The democratic leanings of the Coburg court seriously embarrassed Bismarck, whose policy was further hampered when Coburg became the headquarters of the party in favour of Frederick, duke of Augustenburg, during the Schleswig-Holstein crisis (*see* SCHLESWIG-HOLSTEIN QUESTION). In 1863 Ernest was present at the *Fürstentag* in Frankfurt and began to correspond with the Austrian court, where his cousin Alexander, Graf Mensdorff, was his minister. In 1866, however, after vain attempts at mediation between Prussia and Austria, he put his troops under Prussian command just before the battle of Langensalza.

In 1866, upon the formation of the German empire, Ernest's political role ended. That it had been important is shown by the comment of the emperor William I: "To him in no small degree was due the establishment of the empire." A man of varied tastes, he composed several operas and songs and was a keen sportsman. Ernest died at Reinhardsbrunn on Aug. 22, 1893. In 1842 he had married Alexandrine of Baden; there were no children, and the succession passed to Alfred duke of Edinburgh (d. 1900), second son of Ernest's younger brother Albert, the British prince consort.

Duke Ernest published his memoirs, *Aus meinem Leben und aus meiner Zeit*, in three volumes (1887–89).

See E. Tempeltey, *Herzog Ernest von Koburg und das Jahr 1866* (1898) and *Gustav Freytag und Herzog Ernst von Koburg im Briefwechsel 1853–1893* (1904).

ERNEST AUGUSTUS (1771–1851), king of Hanover from 1837 to 1851, and duke of Cumberland, was born at Kew on June 5, 1771, the fifth son of George III of England. He studied at Göttingen, entered the Hanoverian army and served as a leader of cavalry when war broke out between Great Britain and France in 1793. When Hanover withdrew from the war in 1795 he returned to England, being made lieutenant-general in the British army in 1799. In the same year he was created duke of Cumberland and Teviotdale and granted an allowance of £12,000 a year. A staunch Tory, the duke objected to all proposals of reform, especially to the granting of any relief to the Roman Catholics, and had great influence with his brother the prince regent, afterward George IV, in addition to being often consulted by the Tory leaders.

In 1810 he was severely injured by an assailant, probably his valet Sellis, who was found dead; subsequently two men were imprisoned for asserting that the duke had murdered his valet. Recovering from his wounds, Cumberland again proceeded to the seat of war; as a British field marshal, he was in command of the Hanoverian army during the campaigns of 1813 and 1814, being present, although not in action, at the battle of Leipzig. In May 1815 Ernest married his cousin, Frederica (1778–1841), daughter of Charles II duke of Mecklenburg-Strelitz and widow of Frederick, prince of Solms-Braunfels, a union which was very repugnant to Ernest's mother, Queen Charlotte, and was disliked in England, where the duke's strong Toryism had made him unpopular. The duke resented the refusal of parliament to increase his allowance from £18,000 to £24,000 a year, and retired for some years to Berlin. On the accession of George IV he returned to England but he ceased to play an important part in politics after the accession of William IV (1830).

When William died in June 1837, the crowns of Great Britain and Hanover were separated; and Ernest, as the nearest male heir of the late king, became king of Hanover. He canceled the constitution which William had given in 1833, and the constitution which he sanctioned in 1840 was characteristic of his own illiberal ideas. His reign was a stormy one, and serious trouble between king and people had arisen when he died at Herrenhausen on Nov. 18, 1851 (*see* HANOVER: *History*).

ERNEST AUGUSTUS, DUKE OF CUMBERLAND (1845–1923), only son of George V of Hanover, who was deposed in 1866, following the events of the Seven Weeks' War, and grandson of Ernest Augustus (*q.v.*), king of Hanover from 1837 to 1851, was born at Hanover, Ger., on Sept. 21, 1845, and died at Gmünden, Aus., Nov. 14, 1923. He lived mainly in Austria but on his father's death (1878) he maintained his claim to the Hanoverian throne. He bore the titles of duke of Cumberland and duke of Brunswick-Lüneburg. In 1884 on the death, without issue, of his kinsman William, the last sovereign duke of Brunswick, he also claimed that duchy. Inspired by Bismarck, the federal diet rejected his claim and installed Prince Albert of Prussia as regent of Brunswick. After the latter's death the duke was again excluded (1907) and the regency continued.

In 1902, Ernest Augustus had told the emperor that he was prepared to accept the imperial regime in Germany, and in May 1913, the Guelph-Hohenzollern feud was healed by the marriage of his son ERNEST AUGUSTUS (1887–1953), who had been educated at Eton and held a commission in the Bavarian army, to William II's only daughter, Victoria Louise. Ernest Augustus then renounced his rights in favour of his son, the federal council lifted its veto and the young prince became a sovereign ruler as duke of Brunswick (Nov. 1, 1913). He abdicated, with other German princes, on Nov. 8, 1918. A Titles Deprivation act in 1917, followed by an order in council in 1919, deprived father and son of all their titles and honours in the United Kingdom.

The son of the last duke, another ERNEST AUGUSTUS (1914–), who was educated at Oxford, served in the German army until removed after the attempt on Adolf Hitler's life (July 1944) and was wounded on the eastern front, became a lawyer and administered the ducal estates. These were adversely affected by the events of World Wars I and II, and in 1930 and again in 1950 the family was obliged to sell part of the famous Guelph art treasures.

In 1955, as a lineal descendant of Princess Sophia (*q.v.*), electress of Hanover (and as a Protestant), Prince Ernest successfully claimed British citizenship under the Act of Naturalization of 1705 which had naturalized Princess Sophia "and the issue of her body, and all persons lineally descending from her, born or hereafter to be born." This decision was upheld by the house of lords in 1956. It thus became possible for some 500 other Protestant descendants of the electress to claim British citizenship, while the way was opened for Prince Ernest to petition the crown for a revival of the dukedom of Cumberland, since he was too young to have borne arms against Great Britain in World War I.

(S. B.-R. P.)

ERNESTI, JOHANN AUGUST (1707–1781), German theologian and philologist, whose fame rests upon his works in two areas of scholarship, classical philology and biblical interpretation, was born Aug. 4, 1707, at Tennstädt in Thuringia. He was educated at the celebrated Saxon cloister school of Pforta and at the universities of Wittenberg and Leipzig. After minor appointments he was named professor extraordinarius of ancient literature in the University of Leipzig (1742), professor ordinarius of rhetoric (1756) and professor ordinarius in the faculty of theology (1759). He died at Leipzig on Sept. 11, 1781.

As a philologist, Ernesti edited writings of Xenophon, Cicero, Suetonius, Tacitus, Aristophanes and Homer. Thus he contributed to the resuscitation of classical studies during the 18th century. His methods as a philologist shaped his understanding of biblical scholarship as well. He argued that the Bible should be interpreted according to the canons of interpretation that apply to any ancient writing; one of these canons, to be sure, was that the interpreter should recognize the special qualities of the document before him. The discussion provoked by Ernesti and by his contemporary Johann David Michaelis (*q.v.*) over the validity of such historical and critical study of the Bible lasted throughout the 19th century and into the 20th.

Ernesti's chief work, *Institutio Interpretis Novi Testamenti*, was translated into English by Moses Stuart as *Elements of Interpretation*, 2nd ed. (1824), and by Charles H. Terrot as *Principles of Biblical Interpretation* (1832). (J. J. PN.)

ERNST, PAUL CARL FRIEDRICH (1866–1933), German author, best known as a writer of short stories and a stimulating and uncompromising critic, though he himself would have preferred fame as a dramatist. Ernst was born at Elbingerode, Saxony, March 7, 1866. After studying theology at Göttingen and Tübingen, he went to Berlin where he was attracted to economics. Concerned by the social problems of the metropolis, he gave up his studies to become a Social Democrat journalist, but later turned to literature as the best means of counteracting the materialism of his day. His ideas on the nature and purpose of literature are challengingly expressed in *Der Weg zur Form* (1906) and *Ein Credo* (1912). His ambition was to write great drama and between 1905 and 1925 he wrote 18 plays, including *Canossa* (1908), *Preussengeist* (1915), *Yorck* (1917) and *Chriemhild* (1922). None achieved popular success; they were criticized as "cold" because of their lack of contact with contemporary life. For his plays and his theories, however, Ernst is justly considered the protagonist of a classicist reaction against literary naturalism.

Inspired by the Italian *novella*, Ernst wrote many lively and interesting short stories, on which his fame chiefly rests. They are collected in volumes such as *Die Prinzessin des Ostens* (1903), and *Spitzbubengeschichten* (1920). He also wrote novels, some autobiographical, like *Der schmale Weg zum Glück* (1904), some historical, like *Der Schatz in Morgenbrotstal* (1926). In 1920 appeared his *Erdachte Gespräche*, imaginary conversations on literary and philosophical problems. His last extensive work was *Das Kaiserbuch* (1923–28) a national epic celebrating the German emperors. Ernst died, May 13, 1933, in St. Georgen, Aus.

See R. Faesi, *Paul Ernst und die neuklassizistischen Bestrebungen im*

Drama (1913); G. Rottcher, *Paul Ernst in seiner Auseinandersetzung mit dem deutschen Idealismus* (1953). (W. W. Cs.)

ERODE, a town in Coimbatore district, Madras state, India, lies 63 mi. N.E. of Coimbatore, on the right bank of the Cauvery river. It is a railway junction where the line from Coimbatore to Salem joins that from Tiruchirapalli 88 mi. S.E. Pop. (1961) 74,102. Its chief industries are cotton ginning and pressing based on the proximity of the raw material, and it has for long been a centre of trade. It was raided successively by the Maratha, Mysore and British troops, and several times laid waste, but because of the agricultural wealth of the region it always rapidly recovered upon the establishment of peace. It has an industrial school and a small college. (G. Kn.)

EROS, in Greek mythology, god of love. He is not mentioned in Homer; in Hesiod he is a primeval god, son of Chaos. This cosmic Eros is further elaborated in Orphic cosmogony. A cult of Eros existed at Parium and at Thespiae in Boeotia. He was god not simply of passion but of fertility. Later mythology makes him the son of Aphrodite by Zeus, Ares or Hermes. His brother is Anteros, the god of mutual love, who is sometimes described as his opponent. The chief associates of Eros are Pothos and Himeros (longing and desire); he himself is in constant attendance on the goddess Aphrodite. Later writers (Euripides being the first) assumed the existence of a number of Erotes (like the Roman Amores and Cupidines). Plato and other philosophers allegorize Eros. In Alexandrian poetry he degenerates into a mischievous child. (*See* also Cupid.)

ALINARI

EROS: ROMAN REPLICA OF A 4TH-CENTURY B.C. SCULPTURE BY LYSIPPUS. IN THE LOUVRE, BORGHESE COLLECTION

In archaic art Eros is represented as a beautiful winged youth and tends to be made younger and younger until by the Hellenistic period he is an infant. The rose, the hare, the cock and the goat are frequently associated with him. The most celebrated statue of him, by Praxiteles, was at Thespiae.

See L. R. Farnell, *The Cults of the Greek States*, ii, p. 625 (1896-1909); H. J. Rose, *A Handbook of Greek Mythology* (1933).

EROS, an asteroid of exceptional interest because of its close approaches to the earth. In 1898 the number of known asteroids had mounted to more than 400; and as there was considerable monotony in their appearance and in the nature of their orbits, many astronomers were beginning to question whether the time spent in searching for them and in calculating their movements was worthwhile. Then came a discovery, which might never have been made had the search been abandoned, that gave full compensation for all the time thus spent. On Aug. 13, 1898, G. Witt, at the Urania observatory, Berlin, discovered photographically a faint asteroid of the 11th magnitude that was retrograding at the unprecedented rate of half a degree per day; this unusual motion caused it to be widely observed, and after a short interval Dr. Berberich of the Berlin Rechen-Institut published elements which proved to be very near the truth. The new body was designated at first 1898 DQ, but subsequently received the permanent number 433 and the name Eros (masculine names are given only to asteroids with very exceptional orbits, feminine ones being the rule).

Its period of revolution is 643 days, as compared with 687 for Mars. It was already known that a few asteroids come nearer to the sun at their perihelia than Mars at its aphelion; but Eros' mean distance is considerably less than that of Mars. As its eccentricity is 2/9, its least distance from the sun is only 1.13 astronomical units, so that it can approach the earth within 14,000,000 mi. The distance is rendered less by the fact that the point where Eros crosses the plane of the earth's orbit (descending node) is distant only $2\frac{1}{4}°$ from the perihelion point; since the orbit is inclined some $10\frac{3}{4}°$ to that of the earth, the least distance would be twice as great if the node were 90° away from the perihelion.

Very near approaches of Eros to the earth are rare; they occur only when the planet passes perihelion, about Jan. 22 of any year. It was unlucky that the conditions were the most favourable possible in Jan. 1894, four and one-half years before the discovery; indeed subsequent examination of the immense stores of photographic plates exposed at Harvard College observatory revealed 17 images of Eros at the opposition of 1893–94, and four images on plates taken at Arequipa, Peru, at the opposition of 1896. These images were used to give a better determination of the orbit, but the opportunity was missed of obtaining observations for the solar parallax.

The synodic period of Eros (that is, the average interval between one opposition and the next) is $845\frac{1}{2}$ days, or 2 years and 115 days. Three synodic periods fall short of seven years by about three weeks; it follows that there was an approach to favourable conditions at the opposition at the end of 1900; and a parallax campaign was carried on at that time at the leading observatories of the northern hemisphere, the southern hemisphere being excluded because of the high north declination of Eros (*see* Parallax).

The next favourable opposition occurred in Jan. 1931, when Eros approached the earth to a distance of 16,200,000 mi., the closest approach since its discovery in 1898. A commission of the International Astronomical union under the chairmanship of Sir Harold Spencer Jones, then astronomer royal, made elaborate plans for an international program of observation with the object of taking full advantage of the favourable opposition for a new determination of the solar parallax. As an important preparatory operation the positions of numerous stars along the path of Eros in the sky were specially observed at observatories north and south of the equator in order to provide an adequate number of known star positions to which the position of Eros could be referred on the photographic plates to be taken during the campaign. The observations, at 24 observatories with 30 different telescopes, yielded 2,847 photographic plates for measurement. In 1941, as the result of the extensive discussion of this material, Jones published for the solar parallax the value 8″.790 with probable error 0″.001.

H. N. Russell showed soon after the discovery that the effects of the earth's attraction on the motion of Eros are considerable. The fact that the period of Eros is to that of the earth approximately in the ratio 7:4 causes important long-period terms in the orbital longitude. The principal effect is a heliocentric oscillation with coefficient 747″ and period 41 years. The geocentric displacement at close approaches is magnified about sixfold. Consequently, this shift may give an accurate determination of the earth's mass in terms of the sun's mass, and hence of the sun's distance in terrestrial units. The method is known as the dynamical method of obtaining the solar parallax. In 1950 E. Rabe published the result of a solution in which he made use of observations made between 1926 and 1945. For the earth-moon system he found the mass 1/328,452 times the sun's mass, which corresponds to 8″.7984 for the solar parallax with probable error 0″.0004, in excellent agreement with an earlier result by E. Nooteboom from observations of Eros in the years 1893 to 1914 (8″.799 with probable error 0″.001). Prediscovery observations, noted above, had been used in Nooteboom's work; they contributed significantly to this determination of the solar parallax.

The discordance between the geometrical and the dynamical determination has not yet been cleared up. It is difficult to point to a weakness in the dynamical determination. A possible cause of systematic errors in the geometrical determination might be a difference in colour between Eros and the average reference star. This point was examined by Jones, who concluded that the effect is small. This conclusion is supported by the internal consistency that is a striking feature of the discussion.

The same observations from which the solar parallax is determined during a favourable opposition may also be used for an evaluation of the moon's mass in terms of the earth's mass. Each sidereal month, as the moon describes its orbit around the earth,

the centre of the earth describes a much smaller similar orbit, with radius about 3,000 mi., about the centre of mass of the earth-moon system. If the mass ratio moon/earth used in the ephemeris calculation needs correction, differences between observation and ephemeris should reveal this by an appropriate wavy character. The observations during the 1930–31 opposition gave for the moon's mass 1/81.30 times the earth's mass. This is the mean result from several different treatments of the same material. The result depends in part on the value of the solar parallax adopted.

Attractions by Mercury, Venus and Mars affect the motion of Eros in a way similar to the attraction by the earth. Corrections to the masses of these planets were obtained by Rabe simultaneously with the determination of the mass of the earth-moon system. The reciprocals of the masses with their probable errors so found, in terms of the sun's mass, are: Mercury 6,120,000 ± 43,000; Venus 408,645 ± 208; Mars 3,110,000 ± 7,700. The mass determinations of Mercury and Venus are more accurate than any previous determination; that of Mars is in good agreement with that from the periods and mean distances of the orbits of the Martian satellites.

Another close approach, almost equal to that in 1931, occurred in 1938, when no repetition of the large-scale effort of seven years earlier was carried out. The next favourable opposition, in 1975, will bring Eros near its absolute minimum distance of 14,000,000 mi. from the earth.

At the most favourable oppositions Eros appears somewhat brighter than the 7th magnitude; at the least favourable ones it is somewhat fainter than the 11th. Since shortly after its discovery it had been known to exhibit variations in brightness in a complicated pattern. Sometimes the brightness does not change appreciably; at other times the range is as much as 1.5 magnitude, corresponding to a ratio of 4:1. The variation has been found to be due to the irregular shape of Eros combined with a period of rotation of 5 hrs. 16 min., during which there are two unequal maxima and minima. Its appearance in long-focus telescopes viewed by experienced observers at the time of closest approach in 1931 confirmed this interpretation. The observations led to the conclusion that the maximum cross section of Eros is about 15 mi. by 4 mi. The character of the light variation changes with the angle between the axis of rotation of Eros and the direction from the earth to Eros. (A. C. DE C.; D. BR.)

EROSION OF LAND. The process of removal of portions of the earth's crust by natural agencies, of which the most important is water. The wastage of the sea coast, or coast erosion, is described in the article BREAKWATER; it is brought about in the main by the action of the sea waves but also, in part, by the disintegration or degradation of cliffs by atmospheric agents such as rain and frost and by tidal scour. In rivers and estuaries the erosion of banks is caused by the scouring action of the river stream, particularly in times of flood and, in the case of estuaries, also by the tidal flow on the ebb tide when river and tidewater combine in their erosive action. (*See* RIVER AND RIVER ENGINEERING.) The land surface generally is subjected to a continuous process of erosion by the action of rain, melting snow, ice and frost, the resulting detritus and sediment being carried by small streams into the rivers and thence to the ocean (*see* GEOLOGY: *Physical Geology; Geomorphology;* SOIL EROSION AND CONSERVATION). In some arid and desert tracts wind has an important effect in bringing about the erosion of rocks by driving sand; and the surface of sand dunes, where not held together and protected by vegetation, is subject to erosion and change by the drifting of blown sand. *See* WIND EROSION AND DEPOSITION; *see* also references under "Erosion of Land" in the Index volume.

ER RIF (AR RIFF now ALHUCEMAS), a Mediterranean province of Morocco embracing the central part of the Rif mountains, includes three types of landscape: (1) the lofty highlands (Mt. Tidiguin, 8,045 ft.) with a heavy rainfall and thickly forested with cedars; (2) lower ranges fringing the coast with steep cliffs; and (3) the arid eastern Rif, with the raised Targuist basin draining to Alhucemas bay by the Rhiss and Nekor wadies. On the bay lies Achdir, the capital of Abd-el-Krim (*q.v.*) from 1921 until its capture in 1925 by the Spaniards. The latter then founded the town and port of Alhucemas (*q.v.*), the chief town of the province. Berber is spoken almost throughout the province, whose population at the 1960 census was 191,495. (G. L. M.)

ERROLL (OR ERROL), **FRANCIS HAY,** 9TH EARL OF (*c.* 1564–1631), Scottish nobleman, a leader of the militant Roman Catholic party in Scotland, succeeded to the earldom in 1585. He was converted to Roman Catholicism at an early age and soon became an ally of the earl of Huntly (*see* HUNTLY, EARLS AND MARQUESSES OF). Between 1588 and 1597 he and his associates were involved in a series of treasonable activities whose object was the restoration of Roman Catholicism in Scotland, and ultimately in England, by the aid of Spanish money and arms. Twice, in 1589 and 1594, they engaged in armed rebellion and in the latter year won at Glenlivet a victory over the royal forces commanded by the earl of Argyll. Both these rebellions, however, dispersed when the king took the field in person. On several occasions Erroll and his friends were forced to make a nominal submission to the Reformed Kirk (Church); at other times they were ordered into ward and for a short period they were exiled and their estates and titles forfeited. But in spite of their uniform lack of success and the alarmed hostility of the English government, James VI's friendship for Huntly, his determination to preserve a Catholic party as a counterpoise to the ministers of the kirk and his fear of endangering the English succession by alienating Roman Catholic opinion prevented any violent retribution overtaking the earls. After his restoration in 1597 Erroll played little part in public affairs, though he retained the king's favour and in 1604 was one of the commissioners appointed to treat for a union with England. His religious convictions remained unchanged and from 1608 to 1611 he was imprisoned as an "obstinate papist." He died at Slains, Aberdeen, on July 16, 1631. (J. K. BA.)

ERRORS, THEORY OF. A physical occurrence can seldom if ever be reproduced exactly. In particular the carefully staged physical occurrences known as scientific experiments are never capable of exact repetition. For example, the yield of a product from a chemical reaction may vary quite considerably from one occasion to another due to slight differences of conditions such as temperature, pressure, concentration and agitation rate. Other errors will be introduced because of the impossibility of obtaining an entirely representative sample of the product for analysis. Finally, no method of chemical analysis of the sampled material will be exactly reproducible. The application of probability theory to the treatment of these various discrepancies is called the theory of errors.

Frequency Distribution.—The characteristics of a varying quantity such as the yield of a chemical reaction or the level of blood sugar of a mouse may be described by a frequency distribution of values. In fig. 1 the step diagram is a frequency distribution of 200 percentage yields from a chemical process for which conditions were maintained as nearly constant as possible. The heights of the shaded rectangles indicate the frequency of occurrence of the various yields. The diagram shows that on 7 occasions the yield was between 61 and 62, on 17 occasions between 62 and 63, on 29 occasions between 63 and 64, and so on.

If the number of yield values available were to be increased a finer interval of classification could be employed, and in the limit the diagram might approach a smooth curve such as that shown by the continuous line. The smooth curve represents the distribution of the hypothetical infinite population of outcomes that could result from the experiment.

Probability Density Function.—If this population distribution were to be scaled so that the total area under the curve was equal to unity, then the proportion of yield values greater than any value y would be equal to the area to the right of y. For instance, the heavily shaded area in fig. 1 would provide the proportion of yields greater than 68. The curve so standardized is called the probability distribution of y, or more precisely the probability density function.

Of particular interest is the mean value of the population, denoted here by μ and sometimes thought of as the "true" value. The amount $\epsilon = y - \mu$ by which a particular observation y differs from μ is called the error in y. The distribution of the error

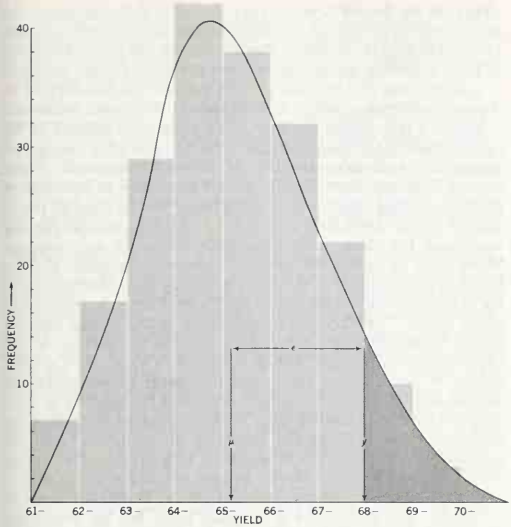


FIG. 1.—A FREQUENCY DISTRIBUTION AND THE UNDERLYING PROBABILITY DISTRIBUTION (*see* TEXT)

ϵ is thus the distribution of y itself referred to the origin μ. If proper precautions are taken in the conduct of the experiment, an observed outcome y may be regarded as made up of the true value μ together with an error ϵ randomly drawn from the error distribution.

Gaussian Distribution.—When the over-all error ϵ is compounded of a number of component errors $\delta_1, \delta_2, \ldots, \delta_s$ it is possible to deduce on theoretical grounds the approximate nature of the distribution of ϵ. The Central Limit Theorem originally enunciated by P. S. Laplace in 1812 says that if

$$\epsilon = \alpha_1\delta_1 + \alpha_2\delta_2 + \cdots + \alpha_s\delta_s$$

where $\alpha_1, \alpha_2, \ldots, \alpha_s$ are constants, then, almost irrespective of the nature of the distribution of the component errors $\delta_1, \delta_2, \ldots, \delta_s$, the distribution of the over-all error ϵ will tend to a particular form as the number s of error sources increases. This limiting form is called the normal or Gaussian distribution or sometimes the error function. The ordinate $p(\epsilon)$ of this probability density function is given by

$$p(\epsilon) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\epsilon}{\sigma}\right)^2}$$

The distribution is shown in fig. 2. The parameter σ which determines the spread of the error distribution is called the standard deviation (*q.v.*). Although this normal distribution is never exactly realized in practice, nevertheless it often provides a sufficiently close approximation to the distribution of errors arising from multiple sources.

Application to Experimental Design.—The influence and importance of the theory of errors in the design of experimental investigations and the analysis of the resulting data may be illustrated by an example.

A particular functional form $\mu = f(x, \theta)$ (for example, $\mu = 1 - e^{-\theta x}$) is tentatively entertained by an experimenter as describing the relationship between the mean value μ of chemical yield and the temperature x. For simplicity only a single variable (temperature) and a single parameter θ are considered but a similar development is available for any number of variables and any number of parameters. The experimenter's objectives are to check the adequacy of the proposed form of relationship, to estimate the value of the unknown parameter θ, and to determine the accuracy of this estimate. Typically he performs a modest number n of experiments at selected temperatures, taking care to repeat at least some of the experimental runs under truly independent conditions.

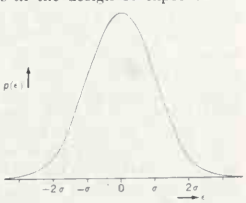


FIG. 2.—THE NORMAL ERROR DISTRIBUTION

Two fundamental problems confront the experimenter: (1) how, after the experiments have been run, he can extract all the information relevant to his objectives from the partially discrepant data; and (2) how, in advance of the experiment, he should choose his experimental conditions (the levels of temperature x) so that the resulting data will throw most light on the questions at issue. The first is the problem of statistical analysis of the data. The second is the problem of the statistical design of the group of experiments. Developments in the theory of errors provide solutions to both these problems.

The error committed in the i^{th} experimental run is $\epsilon_i = y_i - f(x_i, \theta)$. If the magnitude of the error in one experimental run does not influence that in another and if the errors all follow the same distribution then the logarithm of the probability density for all the errors simultaneously is given by the sum $\sum_i \ln p(\epsilon_i) = \sum_i \ln p\{y_i - f(x_i, \theta)\}$. This expression is a function of the observations y_i, the levels x_i of the temperature and the value θ of the parameter. It includes within it the postulated functional form and the probability mechanism necessary to relate these quantities. After the experimental runs are performed, the levels x_i of temperature actually employed and the levels y_i of yield actually observed are known and fixed; $\sum \ln p(\epsilon_i)$ is then a function $l(\theta)$ of θ alone, called the log-likelihood function. As was first pointed out by R. A. Fisher, this function contains all the information available from the data on the questions at issue, and the statistical analysis must therefore be based on the mathematical and graphical analysis of this function.

A sketch of the likelihood function itself shows the relative plausibility of the various values of θ in the light of the data. Of particular interest is the value $\hat{\theta}$ of θ which makes $l(\theta)$ a maximum. This maximum likelihood value provides the required estimate of θ. The reliability of the estimate can also be determined from the likelihood function and is related to the rapidity with which this function falls off on either side of its maximum value. Finally, the actual level of the likelihood function at the maximum, when related to the value to be expected from the degree of closeness of repeated observations, throws light on the adequacy of the proposed form of relationship.

The problem of experimental design, moreover, is essentially that of using appropriate previous knowledge to select a set of experimental conditions (a set of temperatures x_i) that can be expected to maximize the information which the likelihood function will contain on the questions at issue.

In those frequent situations where the normal distribution can provide a useful representation of the error distribution, the value of the likelihood function is

$$l(\theta) = \text{constant} - \frac{1}{2\sigma^2} S(\theta)$$

where $S(\theta)$ is the sum-of-squares function $\sum\{y_i - f(x_i, \theta)\}^2$. Study of the likelihood function is then closely related to study of the sum-of-squares function. Important elements of this study are conveniently carried out by computing an analysis-of-variance table. Furthermore, the method of maximum likelihood then yields the important method of least squares, discovered by K. F. Gauss about 1796. The techniques of statistical design and analysis for these situations have been vigorously explored and are in widespread use in scientific work. *See* LEAST SQUARES, METHOD OF; STATISTICS; STATISTICS, MATHEMATICAL.

BIBLIOGRAPHY.—R. Carnap, *Logical Foundations of Probability* (1950); O. L. Davies (ed.), *Design and Analysis of Industrial Experiments* (1956); R. A. Fisher, *Statistical Methods for Research Workers* (1925); R. B. Lindsay, *Introduction to Physical Statistics* (1941); J. B. Scarborough, *Numerical Mathematical Analysis*, 4th ed. (1958).

(G. E. P. B.)

ERSCH, JOHANN SAMUEL (1766–1828), German bibliographer, is notable for his accuracy before rules of bibliography were generally established. Ersch was born at Grossglogau, Silesia, June 23, 1766; became professor of geography and statistics and librarian at the universities of Jena (1802) and Halle (1803); and died at Halle, Jan. 16, 1828. In addition to bibliographical work on French and German publications, Ersch founded with J. G. Gruber the monumental, uncompleted *Allgemeine Encyklopädie der Wissenschaften und Künste*, 167 vol. (1818–89).

ERSE (Irish), the native language of Ireland. *See* Celtic Languages.

ERSKINE, HENRY (1746–1817), orator, wit and poet, lord advocate of Scotland, the second son of the 10th earl of Buchan, and brother of the lord chancellor Thomas Erskine, 1st Baron Erskine of Restormel, was born in Edinburgh on Nov. 1, 1746. He was educated at the universities of St. Andrews, Glasgow and Edinburgh, and was admitted to the faculty of advocates in 1768. He was lord advocate in 1783 under Fox's and North's ministry, and again under Grenville in 1806. He retired in 1811 and died at Almondel, Linlithgowshire, on Oct. 8, 1817. Erskine's reputation survives as the finest and most eloquent orator of his day at the Scottish bar; added to a charming forensic style was a most captivating wit, which, as Lord Jeffrey said, was "all argument, and each of his delightful illustrations a material step in his reasoning." Erskine was also the author of some poems, of which the best known is "The Emigrant" (1773).

ERSKINE, JOHN, of Carnock (1695–1768), Scottish jurist and authority on Scots law, son of Lieut. Col. John Erskine, was in 1737 appointed professor of Scots law in the University of Edinburgh. In 1754 he published his *Principles of the Law of Scotland*. He retired from his chair in 1765, and the rest of his life was spent in the preparation of his great work, the *Institutes of the Law of Scotland*, which he did not live to publish. He died at Cardross, Perthshire, on March 1, 1768. Erskine's *Institutes* (1773; many later editions) has always been esteemed of the highest authority on the law of Scotland. The *Principles*, being more concise and direct, retained its place as the textbook on Scots law and was frequently re-edited.

ERSKINE, JOHN, 5th Laird of Dun (1509–1589), Scottish reformer who stood high in the favour of Mary, queen of Scots, was born in 1509 and educated at King's college, Aberdeen. At the age of 21 Erskine was the cause—probably by accident—of a priest's death, and was forced to go abroad, where he came under the influence of the new learning. It was through his agency that Greek was first taught in Scotland at a Scottish school at Montrose. Erskine was a close friend of the reformer George Wishart (from whose fate Erskine was saved by his social standing and military reputation) and of John Knox, whose advice openly to discountenace the Mass was given in the lodgings of the laird of Dun. Erskine frequently acted as intermediary both between the queen regent and the reformers, and among the reformers themselves. In 1560 he was appointed—though a layman—superintendent of the Reformed Church of Scotland for Angus and Mearns; and in 1572 he gave his assent to the titular episcopacy proposed at the Leith convention. He was four times elected moderator of the general assembly (first in 1564), and he was among those who in 1578 drew up the *Second Book of Discipline*. From 1579 he was a member of the king's council. He died in 1589.

See V. Jacob, *The Lairds of Dun* (1931). (H. Wa.)

ERSKINE, THOMAS ERSKINE, 1st Baron (1750–1823), British lawyer, a distinguished advocate whose most important and lasting contributions were in the defense of personal liberties, was born Jan. 10, 1750, in Edinburgh, the youngest son of Henry David Erskine, 10th earl of Buchan. Upon completion of his preliminary education, Erskine desired to enter a learned profession, but because of the straitened financial circumstances of his family was compelled to seek a career in the navy. He became a midshipman in 1764 but left the service in 1768 and purchased a commission in a regiment of the 1st Royals. He contributed to his self-education by extensive reading, particularly in English literature. His unsigned pamphlet, *Observations on the Prevailing Abuses in the British Army* (1772), gained a wide audience. Erskine found opportunities for advancement in the army no more favourable than in the navy. Encouraged by the friendly interest of Lord Mansfield, he decided to enter the law. He was admitted to Lincoln's inn in 1775, and in 1778 received an honorary M.A. degree from Trinity college, Cambridge, and was called to the bar.

Within a few months his fame and fortune were assured by his performance in the case of Baillie, who had published charges of corruption in the administration of Greenwich hospital. Those accused instituted a proceeding to show cause why an information for criminal libel should not issue. Erskine was retained as junior counsel by the defendant, and in his first appearance at the bar vindicated his client with remarkable eloquence and courage. When asked to explain his bold and successful argument, Erskine is said to have replied: "I thought I heard my little children plucking at my robe and crying out to me, 'Now, Father, is the time to get us bread.'" Erskine very quickly rose to the leading position in the English bar. He successfully assisted the defense in the court-martial of Admiral Augustus Keppel in 1779. His notable defense of Lord George Gordon on the charge of instigating the Gordon riots of 1780 substantially destroyed the doctrine of constructive treason in English law. Erskine appeared in most of the important cases involving the disruption of commercial relations accompanying the outbreak of hostilities with France. He became the pre-eminent English advocate in cases of criminal conversation.

In 1784 he unsuccessfully defended Dr. W. D. Shipley, dean of St. Asaph, against a charge of criminal libel; but his contention that it is for the jury, not the judge, to determine whether a publication is libelous, was vindicated by the passage of Fox's Libel act of 1792. In 1789 he won an acquittal for John Stockdale, a bookseller, who was charged with criminal libel for selling a pamphlet criticizing the trial of Warren Hastings. Erskine's speech on that occasion is one of the monuments in the literature of English freedom. His unsuccessful defense of Thomas Paine in the case arising out of publication of the latter's pamphlet, *The Rights of Man*, cost Erskine his position as attorney general to the prince of Wales. His masterful representation of such clients as Thomas Walker, Thomas Hardy, Horne Tooke and John Thelwall on charges of treason and related offenses placed a powerful check on the repressive measures of the Pitt ministry taken in response to the insecurity and hysteria engendered in England by the French Revolution and its aftermath. In 1800, by establishing the defendant's insanity, he successfully defended James Hadfield, who had attempted to assassinate George III. His speech at the trial is an important contribution to the law of criminal responsibility.

Erskine, who was an intimate of the Whig leaders Charles James Fox and R. B. Sheridan, sat in the house of commons from 1783 to 1784 and from 1790 until he became a peer in 1806. Erskine's parliamentary career was not distinguished, being almost wholly devoid of the forensic triumphs that marked his practice at the bar. In 1806 he became lord chancellor during the so-called Ministry of All the Talents. Erskine was not well qualified for the position either by training or experience, and his short tenure of 14 months in office was not sufficient to permit an outstanding contribution. The latter years of his life were marked by private sorrows and misfortunes, during which period he largely withdrew himself from public affairs. Near the close of his life he again commanded widespread attention by his prominent role in defense of Queen Caroline during her trial before the house of lords. He died at Almondell, Linlithgowshire, on Nov. 17, 1823, of pneumonia, caught on a voyage to Scotland.

Erskine's great achievements were as a jury lawyer. His speeches are characterized by vigour, cogency and lucidity, and often by great literary merit. They consistently reveal a courageous and passionate dedication to the causes he represented. It was the judgment of Lord Campbell that, as an advocate, Erskine was "without an equal in ancient or in modern times."

Erskine's published works include his pamphlet, *The Unconditioned Freeness of the Gospel in Three Essays* (1778); a student declamation published in 1794 in a pamphlet entitled *Sketch of Erskine, With Anecdotes; A View of the Causes and Consequences of the Present War with France* (1797); *Armata, a Fragmen*

(1817), written in the style of *Gulliver's Travels;* a burlesque of Gray's "Bard" entitled "The Barber" in *Poetical Works* (1823). Numerous collections of Erskine's speeches have been published, the most useful of which include that edited by James Ridgway in 1847 and that edited by James L. High in 1874.

BIBLIOGRAPHY.—Relevant materials on Erskine's life and career include Lord Campbell's *Lives of the Lord Chancellors,* vol. vi (1845); Sir John Moore's *Diary* (1904); A. Fergusson's *Henry Erskine* (1882); H. Duméril's *Lord Erskine: Étude sur le barreau anglais à la fin du XVIII^e siècle* (1883); Lord Brougham's *Memoir,* prefixed to Erskine's *Speeches* (1847); Sir Samuel Romilly's *Memoirs* (1840); the *Papers* of John Wilson Croker (1884); Lord Holland's *Memoirs* (1852). A more recent biography in a popular vein is L. P. Stryker, *For the Defense, Thomas Erskine* (1947). (F. A. A.)

ERYSIPELAS (ST. ANTHONY'S FIRE), an acute bacterial infection of the skin, usually of the face, caused by hemolytic streptococci. The bacteria enter through an abrasion or a wound. The involved region becomes red and swollen and may show small vesicles. The inflammation usually extends radially from the original site; bacteria can readily be isolated from the peripheral region but are seldom present in the central portion. The incubation period is usually two to three days and the usual course varies from four to ten days. Secondary complications such as blood poisoning (septicemia) and pneumonia usually account for the fatalities. Recovery from the disease does not confer immunity to later attacks. Treatment with immune serum may reduce the severity of the disease but does not prevent recurrences. The discovery and use of the sulfa drugs and the antibiotics made treatment more effective. A particularly fatal type of erysipelas occurs occasionally in newborn infants. Another type, puerperal erysipelas, may be confined to the vulva or may extend to the vagina and uterus to give a characteristic puerperal infection. Surgical erysipelas extends from an operative wound. *See* SKIN, DISEASES OF. (F. L. A.; X.)

ERYTHRAE, an Ionic city on the Mimas (now Karaburun) peninsula in the Izmir *il* (province) of Turkey. Its foundation was ascribed to Cretans and later to Ionians. The original site has not been certainly fixed; from the 4th century B.C. it was on the west coast of the peninsula at a place now called Ildir, where traces of the wall circuit, theatre and citadel are visible. Erythrae contributed only eight ships to the Ionic fleet at the battle of Lade (*c.* 494 B.C.). Defying Athens about 453 B.C., it was compelled to receive a garrison. It revolted with Chios in 412, but welcomed Conon in 394. Freed from Persian rule in 334, Erythrae supported Antigonus I Monophthalmos. It was plundered by Galatians under Leonnorios about 275 B.C. A free city in the Roman province of Asia, it was noted for its pest-free vines which produced a soft, bouquetless wine, its hardy goats, timber and millstones. It was also famous for its prophetic Sibyls (Herophile and Athenais). (Jo. M. C.)

ERYTHRINA, a genus of trees of the family Fabaceae growing in Mexico and Central and South America; the seeds contain a number of alkaloids with a curarelike action. One of these, erythroidine, was investigated as a substitute for curare (*q.v.*).

ERYTHRITE is a mineral composed of a hydrated cobalt arsenate. Its colour is diagnostic and varies from crimson to peach blossom. The name is from the Greek for red. Known as cobalt bloom, the mineral is used chiefly as a guide in searching for cobalt-nickel-silver ores. Beautiful crystals occur at Schneeberg, Saxony, and Cobalt, Ont., and in Morocco. Erythrite also has been found in Arizona, California, Idaho, Nevada and New Mexico. The mineral has the formula $Co_3(AsO_4)_2.8H_2O$ and forms a series with annabergite, the nickel equivalent, and vivianite, the phosphate equivalent. Some iron or nickel or both replace part of the cobalt. It occurs most commonly as an earthy mass or crust. Crystals are rare and usually have a needlelike or bladed habit. The crystals form frequently as radiated groups. Erythrite crystallizes in the monoclinic system. It has a perfect cleavage, can be cut with a knife, is soft (hardness is 1.5 to 2.5) and has a specific gravity of about 3; it is formed chiefly from the weathering of cobalt arsenides, and occurs as a pseudomorph of smaltite ($CoAs_2$). *See* COBALT. (G. T. FT.)

ERYTHROXYLACEAE, the coca family, a small group of dicotyledonous shrubs or small trees that consists of three genera: *Erythroxylon,* with about 200 species in the tropics of both hemispheres; *Nectaropetalum,* with 4 species in the American tropics; and *Aneulophus,* with 1 species in western Africa. Economically, the most important member of the family is coca (*q.v.*), a South American shrub, the leaves of which yield the drug cocaine (*q.v.*).

ERZBERGER, MATTHIAS (1875–1921), German parliamentarian and statesman, leader of the left wing of the Centre party and a signatory of the Armistice of World War I, was born at Buttenhausen in Württemberg on Sept. 20, 1875, the son of a craftsman. After two years as a school teacher Erzberger, in 1896, joined the Catholic Centre paper, *Deutsches Volksblatt,* in Stuttgart. A gifted journalist and brilliant speaker, he worked his way up in the Württemberg Centre party. As chairman of the South German Union commission he opened the first Christian trades union congress (Mainz, 1899), an occasion on which the Christian unions were founded.

Erzberger became *Reichstag* member for Biberach in 1903 and settled in Berlin in 1904. From 1903 to 1914 he wrote the annual reports *Die Zentrumspolitik im Reichstag.* Soon an expert on budget, colonial and military questions, he published many important letters to the press, gradually establishing himself as the leader of the party's left wing. In 1906 he began a sensational attack on Bernhard von Bülow's government over the conditions in Germany's African territories. This caused the resignation of Prince Ernst von Hohenlohe-Langenburg, the acting director of the colonial department, and, in Dec. 1906, the dissolution of the *Reichstag.* In the so-called Hottentot elections which followed, the Centre party lost its key position vis-à-vis the Bülow bloc (Conservatives and Liberals); Erzberger, whom the official leadership of his party blamed for this, had temporarily to restrain his political activities.

During World War I Erzberger organized the information bureau of the imperial naval establishment and the "central office for foreign services," mainly responsible for propaganda in neutral countries. Erzberger won growing influence and tried to mobilize international Catholicism behind the Central Powers. He took on numerous diplomatic missions. In vain he attempted, with Bülow, now German ambassador in Rome, to prevent Italy's entering the war in 1915. For his far-flung and independent policies and propaganda he was much criticized. At first in favour of extensive annexations by Germany, by 1917 he strongly advocated a negotiated peace, a view set out in his *Der Verständigungsfriede* (1917). He opposed the government's Polish policies and more particularly the unlimited U-boat war, which he recognized as ineffective. Erzberger was decisively involved in the *Reichstag* resolution of July 19, 1917, proposing a negotiated peace with no territorial gains, and also in the events which caused the resignation of the chancellor, Theobald von Bethmann Hollweg, whom he wanted replaced by Bülow. As a leader of the new *Reichstag* majority (Centre, Social Democrats, Progressive People's party), he aimed at democratic constitutional reform. While approving the treaty of Brest-Litovsk (1918), he demanded that the eastern peoples should receive a true opportunity for self-determination. Even during the war, in his book *Der Völkerbund* (1918), he advocated a league of nations.

In Oct. 1918 Erzberger became state secretary without portfolio under Prince Maximilian of Baden. Under pressure from the chancellor and the army leadership, he undertook to lead the German deputation to the truce commission of Compiègne. There, on Nov. 11, he signed the Armistice and later directed its implementation. In Feb. 1919 he became minister without portfolio under Philipp Scheidemann in Germany's first republican government and vigorously pressed for the acceptance of the Versailles treaty. From June 1919 to March 1920 he was vice-chancellor and finance minister, in Gustav Bauer's ministry. He established a central tax office, strengthened Germany's financial sovereignty and introduced higher taxes.

These reforms, which were opposed both by the former propertied classes and by the federalists, signified a strengthening of the *Reich* and a step toward a unified state; so did the national-

ization of the railways, put through under Erzberger.

In a lengthy lawsuit against Karl Helfferich, the former secretary of the treasury who, in a pamphlet *Fort mit Erzberger* (1919), accused him of dishonesty and of merging financial interests with political activity, Erzberger suffered unrestrained slander from the extreme right. The campaign was aimed against him as the minister who had signed the Armistice and counted as the protagonist of the republican-democratic system. Clumsiness and inadequate support from his own party made Erzberger an easy target. Helfferich came out the moral victor, though fined a paltry sum "for persistent slander." When the verdict was published, in March 1920, Erzberger resigned his ministry. He was set aside by his party and did not exploit his renewed mandate in the elections on June 1920. While on vacation in the Black forest, he was shot dead on Aug. 26, 1921, by members of a nationalist organization. The murderers escaped abroad and were brought to justice only after World War II.

Erzberger was an important personality of the Centre, but often weakened its position and had strong opponents within the party. Politics was his element; a brilliant speaker with a fine sense of political developments, always willing to accept responsibility, he was an acknowledged leader. Yet as a fast and prodigious worker he displayed a certain superficiality and restlessness. Always an optimist, he underrated political opposition; trusting and open, he was notoriously indiscreet.

Among Erzberger's many political works the most important are: *Die Kolonialbilanz* (1906), *Der Toleranzantrag der Zentrumsfraktion des Reichstags* (1906), *Das Deutsche Zentrum*, 2nd ed. (1912), *Der stille Kulturkampf* (1912), *Erlebnisse im Weltkrieg* (1920) and *Christlicher Solidarismus als Weltprinzip* (1921).

BIBLIOGRAPHY.—For the best appraisal *see* K. Epstein, *Matthias Erzberger and the Dilemma of German Democracy* (1959), with bibliography, "Erzberger and the German Colonial Scandals, 1905–1910," *English Historical Review* (Oct. 1959), and "Erzberger's Position in the Zentrumsstreit before World War I," *Catholic Historical Review* (Jan. 1958). *See* also *Der Erzberger-Prozess* (1920); E. Bauer, *Erzberger*, 2nd ed. (1925); K. Bachem, *Vorgeschichte, Geschichte und Politik der deutschen Zentrumspartei* vol. ix (1932); A. Hagen, *Gestalten aus dem schwäbischen Katholizismus*, part iii (1954); J. Joos, *So sah ich sie* (1958); E. Matthias and R. Morsey (eds.), *Der Interfraktionelle Ausschuss 1917–18*, 2 vol. (1959). (R. Mo.)

ERZINCAN, the chief town of the *il* (province) of Erzincan in eastern Turkey, is situated near the western end of a rich plain. Pop. (1960) 36,420. Surrounded by orchards and gardens, it is near the right bank of the Euphrates (Firat) river, 100 mi. W.S.W. of Erzurum. One wide street traverses the town from east to west, but others are narrow and unpaved except near the government buildings and the modern mosque of Haji Izzet Pasha to the north. The town occupies an important position on the railway and main road from the Caucasus to central Anatolia. There are barracks, a military hospital and a clothing factory to the north at Kara-kulak and along the foothills. The manufacture of silk, cotton and copper utensils provides the principal industrial occupations.

Erzincan occupies the site of a nearby town which contained a temple of Anaitis (Anahita). The medieval Arsinga was important in the 4th century when St. Gregory the Illuminator lived there. The celebrated monastery of St. Gregory lies 11 mi. S.W. of the town. The district passed from the Byzantines to the Seljuks (1071) and then to the Mongols (1243). The battle in which the Mongols defeated the sultan of Rum was fought on the plain. The town was added to the Ottoman empire by Mohammed II (1473). In the latter half of the 19th century, Erzincan with its leather and textile (woolen) industries was an important centre of eastern Anatolia. However, the Armenian riots and occupation by the Russians in World War I had damaging effects and emigration much reduced the population. The town was largely destroyed by earthquakes in 1047, 1457, 1583, 1666 and 1784, but it was rebuilt each time. An earthquake in 1939 destroyed most of the town and its population.

ERZINCAN IL (area 4,607 sq.mi.; pop. [1960] 243,005) is almost surrounded by lofty mountains. The climate is severe in winter (January average temperature at Erzincan 24° F.) but warm in summer (July average temperature 74.5°). The plain is highly productive; wheat, fruit, vines and cotton are grown and cattle and sheep are raised. Water is abundant everywhere and there are sulfur springs and iron. (N. Tu.; S. Er.; E. Tu.)

ERZURUM, a town and capital of the *il* (province) of Erzurum in eastern Turkey, is situated at the eastern end of a wide valley, 450 mi. E. of Ankara and 110 mi. S.E. of Trabzon. Pop. (1960) 90,069. It is surrounded by bare and sparsely populated mountains which rise in the north to an altitude of about 10,000 ft. Erzurum is an important military station and fortress vital to the eastern defenses of Turkey. The town falls within a protected military area and can be visited by foreigners only with an official permit. No traces remain of the stone city wall known to have existed in medieval and later times but the town is still surrounded by an *enceinte* which consists of an earthern rampart with ditches and four principal gates, constructed between the wars of 1855 and of 1877. Erzurum has suffered frequent earthquakes and that of 1859 destroyed 4,500 houses and nine minarets. Nevertheless some important historical buildings have survived and are grouped around the remains of a raised citadel. The Cifte Minare is a fine Seljuk *medrese* built in the 12th century with a circular *turbe* behind it. Adjoining is the Ulu Cami (the Great mosque) and the Yakutiye *medrese*. A modern boulevard slopes down from the citadel to the Erzincan gate, passing government buildings and a public garden surrounding a 16th-century mosque. Erzurum has an airfield (9 mi. from the town), a cement factory and several hospitals. It is a main junction on the Transit route from Iran, through Dogubeyazit, which continues northwestward over high mountain ranges to the Black sea at Trabzon. The town is also linked to Ankara and Istanbul by rail. Regular motor-coach service connects it with Tabriz and Teheran. Local craftsmen still excel in metalwork and saddlery.

Erzurum has been identified with the Byzantine Theodosiopolis of the 5th century which ranked as an important fortress. Later its possession was disputed by the Sasanids and from A.D. 655 by the Arabs who occupied it for nearly a century. The emperor Constantine V razed the walls and transported many of the Christian inhabitants to Thrace. The town played an important part during the Arab wars of Leo and Porphyrogenitus (886–959). It appeared as a Byzantine dependency under the name of Karin in 978 and later as a part of the Bagratid Armenian kingdom. It fell to the Seljuks in 1071 and enjoyed prosperity in the early 13th century under the sultans of Rum who knew it as Arz-ar-Rum ("frontier of Greece"), from which its present name is derived. In 1515 it passed to the Ottoman empire.

Erzurum has been the scene of two major battles against the Russian armies. In 1877 Ahmet Muhtar Pasha successfully defended it against a Caucasian force under General Heimann whose intention was to cut off the Turkish line of retreat from Kars. The second battle, which was a brilliant feat of arms, ended in the capture of the fortress by the Russians in Feb. 1916. Erzurum remained in Russian possession until 1917. (*See* TURKEY: *History.*)

ERZURUM IL is divided into 13 *ilçeler* or districts. Pop. (1960) 568,864. The winter climate is severe with temperatures reaching about −24° F., accompanied by blizzards. Summer temperatures rarely exceed 70° F. The *il* is traversed by the Kara Su, a headstream of the Euphrates. The plain produces wheat, barley, millet and vegetables and latterly also sugar beets.

See H. F. B. Lynch, *Armenia* (1901); W. E. D. Allen and P. Muratov, *Caucasian Battlefields* (1953). (S. H. Ll.)

ESARHADDON, king of Assyria from 681 to 669 B.C., who had already been proclaimed successor to Sennacherib (*q.v.*) although a younger son, came to the throne after the murder of his father; it has been suggested, against the weight of evidence, that Esarhaddon himself was the regicide. His first task was to quell a rebellion in favour of his brothers. Marching from the west he met the rebels in Hanigalbat (western Assyria), and after part of their forces had deserted to him defeated the remainder and restored order.

The rebellion had unsettled the Sealands (southern Babylonia), making military action necessary against the paramount chieftain of the Chaldean tribe of Bit-yakin, who was claiming independence

and had attacked the Assyrian governor of Ur. The insurgent fled to Elam, but there a ruler friendly to Assyria had succeeded, and he was executed. His brother, who submitted to Assyria, was then appointed ruler over the Sealands, proving a loyal vassal.

The cities of northern Babylonia, which had suffered severely under Sennacherib, were shown particular favour under Esarhaddon. Land was restored to displaced citizens who could make good their claim, and in 678 B.C. military action was taken against the Chaldean tribe of Bit-dakkuri, which had encroached on the lands of Borsippa and Babylon.

In the west Sidon, in association with maritime allies in the Gulf of Antioch, in 678 B.C. withdrew its allegiance, possibly at Egyptian instigation. The city was taken by assault and with the surrounding area made into an Assyrian province. Tyre remained loyal to Assyria at this time, receiving an accession of territory in return for increased tribute, but subsequently, inspired by Egypt, with which it was bound by commercial interests, revolted. The city, well supplied by sea, withstood siege for many years.

Throughout Esarhaddon's reign relations with Elam were, except for an Elamite raid on Babylonia in 675 B.C., good. This resulted in part from Esarhaddon's avowed policy toward Gambulu, the buffer state east of the lower Tigris, whose ruler he assisted in strengthening his defenses against Elam. In Media the chieftains accepted Assyrian suzerainty, three of them invoking help against domestic rebellions. An Assyrian expedition penetrated to the salt deserts of Iran, exacting nominal submission and heavy booty.

Farther north the pressure of Cimmerians (*q.v.*) and Scythians was being increasingly felt. The kingdom of Urartu was rapidly falling under the control of the Scythians; Esarhaddon claimed to have defeated them in the region south of Lake Urmia, but any success there was ephemeral. The indications are that Esarhaddon later made a marriage alliance with the Scythians to strengthen his position in the north. The pressure of the Cimmerians was heaviest in the northwestern provinces of Cilicia and Tabal, and despite initial successes in the area Esarhaddon finally lost control of much of this territory.

The most striking political development of Esarhaddon's reign was his campaign against Egypt. After initial operations of limited success the major operation was conducted in 671 B.C. when Memphis fell, Tarku (Tirhakah), king of Egypt, fled to the south and the whole land surrendered. Twenty-two native rulers were appointed under Assyrian suzerainty, and Esarhaddon withdrew. In 669 B.C. Esarhaddon was marching to deal with fresh unrest in Egypt when he died.

Esarhaddon had made detailed arrangements for the succession, two sons, Ashurbanipal (*q.v.*) and Shamash-shum-ukin, having been nominated to the thrones respectively of Assyria and Babylonia and vassal princes bound by oath to support them. This, and his settlement of Babylonia, are marks of Esarhaddon's statesmanship. *See* also BABYLONIA AND ASSYRIA.

See *The Cambridge Ancient History*, vol. iii, chapter iii (1925); R. Borger, *Die Inschriften Asarhaddons, Königs von Assyrien* (1956); H. W. F. Saggs, *The Greatness That Was Babylon*, pp. 123–127 (1962).
(H. W. F. S.)

ESAU, identified in Hebrew tradition with Edom, the eponymous ancestor of the Edomites, was the son of Isaac and Rebekah and elder twin brother of Jacob (Gen. xxv, 24–26). This narrative probably represents an earlier superiority of Edom over Israel, further attested by the view that Edom was a settled state before the Hebrew conquest of Canaan. The change in comparative status is reflected in two narratives of Genesis; in the one Esau sells his birthright to Jacob for a mess of red pottage (Gen. xxv, 21–34), whence the name Edom ("red"). According to the other, Jacob, in alliance with Rebekah, succeeded in cheating Esau out of the paternal blessing (Gen. xxvii). That Israel regarded Edom as living on a lower social order appears from Gen. xxv, 27, where Esau is a representative of the primitive "food collector" or wandering hunter, and Jacob is a type of the "civilized" pastoral nomad. The greater fertility of Israelite territory is illustrated by the blessing of Jacob and the corresponding curse of Esau (Gen. xxvii, 26–29, 39–40).

ESBJERG, a seaport in Denmark in Ribe *amt*, lies on the west coast of Jutland opposite Fano island. Pop. (1960) 55,171. Esbjerg grew rapidly after the harbour was constructed in 1874, and was granted a charter in 1899. It has municipal schools, teachers', technical and commercial colleges and a folk high school. There is a small airport. Esbjerg is Denmark's largest fishing port with 6 mi. of quay and more than 500 fishing vessels. Beside fish and fish products large quantities of meat and dairy products are exported, chiefly to Great Britain; fertilizers, foodstuffs, coal, oil and industrial goods are imported. (E. BT.)

ESCALATOR. An escalator is a passenger conveyer in the form of a moving stairway in which the steps move as a unit upward or downward under power, at an angle. The escalator consists of a series of steps (treads and frames) attached to chains, all contained within a steel truss. The chains pass over sprocket-wheel assemblies at the top and bottom. An electric motor is geared to the top sprocket wheels to produce a continuous, even movement of the steps, which are supported by rollers traveling on inclined tracks. Side enclosures or balustrades are provided to form a railing and to conceal the mechanism. Along the top of each balustrade is a handrail which moves in synchronism with the steps.

The purpose of the escalator is to carry large numbers of persons continuously from one elevation to another. The width between balustrades usually accommodates one to three files of passengers.

The early type of escalator was in the form of a conveyer or an endless belt on which cleats were provided for riding stability. Jesse W. Reno was issued a patent in 1891 for an inclined elevator. Later the cleats were modified to give a larger angle of incline and thereby form a more comfortable step and this design became known as the Reno type escalator.

Credit is given to Charles D. Seeberger for bringing out, about the same time, a similar type of conveyer with horizontal steps. However, these necessitated stepping on and off to one side, thus introducing a hazard.

The Otis Elevator company eventually acquired both of these inventions. Installations were first made in 1900 at the Paris exposition, in the New York city subway and in Bloomingdale's store, New York city. Both designs had weaknesses; the Reno cleat type necessitated a relatively small angle of incline of the escalator and consequently more space, and the Seeberger type presented the difficulty of stepping on and off.

In 1921 the Otis Elevator company evolved and sold the first horizontal-cleat type step which incorporated the best step features of both the Reno and Seeberger escalators. By means of a combplate at the landings, the step cleats mesh with the prongs of the combplate, thus decreasing considerably the hazard in stepping off. The angle of incline was adopted as 30° which appeared to give a reasonable depth of step tread and height of step riser.

Later, in 1932–33, the Westinghouse Electric Elevator company introduced an electric stairway in which were incorporated narrower cleats in the horizontal steps and metal balustrading. Vari-

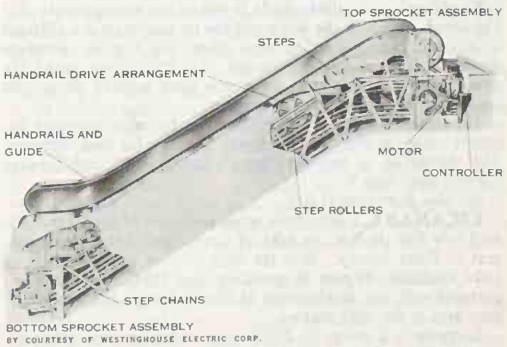
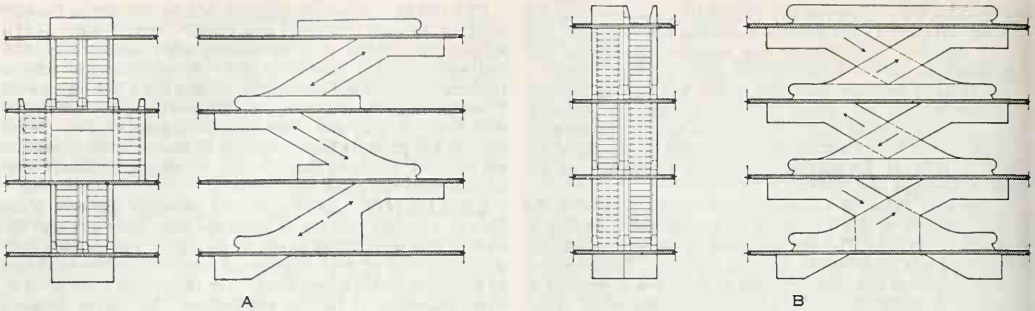


FIG. 1.—CUTAWAY VIEW OF AN ESCALATOR



BY COURTESY OF WESTINGHOUSE ELECTRIC CORP.

FIG. 2.—DIAGRAMS OF (A) PARALLEL, AND (B) CRISSCROSS, TYPES OF ESCALATORS

ous safety devices were introduced to stop the steps in case of overspeed, breakage of chains, etc. Quietness, smooth riding and improved appearance did much to expand the use of this type of conveyer so that it found general use.

The advantages of the escalator as compared with the elevator are its larger handling capacity for short distances; its immediate availability for transportation; its comparatively small space requirements; and its low operating cost.

The escalator is commonly built in two standard sizes: 32 in. and 48 in. as measured between the balustrades. These have respective ratings of 5,000 and 8,000 persons per hour when operating at a speed of 90 ft. per minute. The two commonly used operating speeds in the United States are 90 and 120 ft. per minute. The higher speed has the advantage of increased capacity and faster transportation.

Certain design characteristics are incorporated to provide a maximum degree of safety. The materials used are noncombustible; the newels carrying the moving handrails extend beyond the steps at both top and bottom to lead passengers off and on; emergency stop buttons are located in one of the balustrade newels at top and bottom; a key is required to start the escalator; the first step at top and bottom is at floor level for easy stepping on and off; a service brake is provided on the motor shaft to stop the escalator whenever power is shut off by one of several safety devices; a floating lower sprocket assembly shuts off power in case an obstacle jams the steps; and the finger section of the combplate is smooth to allow a person's feet to slide off in case he does not step off.

Applications of escalators are found principally in multifloor stores, subways, office buildings, transportation terminals, banks, hotels and schools.

In stores escalators may be the chief means of transportation, since traffic demand is reasonably continuous and heavy. Their convenience is often a factor in many applications since they eliminate the need for waiting.

Escalators are installed usually in one of two arrangements: (1) Crisscross—in which the down and the up escalators are adjacent usually and in planes 60° to each other; this requires minimum space and is the usual arrangement. (2) Parallel—in which the down and the up escalators are in the same inclined plane; this arrangement occupies more floor space.

The escalator is primarily a mass mover for relatively short distances while the elevator is a mover of many for long distances. The distinction is between "mass" and "many" and between "short" and "long."

See also ELEVATOR (Lift). (H. C. HK.)

ESCANABA, a city of the upper peninsula of Michigan, U.S., on Little Bay de Noc, an inlet of Green bay (Lake Michigan); seat of Delta county. It is the only iron-ore shipping point on Lake Michigan; its port, in operation since 1864, increased in importance with the development of steelmaking in the South Chicago area in the 20th century.

Escanaba is a centre of distribution, recreation and forest administration and has marine terminals for storage and shipment of coal and petroleum products. Important products are veneers, pulpwood, paper, lumber, dimension stock, wood novelties, garments, truck cranes, road stabilizers, log loaders, castings, welding machines, rustic fencing, stone, gravel, concrete products and fish.

Sawmills were operating in the area as early as 1830 and development of the settlement accelerated in 1863, when the first ore dock was constructed. Escanaba was chartered as a city in 1883. The city's name is an Indian word meaning "flat rock." In 1922 it adopted a council-manager form of government.

For comparative population figures *see* table in MICHIGAN: *Population.* (J. Wo.)

ESCHATOLOGY (the "doctrine of last things") is a theological term, coined in modern times but derived from New Testament phrases themselves harking back to the Old Testament. In one form or another, all religions and philosophies have some teaching about "last things," about the ultimate destiny of man, society and the cosmos, as they do about "first things," or origins. This fact has been widely noted, but it should also be noted that this widespread appearance of a common theme in human spirituality hides a profound difference in how the theme ("last things") is understood and developed.

By and large, eschatologies have appeared in two radically divergent forms, distinguished by their attitude toward time and history. On the one hand, ultimate destiny may be seen as consummatory of nature; on the other hand, it may be seen as a fulfillment, rectification or transfiguration of history. Nature again may be understood as natural vitality, engrossed in cyclical rhythms, or as eternal spirit, in a realm beyond the temporal. Together, thus, there are three forms of eschatology, reflecting three basic modes of understanding reality—naturalistic, eternalistic and historistic. Naturalistic eschatologies are characteristic of so-called primitive religions and of mass religiosity in many higher cultures. Eternalistic eschatologies have been developed most typically in Greek philosophy and Hindu spirituality and seem to arise in cultures that have reached a certain degree of spiritual sophistication. Historistic eschatologies, in the full sense of the term, however, have appeared only in the sphere of Hebraic religion (Judaism, Christianity, Islam) and in a truncated form perhaps in Zoroastrianism as well. Through the spread of a Christianized western culture over many parts of the world, the historistic eschatology of the Hebrews, often in distorted and drastically secularized forms (Condorcet, Hegel, Comte, Spencer, Marx), has permeated cultures based on very different ontological and eschatological concepts.

Eschatological teachings, whether explicit or merely implicit in the ways and outlook of the culture, are everywhere intimately connected with concepts or beliefs about the divine, about man, about ethics, about history and about salvation.

NONHISTORICAL ESCHATOLOGIES

Naturalistic.—In primitive religion, man understands himself as part of nature, which itself is seen as the locus of all ultimate sanctities. Nature is engrossed in cyclical rhythms, and the pat-

tern of eternal recurrence becomes the context of naturalistic eschatology. Wrongness is experienced as alienation from nature, and through cult, magic and myth the attempt is made to re-establish the organic unity with nature which is blessedness. The vision of last things is a vision of such organic unity, beyond the threat of time and season, and is generally expressed in concrete and graphic imagery characteristic of the culture. On occasion, the achievement of the end-state is regarded as the work of a heroic, divine or semidivine being (sometimes a bird or a beast) and is expected only after a period of conflicts, cataclysms and disasters (but here many investigators suspect the influence of Christian or Muslim ideas). Whatever may be the sources of these conceptions, they fit without much difficulty into the underlying nature pattern of the eschatology.

Eternalistic.—Eternalistic eschatologies appear in various forms in Greek and oriental spirituality and have deeply influenced the religions of the west, based though these latter are on the very different Hebraic outlook. Most eternalistic eschatologies find their background in the conception of time as the endless cycle of eternal recurrence. It is from this "vain repetition" that they seek to escape: the "last thing" hoped for is the deliverance of the individual from the unreal realm of the empirical, temporal and historical to the timeless realm of spirit. Prominent in all schools of Chinese thought, for example, is the belief that the universe is in a constant state of flux, either of oscillation or of cyclic movement in a closed circuit. Indian thought emphasizes the belief in kalpas, or world periods of four phases, through which successive worlds appear, flourish, disintegrate and die. Similar doctrines are familiar from Greek and Hellenistic literature.

In Indian thought, not only does each particular human life pass through a phased cycle but there is a repetitive succession of lives in the cycle of rebirths (metempsychosis, transmigration of souls). Original Buddhism denies the substantial identity of the soul but strongly emphasizes the succession of lives governed by the law of karma, according to which the quality of each life is the determining factor for the next. Temporal existence in the world, the round of birth, death and rebirth (samsara), is conceived of as misery and evil; deliverance is sought through self-extrication from the unreal world of sense, time and history and through self-elevation to the timeless realm above and beyond. Means of deliverance are found in cultic devices, asceticism, mystical illumination and "true knowledge." The end-state which is striven for is sometimes, especially in early Buddhism, indistinguishable from nonbeing (nirvana). In later Buddhism, as in most Indian cults, the essential character of this escape-from-time eschatology is overlaid by heavy accretions of popular religion, but it is never quite extinguished.

Greek religion presents the most varied aspects. The Homeric Hades seems to represent a sadly attenuated continuation of earthly life, but there is also early reference to the Elysian fields, where the afterlife is much happier. In the civic religion of the polis, the individual's destiny was supposed to be absorbed in and expressed through the destiny of the community; but it may be questioned whether this ever proved fully satisfactory. By and large, Greek spirituality seems to be pervaded throughout by a deep melancholy about the transitoriness of life and the emptiness of all things given over to birth and death. Against the "sharp tooth of time" the Greek sought refuge in some form of perpetuation—the perpetuation of his fame through epic and story; the perpetuation of youth, beauty and perfection through art; the perpetuation of life through identification with the immortal god in the many mystery cults; the perpetuation of the mind through the philosophic discipline which dissolves the temporal into the eternal; and finally, perpetuation of being through the survival of the soul, by nature immortal.

The generic likeness of Greek eschatologies to those of the far east, especially India, has often been noted; but it need not be attributed, as it usually is, simply to direct influence, since some such outlook seems to be natural to men at a certain state of cultural development. The world of sense, of time and history, is a world lost in unreality and emptiness (the "vanity" of Ecclesiastes); deliverance is deliverance from this world to a superworld above and beyond, a world of eternal being. The end-state is thus a state in which the soul returns to its own proper abode of timeless eternity.

Zoroastrianism constitutes a perplexing problem. Its view of reality differs sharply from that of the Hindus or the Greeks, since, despite all the cyclical periodization it sometimes indulges in, its view of time and history is not one of eternal recurrence but rather one of linear movement toward a consummation. The world is the theatre of a grandiose cosmic struggle between the forces of good and the forces of evil; everyone is called to take his place behind Ahura Mazda on behalf of Good and Truth against Ahriman and his hosts of Evil and the Lie. All hope is directed toward an end-time in which the Saoshyant, the "bringer of good," himself of the seed of Zoroaster, will appear, disperse the powers of evil, resurrect the dead, renew the world and usher in an age in which all men "will rejoice for ever and ever" and "become of one voice giving praise with a loud voice to the Wise Lord and the Bounteous Immortals." Though in developed form Zoroastrian eschatology may be a later (Achaemenid) development, there can be little doubt that even in its earliest forms it reveals features resembling Jewish and Christian teaching far more closely than anything coming out of India or Greece.

HISTORICAL ESCHATOLOGIES

Old Testament Eschatology.—Historistic eschatology is native to the Old Testament and thus enters into the structure of faith of those religions that stem from it (Judaism, Christianity, Islam). In one sense, of course, considering the literary history of the Old Testament writings, with their many and various times, places and circumstances of origin and redaction, it may be impossible to speak of an Old Testament eschatology, since there are many motifs, diversely developed. Yet there is unity, and this unity consists in the conviction that the present (historical) period of wrongness is a falling away from an original (protological) rightness (the rightness of God's creation) and is destined for a return to a final (eschatological) rightness in which God's purpose will at last be fulfilled. Old Testament eschatology is unique in that it is bound to the course of history as its goal and end, which history it sees as *Heilsgeschichte* (redemptive history—the history of redemption, the history that redeems). It is therefore closely connected with God's management of history (providence), especially through the individual or community he has chosen as his instrument (election, covenant).

Old Testament eschatology is basically corporate; it is the elect community (Israel) as People of God that is destined for restoration at the end-time, and by way of the elect community all of mankind—indeed, the entire cosmos. Particularistic and universalistic motifs are thus combined, though in different forms and to different degrees in different contexts. The end-time is a time of both judgment and salvation; the judgment is a judgment on historical action, and the salvation is a fulfillment or transfiguration of historical life. What is looked for is not an escape from the world, time and history, but a new world, a new time, a new history ("new heavens and a new earth"; Isa. lxv, 17, lxvi, 22), no longer against God, as in the present age, but at one with God, under his life-giving rule (the Kingdom [*i.e.*, kingship] of God). The mediator of this salvation is, in many cases, the anointed (Messiah) of God, a king of the seed of David, who will bring God's blessing and peace to Israel and the nations. There will be an ingathering of the dispersed of the House of Israel (Isa. xl, 11, xliii, 5, lvi, 8; Jer. xxxi, 8, 10; Ezek. xxxiv, 13), and the peace of Paradise will return to the world (Jer. xxxi, 12–14, 23–25; Zech. viii, 11–13; Hos. xiv, 5–7). In Second Isaiah (*see* ISAIAH) these various strands seem to be gathered up; God's judgment to be overcome only by atonement and forgiveness, God's salvation for Israel, and through Israel for the nations, in a new life under God.

Old Testament eschatology becomes individualized at a relatively late stage, although it must not be forgotten that the destiny of the individual had always somehow been included in the salvation of the community. When, however, with the Exile, the com-

munity could no longer satisfactorily serve as a total context of being and meaning for the individual, the problem of individual eschatology emerged, just as did that of individual responsibility. Perhaps the earliest undisputed passage speaking of individual resurrection is the celebrated verse, Dan. xii, 2 ("And many of those who sleep in the dust of the earth shall awake, some to everlasting life, and some to shame and everlasting contempt"), which dates from the 2nd century B.C. It should be noted that it is not the immortality of the soul that is spoken of but the resurrection of the dead, who have remained dead ("sleep in the dust") until this end-time (when "your people shall be delivered" [xii, 1]). This remains the hope and expectation normative in Judaism and Christianity.

There has been much inquiry as to the sources of Old Testament eschatology. Its kinship with the Zoroastrian scheme is obvious, but the relation is not clear. Other influences have been sought. It would be well to distinguish two questions often confused: (1) the sources of the multitude of images, symbols and motifs in which the Old Testament eschatological hope is expressed, and (2) the intrinsic character of this hope. The first question cannot be answered with any assurance, but it can hardly be doubted that the sources are many and diverse. The second must be answered with a reassertion of the uniqueness of the Hebraic eschatological vision, always bearing in mind, however, the extraordinary similarity of the Zoroastrian scheme of last things.

Judaism.—Old Testament eschatology, both particular and universal, both corporate and individual, entered into the formation of Judaism, but in different ways in the different tendencies that went to make up the Judaism of the ancient world. Most groups, aside from the Sadducees, accepted the doctrine of the resurrection; indeed, in Pharisaic teaching it became a primary article of faith, as it has continued to be in any form of Judaism that makes any pretensions to orthodoxy. The three-phase pattern—first age, present age, age-to-come—also became part of the tradition of faith. A significant difference, however, emerged, perhaps only gradually, between Pharisaic orthodoxy and the apocalyptic visionaries. The former tended to stress the God-obedient life in the present age, while always hoping for and expecting the imminent advent of the Messiah, the resurrection of the dead, the final judgment and the age-to-come. The apocalypticists, on the other hand, virtually wrote off the present age as irredeemably evil, saw the new age suddenly breaking in as something totally different (replacing, not redeeming, man's life in history) and claimed to have a revealed scheme of periodized world-ages giving a sure knowledge of the imminent end. Apocalyptic visions are found in some of the Old Testament writings, especially Daniel, but the main literature comes in the intertestamental period (*see* APOCALYPTIC LITERATURE). Other groups in the ancient Jewish world, such as the Essenes and the Qumran community, combined these two forms of Jewish eschatology in various ways. Also about this time strong Hellenistic influences began to be felt, bringing with them alien soul doctrines and gnostic teachings, which were soon syncretized with the earlier biblical ideas.

Early Christianity.—In the New Testament there are presumably three elements: (1) the authentic tradition of Jesus' words and deeds; (2) the elaboration of this tradition in the corporate tradition, written or oral, of the New Testament community; and (3) the theological reflections of particular writers. These elements are never easy, and often impossible, to distinguish.

Jesus' own eschatological vision was undoubtedly that of the Judaism of his time, with perhaps a heightened sense that the age-to-come was dawning and the inauguration of God's sovereign rule was imminent. As to Jesus' own understanding of his messianic role (the so-called problem of Jesus' messianic consciousness; *see* JESUS CHRIST: *The Mission of Jesus*), there is wide disagreement and little probability of an unequivocal answer. The Jewish eschatological hope involved the belief that the Messiah himself would be known and acknowledged as such only when he had revealed his identity through his work of salvation. If this is to be applied to Jesus, it would mean that during his life he was *Messias designatus* (as termed by S. Mowinckel), to be enthroned only with the accomplishment of his mission.

Whatever may be the case with Jesus himself, it is clear that the earliest New Testament community saw in him, upon his crucifixion, resurrection and ascension, the Messiah of God, the eschatological saviour, he in whom, through whom and by whom the new age, the veritable age-to-come, had at last been opened. He had come in the "fullness of time" (Gal. iv, 4), had gone and was to come again very soon. The Christians, living already in the new age (hence freedom from the Law), yet also in the old, were to wait for his return in power and glory, when the forces of evil, already conquered, would be wiped out, the dead resurrected, all creation judged and renewed and God's plan finally consummated. In this faith, in this tension "between the times," the New Testament community lived, hoped and bore its witness.

With the delay of the return (parousia), many problems arose. These problems just appear in Mark, are carried forward in Matthew and are "solved" in Luke, in the sense that the interim period begins to take on the aspect of an age of the church, following after the age of Israel and the age of Jesus. What effect the delay of the parousia had on early Christian belief is much disputed, but it can hardly be doubted that it facilitated the reinterpretation of the originally thoroughly Jewish messianic eschatology in an existentialist direction, on the one side, and a gnostic-spiritualistic direction on the other. Paul and the Gospel of John already contain strong elements of the former: the eschatological moment is now, the moment of confrontation with the demand of faith, and judgment and resurrection come now, upon the decision of faith. Yet the tension between the already and the not yet, the orientation toward the future consummation, is never wholly lost; indeed, in Paul at least, it is often strongly emphasized (*see, e.g.,* I Cor. xv, 24–28; I Thess. iv, 13–18). More serious was the Hellenization of the early Christian eschatology, in which time concepts are gradually replaced by those of space (the age-to-come by another world), the resurrection of the body is played down, and the historical structure of the eschatology is spiritualized into a kind of allegory of the odyssey of the soul. Beginnings of this Hellenization are to be detected in the New Testament, as indeed in the Judaism of the time.

Scholars have interpreted New Testament eschatology in surprisingly different ways. Albert Schweitzer has found it to be consistently futuristic; so also, later, does Martin Werner. Rudolf Bultmann sees it as an eschatology of existential decision in the present, opening the future. C. H. Dodd, on the other side, has spoken of a realized eschatology, an eschatology in which the eschatological act has already been realized, with little or nothing left for the future. (Dodd subsequently somewhat modified his position, declaring the term "realized" to be subject to misunderstanding and speaking of an "inaugurated eschatology.") Oscar Cullmann attempts a synthesis of the futuristic, realized and existentialistic interpretations in an over-all conception of time and history which he believes does justice to the faith and outlook of the early Christian community.

The multidimensionality of New Testament eschatology is graphically illustrated in the varying yet related uses of the term "last days" (last day, last time). This term is used: (1) for time contemporary with the events reported (Acts ii, 17; also John xi, 24–25); (2) for the time of preparation of the final consummation (II Tim. iii. 1); and (3) for the final consummation itself, the parousia (John vi, 40; I Pet. i, 5). All these eschatological "times" are taken up in the one great eschatological event.

Later Developments.—Later developments in Christian eschatological thinking move along three lines. With some (Irenaeus, Tertullian), the old eschatology remains a vital element in their vision of God's redemptive work in history. Others retain the eschatological doctrine, but as a kind of foreign body in an otherwise unhistorical, uneschatological theology (Latin and Greek apologists of the 3rd century, Lactantius). Alexandrian Platonists, for their part, seem determined to spiritualize away the eschatological tradition so foreign to their presuppositions of thought, and this brings them perilously close to gnostic ideas (Clement, Origen; though there has arisen a tendency to see in Origen a much more biblical thinker than was formerly the case).

Augustine, the great teacher of western Christendom, produced a new synthesis; the neo-Platonist philosopher gave forth a theology of history which takes man's historical enterprise and its eschatological consummation with the utmost seriousness. There is little spiritualization in Augustine, but, on the other end, there are the beginnings of the ecclesiastical interpretation. The millennial (thousand-year) rule of the saints, which in the New Testament (Rev. xx) figures as a preliminary period of earthly rule by Christ and his chosen ones, is now presented as being carried through by the church ("The church now on earth is both the kingdom of Christ and the kingdom of heaven" [*The City of God,* xx, 9]). This interpretation became standard doctrine in the middle ages, meeting with sporadic resistance from heretical millenarian movements. (*See* also MILLENNIUM.)

The Middle Ages.—In medieval Christendom, the New Testament eschatology was given its place in a dogmatic system of which the philosophical foundations were at first Platonistic and, later in the west, Aristotelian. Traditional conceptions about the parousia, resurrection and the like were combined with Greek notions about the soul and its immortality, giving rise to a duplicated eschatology (post-mortem, post-resurrection), as in medieval Judaism and Islam, which were subject to the same influences. Medieval Christianity was largely institutional, juridical and sacramental, with little place for the eschatological passion. This passion, however, was not dead; it lived in certain heretical movements, but especially in the strange speculations of Joachim of Floris, the 12th-century Calabrian monk who foretold the imminent advent of the Age of the Spirit, with its "unwritten gospel," following upon the two previous ages, the Age of the Father (Old Testament) and the Age of the Son (New Testament) (*see* JOACHIM).

The Reformation.—The early Reformation saw itself, in eschatological terms, as engaged in a struggle with Anti-Christ for the purification of the church in the face of the imminently approaching end-time. Especially in Luther, but also to some degree in Zwingli and Calvin, there is a strong apocalyptic tension, nurtured by the crisis of the time and by the return to Scripture. Reformation eschatology was much influenced by the dominant role Paul's writings played in the new theology.

The period of the Reformation also witnessed a great resurgence of millenarianism on the part of the off-centre sects, often, though inaccurately, described as Anabaptists. The disastrous collapse of these adventures turned the sectaries from an activist, militant millenarianism to one that was quietist and pacifist, from an attitude of forcing the time to one of waiting upon the Lord. The development of pietism in the Lutheran world, and partly also in the Reformed world, led to an analogous position.

The early Reformation soon became consolidated, its religion conventionalized and its theological outline cast in scholastic dogma. Eschatological teaching took on very much the aspect it had had through the middle ages. Yet something of the older feeling remained in the left-wing Calvinist groups, especially among those finding themselves in situations making for political militancy.

Religious Liberalism.—In the religious liberalism that emerged, especially among Protestants and Jews, toward the end of the 18th and through the 19th century, eschatology could find no place. It was regarded as part of the crude, primitive, outworn trappings of traditional religion which could no longer be accepted in an age of enlightenment. In most cases, eschatological ideas were abandoned altogether, and a simple post-mortem immortality of the soul was held forth as man's end. Other theologians reinterpreted the Kingdom of God expectation in ethical, quasi-mystical or social terms. Such was the cultural atmosphere that the eschatological element of the New Testament came to be considered merely peripheral, not at all integral to its essence, which was held to be spiritual and moral. It came as a flash of lightning, therefore, when Albert Schweitzer, in a series of writings just before World War I, proved conclusively that the eschatological hope, often cast in a radically futuristic form, was the very heart of the New Testament faith.

20th-Century Theology.—Theology by the middle of the century had come to be thoroughly eschatological in temper and orientation; the eschatological theme dominates theological thinking. This is, in varying degrees, true of Protestantism, Catholicism and Judaism; even Eastern Orthodoxy, insofar as it speaks through Nikolai Berdyaev, has been deeply affected. The problem for 20th-century theology seems to be double: (1) first, to recapture the true eschatological vision of biblical revelation; and (2) to interpret this vision in terms relevant to conceptions, perplexities and problems of the time. Various theologians deal with these problems in various ways, and there is no consensus, not even among theologians of the same religious confession. Yet on one thing there is agreement: eschatology is not just a part of the biblical faith; it is that faith itself in its ultimate dimension as it looks forward to the "mighty act" of God in which all creation is restored to its true end, which is also its true source of being. *See* also references under "Eschatology" in the Index volume.

BIBLIOGRAPHY.—*Nonbiblical Religions:* Mircea Eliade, *The Myth of Eternal Return* (1955); G. van der Leeuw, *Religion in Essence and Manifestation,* 2nd ed. (1947); R. C. Zaehner, *At Sundry Times* (1958). *Old Testament:* R. H. Charles, *A Critical History of the Doctrine of a Future Life in Israel, in Judaism, and in Christianity,* 2nd ed. (1913); S. Mowinckel, *He That Cometh: the Messianic Concept in the Old Testament and Later Judaism* (1956); Helmer Ringgren, *The Messiah in the Old Testament* (1956). *Judaism:* George Foot Moore, *Judaism in the First Centuries of the Christian Era,* 2 vol. (1927, 1930); Paul Volz, *Die Eschatologie der jüdischen Gemeinde im neutestamentlichen Zeitalter* (1934). *New Testament and Early Christianity:* Oscar Cullmann, *Le Retour du Christ* (1948), in *The Early Church* (1956); William D. Davies, *The Background of the New Testament and Its Eschatology* (1956); William Manson, "Eschatology in the New Testament," *Scottish Journal of Theology Occasional Papers,* no. 2 (1953); J. A. T. Robinson, *Jesus and His Coming* (1958). *Patristic, Medieval and Reformation Periods:* Norman Cohn, *The Pursuit of the Millennium* (1957); G. W. H. Lampe, "Early Patristic Eschatology," *Scottish Journal of Theology Occasional Papers,* no. 2 (1953); Heinrich Quistorp, *Calvin's Doctrine of Last Things* (1955); T. F. Torrance, *Kingdom and Church: a Study in the Theology of the Reformation* (1956). *Modern Discussion:* Karl Barth, *The Epistle to the Romans,* 6th ed. (1933); Rudolf Bultmann, *History and Eschatology* (1957); Emil Brunner, *Eternal Hope* (1954); Romano Guardini, *Die letzten Dinge,* 4th ed. (1956); Reinhold Niebuhr, *Faith and History* (1949); W. A. Whitehouse, "The Modern Discussion of Eschatology," *Scottish Journal of Theology Occasional Papers,* no. 2 (1953).

(W. HE.)

ESCHEAT, a term signifying the right of the sovereign to property of a person who dies intestate and without lawful heirs, or of a person such as an alien barred from ownership of property, or to unclaimed intangible property such as bank accounts.

Historically, escheat meant the right of a lord to the inheritable estates in land of a tenant who died without lawful heirs. The comparable consequence for personal property was accomplished by the doctrine of *bona vacantia,* unclaimed goods that became property of the crown. Besides escheat for defect of heirs, there was also escheat *propter delictum tenentis,* or by the corruption of the blood of the tenant through attainder (*q.v.*) consequent on conviction and sentence for the offense of treason or felony. The real estate therefore escheated to the lord from whom he held, subject to the superior right of the crown to the forfeiture of the lands—forever in the case of treason, and for a year and a day in the case of felony. All this was abolished by the Felony act, 1870, which provided for the appointment of an administrator to the property of the convict.

Because of the gradual elimination of mesne or intermediate lordships after the statute of *Quia Emptores* in England in 1285, almost the only lord entitled to escheat in modern times has been the crown or, in parts of the United States, its successor, the state. Technical escheat was abolished in England in 1925, and the crown is entitled to real property as well as to personal property under the *bona vacantia* doctrine. (*See* REAL PROPERTY AND CONVEYANCING, LAWS OF; CONFISCATION AND EXPROPRIATION.)

In those states of the U.S. which treat the state as the successor to the crown, escheat of land may have its historic meaning, but in most of the U.S., where land is said to be allodial, or held absolutely, the state seems to be treated as the ultimate heir. This theory also prevails in many of the civil-law countries. In some

states of the U.S., escheated property does not pass to the state for general use but is held in trust for a public benefit, such as schools, poor relief or prisons.

In modern times many factors in the laws of inheritance affect escheat. The inclusion as heirs of half-bloods, illegitimate children, adopted children, the surviving spouse and, in some states, even other relatives by marriage have tended to make failure of heirs an extremely rare occurrence. On the other hand, the trend in some jurisdictions to reduce the category of relatives who may inherit to those of a more immediate class has tended to increase the possibility of escheat. The most important factor contributing to the insignificance of escheat is the existence of extremely wide powers of testation. (*See* INHERITANCE; INTESTATE SUCCESSION.)

The most recent legislative developments in escheat have resulted from the discovery that there are large credits in the hands of institutional debtors, such as banks and insurance companies, representing unclaimed bank balances, dividends, interest and principal on bonds, life insurance proceeds, pension funds and unpaid wages due an employee who leaves his job before the stipulated pay day. Legislation exists in many states of the U.S. which provides that after the lapse of a specified period of time, usually 10 or more years, funds in inactive accounts become the property of the state and not of the bank or institution indebted to the original claimant. These statutes usually require the holder of the funds to advertise for the owner before the funds are turned over to the state and also provide for a long period during which a claimant may come forward and claim the escheated property or its value from the state. (A. DM.)

ESCHENBURG, JOHANN JOACHIM (1743–1820), German critic, literary historian and translator of Shakespeare, was born at Hamburg, Dec. 7, 1743. He became director of the Collegium Carolinum in Brunswick and was a friend of Lessing. He published translations of works by the principal English writers on aesthetics, including Dr. Charles Burney, Joseph Priestley and Richard Hurd, and also the first complete translation (in prose) of Shakespeare's plays (*William Shakespeare's Schauspiele*, 13 vol., 1775–82). This is virtually a revised edition of the incomplete translation published (1762–66) by C. M. Wieland. Eschenburg died at Brunswick, Feb. 29, 1820.

See H. Schroder, *Eschenburg und Shakespeare* (1911).

ESCH-SUR-ALZETTE, the second largest town of Luxembourg, lies on the upper Alzette, 10 mi. S.W. of the capital and close to the French frontier. Pop. (1960) 27,954. Before 1870 Esch was just a village, but it grew rapidly as a result of the large-scale exploitation of the local phosphoric iron ore, the use of which for steel production was made possible by an invention of the British metallurgist S. G. Thomas (*q.v.*). Esch is the chief centre of the Luxembourg iron and steel industry. Three large integrated steelworks are located on the outskirts of the town, and others lie within a few miles of it. Basic slag fertilizer is a by-product. (K. C. E.)

ESCHWEILER, a town of Germany, which after partition of the nation following World War II was included in the *Land* of North Rhine-Westphalia of the Federal Republic of Germany. It lies 15 km. (9 mi.) E.N.E. of Aachen by road. Pop. (1961) 39,590. Eschweiler is on the edge of the Eifel mountains and it is the chief town of the Aachen coal basin. Iron, steel and zinc goods are made and other industries include tanning, distilling and brewing. The town was in the duchy of Jülich until the French Revolutionary Wars and was annexed by Prussia in 1815.

ESCOBAR Y MENDOZA, ANTONIO (1581–1669), Spanish Jesuit, famous as a preacher and moral theologian, was born at Valladolid. He became the butt of Pascal's ridicule in several of his *Provincial Letters*, in which the Jansenist opponent of the Jesuits pilloried Escobar and his confreres as teachers of lax moral principles and as proponents of the principle that "the end justifies the means." These attacks of Pascal laid Escobar open to further witty criticism by such other French writers of the 17th century as Molière, Boileau and La Fontaine, none of whom was particularly friendly to the Jesuits. Unprejudiced scholars have found Pascal's attacks on Escobar quite unfounded, if the Jesuit's words are taken in context. In moral matters Escobar was a strong supporter of the position known as Probabilism (*q.v.*). He died at Valladolid on July 4, 1669. His collected writings comprise 32 volumes on biblical, sacred and moral subjects. (J. F. BN.)

ESCOFFIER, GEORGES AUGUSTE (1846–1935), French culinary artist who became "the king of chefs and the chef of kings," was born at Villeneuve-Loubet, Alpes-Maritimes, on Oct. 28, 1846. He began his career at the age of 12, and when he retired from the Carlton hotel, London, at the age of 74, he counted 62 years of active service, a span considered a record in his profession. The name of Escoffier became of world-wide repute when in 1890 he was given the direction of the kitchens of the newly opened Savoy hotel, London, and he created the *pêche Melba* in honour of the famous singer when she was staying there in 1893. In 1899 he moved to the Carlton hotel, where he was to build up a fabulous reputation for *haute cuisine* during the next 23 years: on one occasion Emperor William II is reported to have said to Escoffier, "I am the emperor of Germany but you are the emperor of chefs." In recognition of his services to the prestige of French cooking abroad, he was awarded the Légion d'Honneur in 1920 and made officer of the order in 1928. He died in Monte Carlo on Feb. 12, 1935.

Besides the renown of his name, said to be greater even than that of Marie-Antoine Carême, Escoffier left several books, the best known of which are *Le Carnet d'Epicure* (1911), *Le Guide Culinaire* with Philéas Gilbert and Emile Fetu (1903), *Le Livre de Menus* (1912) and *Ma Cuisine* (1934).

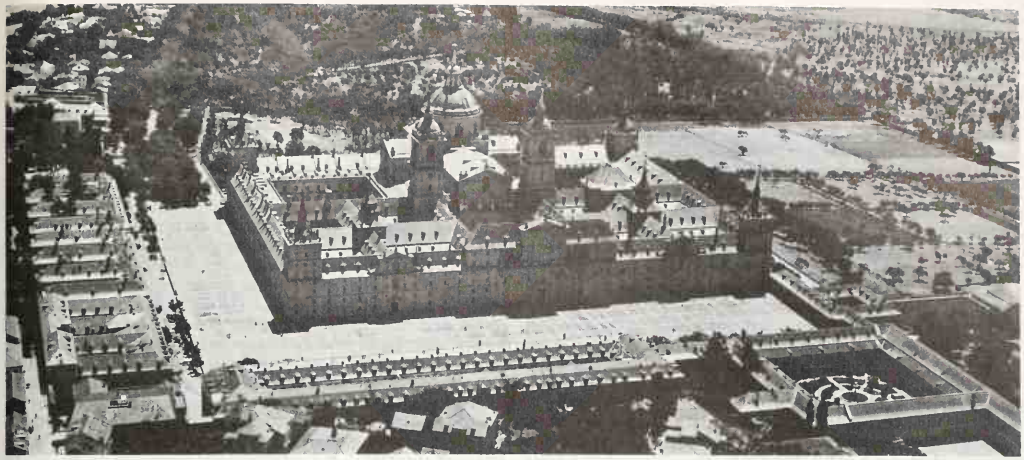
See E. Herbodeau and P. Thalamas, *Georges Auguste Escoffier* (1955). (P. W. HE.)

ESCORIAL (ESCURIAL), a village in the Guadarrama mountains, 32 mi. N.W. of Madrid, Spain, the site of the monastery of San Lorenzo el Real del Escorial, originally Hieronymite but since 1885 occupied by Augustinians. Philip II decided to construct a monastery there, and to make it the royal pantheon where all Spanish sovereigns beginning with the emperor Charles V are buried, except Alfonso XIII (died 1941). One of the largest (about 675 ft. by 525 ft.) and handsomest religious establishments in the world, it was begun in 1563 by Juan Bautista de Toledo, a Spanish architect, who had worked earlier in Italy.

Toledo is responsible for the general plan consisting of a great rectangle of three parts, the centre being occupied by the church. To the south of it lie five cloisters in which are included the royal palace and offices, while the living quarters of the monks occupy an equivalent northern wing. After his death in 1567, Toledo was succeeded by Juan de Herrera, who made extensive revisions in the designs, prepared new plans for the church (1573) and brought the whole building to completion in 1584. Although the church (about 320 ft. by 200 ft.) of Greek-cross plan surmounted by a monumental dome (about 60 ft. in diameter, 320 ft. high at its centre) was clearly inspired by Michelangelo's Cathedral of St. Peter's in Rome, Herrera's building was far more sober. The massive walls of the interior, relieved only by Doric pilasters with no concession to decorative richness, produced a monument austere beyond anything the Italian Renaissance ever envisaged. On the exterior the gigantic scale of the monastery and the severe gray granite walls are overpowering. There Juan de Herrera, the most celebrated of all Spanish architects, established his fame and the style which was to prevail in Spain for a half century.

Philip II engaged numerous Italian and Spanish artists to carry out work for the Escorial. The effigies of Charles V and Philip II with their families which kneel beside the high altar, as well as the statues of saints in the altar itself, were cast in bronze in Milan by Pompeo Leoni. Luca Cambiaso painted frescoes in the vaults of the church, while Pellegrino Tibaldi was the chief decorator of the Cloister of the Evangelists and of the library.

The most interesting parts of the royal palace are the apartments of Philip II, with 16th-century furnishings and personal possessions of the monarch. The section which was redecorated for Charles IV in the late 18th century is notable for its tapestries by Goya and by the French and Flemish schools. A charming example of the decorative arts of the same period is the small palace,



BY COURTESY OF THE SPANISH GOVERNMENT TOURIST OFFICE

MONASTERY OF SAN LORENZO EL REAL DEL ESCORIAL; BEGUN 1563 BY JUAN BAUTISTA DE TOLEDO, COMPLETED 1584 BY JUAN DE HERRERA

the Casita del Principe, located in the gardens below the monastery.

The art collection of the monastery, which ranks among the finest in the world, is exhibited primarily in the sacristy and the chapter halls. It includes Van der Weyden's "Crucifixion," El Greco's "Martyrdom of St. Maurice" and the "Dream of Philip II," important masterpieces by Bosch, Patinir, Titian, Tintoretto, Veronese, Velázquez, Ribera, Claudio Coello and numerous other masters.

The Escorial library, originally founded by Philip II, houses one of the rarest collections in existence among its 4,000 manuscripts, many of them illuminated, and its 40,000 printed books. Some of the choicest items are the Arabic manuscripts, the Mozarabic codices, the early Gothic Cántigas de Santa María and the Lapidarium (both belonging to the reign of Alfonso the Wise) and the Golden Gospels of Otto II (1033–39).

See also RENAISSANCE ARCHITECTURE: *Spain: Herreran.*

See J. A. Gaya Nuño, *El Escorial* (1947). (H. E. W.)

ESCROW, in law, a term generally applied to a deed or other written instrument entrusted to a third party with instructions that it be delivered to the grantee upon the occurrence of a designated event. Until the occurrence of the event, the instrument is said to have been delivered "in escrow"; *i.e.,* as a mere scroll or writing. In England (where the term may also refer to a document retained by the grantor himself) and in most states of the U.S., the escrow becomes a deed, note or bond, as the case may be, when the specified event occurs; its effectiveness does not depend upon actual physical delivery to the grantee by the third party or depositary (known as the escrow agent or escrowee).

If the grantor delivers the instrument to the person intended to benefit, it becomes immediately operative despite any conditions, unless the conditions are contained in the instrument itself. A grantee who secures possession of the instrument from the depositary before the conditions are fulfilled does not take title, and the grantor is generally protected even against innocent third parties who may have relied on the validity of the instrument.

The term escrow originally applied only to instruments for the conveyance of land, but is now used for all written instruments deposited in the manner described above, including mortgages, bonds, promissory notes, life insurance policies, patents, stock certificates, and even money deposited to be held until the performance of a condition.

ESCUDERO, VICENTE (1892?–), considered by many critics the greatest flamenco dancer of his era, was born of gypsy parentage in Valladolid, Spain; the date is unknown and may have been considerably earlier than 1892. He was known in the cafes of Spain during his youth and began his international career in 1920 when he first danced at the Olympia theatre in Paris. His success was immediate and in 1922 he returned to Paris with a small company for concert appearances at the Salle Gaveau. Until 1932 Escudero was occupied with extensive tours in Europe. In that year he made his first tour of the United States. In 1934 he danced the role of Carmelo in La Argentina's production of the ballet "El Amor Brujo" at the Teatro Espanol in Madrid.

During World War II and for some time thereafter Escudero disappeared from public view. But in 1954 he formed a company, returned to Paris for a brilliant season and then toured the United States for 10 months. In 1959 he began a farewell tour of the United States with Carmita Garcia and a company of gypsies, but because of financial difficulties the company was disbanded in Dallas, Tex.

Escudero is credited with being the first to dance to the rhythms of the "Seguiriyas Gitanas," a gypsy *seguidilla,* which he introduced in 1940. (R. M. Hs.)

ESCUDO, the monetary unit of Portugal after 1911, containing 100 centavos. Originally, it was a Spanish gold coin, struck in 1523 by Charles I to replace the gold ducat. The name is derived from the Latin *scutum* ("shield"), and was given to the coin because it bore Charles's arms. This coin remained in circulation in Spain until the middle of the 19th century.

During the 17th and 18th centuries, when Portugal declined in power, its supply of precious metals dwindled and resulted in the introduction of Spanish coins. From 1722 on, the escudo was coined in Portugal and also became the national money of account. This situation continued until 1834 when the escudo was replaced by the rei. The rei soon declined in value and after the revolution of 1910 the escudo was reintroduced. One new escudo was equal in value to 1,000 reis, or to one milreis. The par value of the new escudo was approximately 108 U.S. cents. It declined in value, however, soon after it was introduced and in 1920 was worth only 10.49 U.S. cents. After the collapse of the gold standard in 1931, the value of the escudo was fixed in terms of the pound sterling, 110 escudos being made equal to one pound. In terms of U.S. dollars, the escudo was worth 4.42 cents. When the dollar was depreciated in 1934 the value of the escudo rose correspondingly. Following the devaluation of the pound sterling in 1939 the value of the escudo fell to approximately 4 U.S. cents. This rate of exchange was maintained throughout World War II until the pound was again devalued in Sept. 1949. Although the pound was devalued 30%, the escudo was devalued only 13%. Its value thus fell to 3.48 U.S. cents, or 81.60 escudos to the pound. Throughout the 1950s and early 1960s its value remained comparatively stable, with buying and selling rates within the limits of

28.48 and 29.02 escudos per dollar. One thousand escudos equal one conto. The modern escudo has been coined in various denominations in silver, bronze and copper. (Jo. M. L.)

ESCUINTLA, a department in southern Guatemala and its capital city of the same name. Area of the department, 1,693 sq.mi. Pop. (1957 est.) 154,065. It is located on the Pacific lowland and on the southern slope of the highlands, and includes the country's chief Pacific port, San José.

The department's main products are sugar cane, bananas, coffee, maize, citronella and tropical fruits. Large areas are devoted to the pasture of beef cattle. The department is well served by highway and railroads.

The city of Escuintla is 28 mi. S.W. of Guatemala City, at an elevation of 1,109 ft. Pop. (1957 est.) 11,511. (P. E. J.)

ESDRAS, FIRST BOOK OF (the GREEK EZRA), an apocryphal book of the Old Testament, in the English Bible first in position among the Apocrypha. It includes parts of the canonical II Chronicles, Ezra and Nehemiah, with an interpolated story.

The title of this "book" is ambiguous and confusing. In Latin Bibles since Jerome, in some Greek manuscripts and in the Ethiopic version it is called III Esdras. In the Lucianic recension of the Greek Bible it is called II Esdras. It is I Esdras in the most important Greek manuscripts, in the Old Latin version prior to Jerome and in the English Apocrypha since the Bishop's Bible (1560). It is popularly called Greek Ezra or Apocryphal Ezra. (For a table delineating the complex history of the Ezra literature, *see* ESDRAS, SECOND BOOK OF.)

History.—Although listed among the Apocrypha, I Esdras differs greatly from other such writings since it is the oldest-known Greek translation of a portion of the books of Chronicles–Ezra–Nehemiah. It is but a fragment of the canonical work, beginning abruptly with II Chron. xxxv, 1 and concluding mid-sentence with the first few words of Neh. viii, 13. The Lucianic recension (a revision of the text of the Greek Bible by Lucian of Antioch in the late 3rd or early 4th century A.D.) and the Latin version arbitrarily complete the final sentence of the "book." C. C. Torrey explained its fragmentary character by the assumption that someone, to preserve an important recension of a text in danger of being lost when authorities favoured the present Hebrew text, excerpted probably a single ten-leaved quire portion of a Greek codex containing the interesting text.

The translation was made perhaps in Alexandria, Egypt, at some time about the beginning of the 2nd century B.C. from a Hebrew-Aramaic text much like that of the current Hebrew Bible except for its order of the narrative (*see* below). The anonymous translator skilfully rendered the Semitic original into good, idiomatic Greek, neither a free rendering nor a paraphrase. The language is more elegant than that of II Esdras, its competitor in the Greek Bible (*see* ESDRAS, SECOND BOOK OF). In at least 20 instances I Esdras has preserved a better text than that of the present Hebrew Bible.

Josephus (*c.* A.D. 100), the first witness to I Esdras, uses it as his source for both his account of Josiah and for his description of the return from exile. Although he alters Artaxerxes of I Esdras to Cambyses in the interest of a better chronology, Josephus borrows much of the exact wording of I Esdras, includes the unique episode of the guardsmen (*see* below), and passes directly from Ezra x to Neh. viii, as I Esdras does. During the first five centuries A.D. both Greek and Latin Church Fathers treated the work as canonical and often quote from it. Origen preserved it in his Hexapla (*q.v.*), and Paul of Tella (A.D. 616–617) introduced it from there into his Syriac Hexapla. It is found in codices A and B of the Greek Bible but it is missing from the incomplete Codex Sinaiticus.

Jerome, the first to reject I Esdras, did not concern himself with the Old Latin versions of I Esdras, and the book is missing from the older manuscripts of the Vulgate (*e.g.*, Codex Amiatinus) and from the Complutensian Polyglot (1502). I Esdras was rejected as uncanonical by the Council of Trent (1546), and, as III Esdras, it is today appended to the New Testament in the Vulgate version. It is omitted also from the Syriac Bible (Peshitta), but there is a Syriac translation in the Walton Polyglot Bible (1657). In the English Bible I Esdras now holds first place in the Apocrypha.

Contents.—In two striking particulars I Esdras differs from canonical Ezra–Nehemiah: (1) It includes the story of the three guardsmen at the court of King Darius, along with some appended narrative (I Esd. iii, 1–v, 6), resulting in a transposition of the account of events in the time of Artaxerxes (I Esd. ii, 16–30 = Ezra iv, 7–24); and (2) part of the Book of Nehemiah (Neh. vii, 73–viii, 13a) is made to follow the last verse of Ezra x.

In the story of the three guardsmen (I Esd. iii, 1–v, 6), three young members of the Persian king's bodyguard, apparently at Susa, in a contest of wit and wisdom, declaimed before the king and his princes on the subject of the strongest thing in the world. The winner was promised a rich prize. After wine and the king were proposed, the winner first suggested women and then lauded Truth as mightiest of all. This bit of pagan wisdom literature, conjectured by R. H. Pfeiffer to be a Zoroastrian popular homily including an echo of a hymn in praise of Asha ("Truth"), was composed probably sometime between 400 and 350 B.C. As an Aramaic text it was adapted for Jewish use in the Book of Ezra by the interpolation of "that is Zerubbabel" (I Esd. iv, 13) to identify as Zerubbabel the winner of the contest, who is elsewhere only an anonymous young man (I Esd. iv, 58) and by rewriting the ending of the story. As his prize Zerubbabel asked permission to rebuild Jerusalem (I Esd. iv, 43). To the original story the editor appended an account of the return of Jews from Babylonia under the leadership of Zerubbabel (I Esd. iv, 48–v, 6). Torrey argued that the account of the return was originally part of the chronicler's work, slightly edited, that was later accidentally lost when the story of the guardsmen was excised from the Hebrew Bible.

In I Esdras the text of Neh. i, 1–vii, 5 is omitted entirely. After the account of mixed marriages (Ezra x, 16–44 = I Esd. ix, 15–36) the reading of the Law by Ezra (Neh. viii) is introduced abruptly in I Esd. ix, 37–55. The Semitic prototype used by the translator of I Esdras seems to have included the text of Neh. vii, 6–73 also, for the ending of that section, with some adjustment, is attached to Neh. viii, 1 as translated in I Esd. ix, 37–38 to provide the date and setting for the reading of the Law (Neh. viii = I Esd. ix, 37–55). The long list in Neh. viii is omitted in I Esdras because, as a duplicate of Ezra ii, it was already used in I Esd. v, 7–46. Josephus follows I Esdras in the abrupt transition from Ezra x to Neh. viii. After the first few words of Neh. viii, 13 the fragmentary book of I Esdras breaks off and is concluded. Josephus continues to the end of Neh. viii, with a reference to the Feast of Tabernacles. Such also must have been the case in the codex from which the I Esdras fragment was plucked. *See* also EZRA AND NEHEMIAH, BOOKS OF.

BIBLIOGRAPHY.—C. C. Torrey, *Ezra Studies* (1910), *The Apocryphal Literature*, pp. 43–54 (1945); S. A. Cook in R. H. Charles, *Apocrypha and Pseudepigrapha*, vol. i, pp. 1 ff. (1913); W. O. E. Oesterley, *An Introduction to the Books of the Apocrypha*, pp. 133–141 (1935); R. H. Pfeiffer, *History of New Testament Times*, pp. 233–257 (1949).
(RD. A. B.)

ESDRAS, SECOND BOOK OF (the so-called EZRA APOCALYPSE), an apocryphal book of the Bible, second in position among the Apocrypha in the English Bible. It is numbered IV Esdrae in the Vulgate Bible. II Esdras has no claim to be part of the Old Testament, and is printed in the Latin Vulgate as an appendix to the New Testament. Ch. i–ii and xv–xvi are in fact Christian additions, but the book as a whole has as little right to be assigned to the New Testament as to the Old. It is a product of the apocalyptic movement (*see* APOCALYPTIC LITERATURE), partly Christian (i–ii, xv–xvi), partly Jewish (iii–xiv).

The literature associated with the name Ezra has a complex history; the table seeks to minimize the difficulties. Originally, the books of Ezra and Nehemiah were integral parts of the Hebrew work *Dibre Hayyamim*, now divided into I and II Chronicles, Ezra and Nehemiah.

The "Ezra Apocalypse" has not survived in either its Semitic (probably Aramaic) original or (apart from a few verses found among the papyrus fragments at Oxyrhynchus) in its Greek translation, but only in Latin and other derived versions. Even in the

Latin, however, it is clear that ch. i–ii and xv–xvi were originally separate works. Some Latin manuscripts number Ezra and Nehemiah as I Esdrae i–xxiii, and assign the title II Esdrae to "Ezra Apocalypse" i–ii and the title V Esdrae to "Ezra Apocalypse" xv–xvi, leaving only ch. iii–xiv to be designated as IV Esdrae.

Hebrew	Greek (Septuagint)	Latin (Vulgate)	English (King James and Revised Standard versions)	Current description
Ezra . . .	II Esdras i–x	I Esdrae (or Ezras i–x)	Ezra	—
Nehemiah . .	II Esdras xi–xxiii	II Esdrae (or Ezras xi–xxiii)	Nehemiah	—
Parts of Chronicles, Ezra and Nehemiah. .	I Esdras	III Esdrae	I Esdras	Greek Ezra
—	—	IV Esdrae	II Esdras	Ezra Apocalypse

Ch. vii in most Latin manuscripts lacks some 70 verses (probably deleted because they deny the value of intercessory prayers for the dead) which were restored to the Revised and Revised Standard versions as verses [36]–[105], the former verses 36–70 receiving the additional enumeration [106]–[140].

Contents.—The book may be outlined as follows: Ch. i–ii. Ezra received a revelation announcing the rejection of Israel and the substitution of gentile Christians as the people of God.

iii–v, 20: Ezra wants to know why Zion has been destroyed by Babylon (*i.e.*, Rome). He is told that he cannot hope to understand all things, but the righteous will be vindicated in the Age to Come, which will be heralded by certain signs.

v, 21–vi, 34: Ezra is assured that the righteous who have died will share equally in the resurrection with those living at the last day.

vi, 35–ix, 25: Ezra deplores the fewness of those to be saved. He is shown the future state of the righteous and the wicked, and is told that Messiah will reign on earth 400 years. He and all mankind will then die. The resurrection and the Day of Judgment will follow after seven days.

ix, 26–x, 59: Ezra, mourning for Jerusalem which had been destroyed, is given a vision of the heavenly, incorruptible Jerusalem.

xi, 1–xii, 51: Ezra has a vision of an eagle, which is identified as the fourth beast of Dan. vii, but is reinterpreted (xii, 12) to mean Rome. It is confronted by a lion, who destroys it, and the lion is said to be Messiah.

xiii, 1–58: Ezra sees a man arise out of the sea, who is attacked by a multitude, but he destroys them and welcomes to himself another and peaceable multitude. Ezra is told that the man is Messiah, the belligerents are the gentiles and the peaceable multitude are the lost ten tribes.

xiv, 1–48: Ezra is warned of his impendent assumption into heaven, and is inspired to restore the "lost" 24 books of the Old Testament and to add 70 more—*i.e.*, the apocalyptic literature of the period.

xv–xvi: Prophecies of the destruction of "Babylon," Asia, Egypt and Syria.

Date.—The compiler used much traditional material, but in its present form the main part of the book (iii–xiv) comes from the last decade of the 1st century A.D. The first four visions (iii–x) probably existed separately and may not originally have been connected with Ezra.

Teaching.—The book is a classic example of apocalyptic, displaying an eschatological preoccupation and a love of imagery. It also illustrates the determinism of the period (iv, 36–37) and the doctrine of racial depravity (*cor malignum*) (iii, 21, 25–26; vii, [48], [118]), but not that of inherited sin. At this and several other points, the book shares with the Syriac Apocalypse of Baruch the same Jewish background as St. Paul (*see* BARUCH APOCALYPSES). In its deep dissatisfaction with the current Jewish theology it has been likened to the Book of Job. It was taken over by its Christian editor for the sake of its supposed antiquity and its eschatological teaching, without concern for other elements quite inimical to Christian views (*e.g.*, vii, 28–29). Possibly its clear teaching on rewards and punishments in the afterlife (vii, [76]–[99]) also commended it to Christian readers.

BIBLIOGRAPHY.—B. M. Metzger, *Introduction to the Apocrypha* (1957); W. O. E. Oesterley, *II Esdras* (1933), Eng. trans. and commentary; G. H. Box in R. H. Charles (ed.), *The Apocrypha and Pseudepigrapha of the Old Testament*, vol. ii (1913). (ST. B. F.)

ESENIN, SERGEI ALEKSANDROVICH (1895–1925), Russian poet who described himself as "the last poet of wooden Russia," was born Oct. 3 (new style; Sept. 21 old style), 1895, to a peasant family in Konstantinovo (since renamed Esenino after him) in Ryazan province. Trained as a teacher, Esenin soon embraced a literary career. Influenced in form by Aleksandr Blok and Andrei Bely (*qq.v.*) and in outlook by the peasant poet Nikolai Klyuev (1887–*c.* 1937), he published his first book, *Radunitsa* ("Day of the Dead"), in 1916. He welcomed the Revolution as a social and spiritual transformation which ultimately would lead to a peasant millennium in *Inoniya* ("Otherland," 1918) and *Preobrazhenie* ("Transfiguration," 1919). In 1919 he signed the literary manifesto of the Imaginists and was later regarded as the head of that group.

In the early 1920s he lived a rowdy bohemian life, contracted an ill-starred marriage with the dancer Isadora Duncan (*q.v.*) and visited the U.S. This period coincided with his disenchantment with the Revolution, which gave his poetry new tragic overtones, as in his books *Moskva kabatskaya* ("Moscow of the Taverns," 1924) and *Ispoved kkuligana* ("Confessions of a Rowdy," 1920); and the poems "Kobylyi korabli" ("Mare Ships") and "Cherny chelovek" ("The Black Man"). On Dec. 27, 1925, Esenin hanged himself in a Leningrad hotel having written his last lines in his own blood.

A prolific and somewhat uneven writer, Esenin had a true gift of song. His poignant short lyrics, full of striking imagery, often evoked the Russian countryside. Genuine and deep emotions were also hidden under the outward cynicism and blasphemy of his "tavern" poetry. Daring imagery was coupled with a pure melodic line. Toward the end of his life Esenin showed many signs of developing in the direction of greater, at times Pushkinlike, simplicity. He was very popular both during his lifetime and in the years following his tragic death. Communist critics and party leaders spoke of the dangers of "Eseninism" and of its debilitating effect on the civic dedication of the young, and Esenin was long more or less out of official favour, though selections of his poetry occasionally appeared. Several collected editions of his works became available about 1956–60. English translations of Esenin's poems are found in various magazines and anthologies.

BIBLIOGRAPHY.—F. de Graaff, *Serge Ésénine, sa vie et son oeuvre* (1933); P. Pascal, "Ésénine, poète de la campagne russe," *Oxford Slavonic Papers*, vol. ii, pp. 54–71 (1951); J. Lavrin, *From Pushkin to Mayakovsky* (1948); R. Poggioli, *The Poets of Russia: 1890–1930* (1960). (G. ST.)

ESHER, an urban district of Surrey, Eng., on the Portsmouth road, 16 mi. S.W. of Charing Cross, London. Pop. (1961) 60,586. Esher lies on rising ground by the Mole, 3 mi. from its junction with the Thames at East Molesey. Though mainly modern in appearance, Esher has Saxon origins and is mentioned in Domesday. Of the 15th-century Esher place, founded by William of Waynflete, only the gateway remains.

To the south is Claremont, rebuilt in the Palladian style by Lancelot "Capability" Brown for Lord Clive in 1769 on the site of a house of Sir John Vanbrugh. In 1816 the mansion was presented to Charlotte, the granddaughter of George III and heiress presumptive, on her marriage to Prince Leopold of Saxe-Coburg (afterward Leopold I of Belgium); she died there in 1817. King Louis Philippe died there in exile in 1850. In 1882 Queen Victoria granted the property to her youngest son, the duke of Albany. Claremont was bought in 1926 for a girls' boarding school; the woods and lake belong to the National trust.

The urban district includes Esher, East and West Molesey, Long Ditton, Thames Ditton, Cobham, Claygate, Stoke D'Abernon (whose ancient church contains brasses to the D'Abernons, 1277 and 1327) and Oxshott, as well as Sandown Park racecourse (opened in 1875). The area is mainly residential with few industries and many open spaces, such as Ditton marsh, Oxshott

heath and Littleworth, Esher and Cobham commons.

ESHNUNNA, an ancient city marked by a mound, the modern Tall Asmar, lying in the Diyala river valley about 20 mi. N.E. of Baghdad in Iraq. The mound was excavated in the course of six expeditions organized by the Oriental Institute of The University of Chicago under the direction of Henri Frankfort.

In limited areas of the mound virgin soil was reached. From these soundings it appears that the site was not occupied till late Protoliterate times (prior to 3000 B.C.), but the Diyala valley was settled in earlier periods, as the pottery remains attest, so that the earliest settlement may not have been found. The city expanded throughout the Early Dynastic period and during its later years (*c.* 2400 B.C.) contained a large administrative building called "the northern palace" by the excavators. During the 3rd dynasty of Ur the city was the seat of an *ensi* (governor) of the area. Following the collapse of Ur, Eshnunna became independent as one of several contending principalities. Eshnunna maintained this independence, sometimes with success—as, when under Bilalama (*c.* 1980 B.C.) and later under Naramsin the city controlled the Diyala valley and beyond—and sometimes without—as under Abdierah (*c.* 1890 B.C.), who lost control of the lower Diyala to the city of Kish, or under Ibalpiel II (*c.* 1780 B.C.), who had to accept the hegemony of Assyria and Eshnunna was reduced to a city-state. The city was decisively defeated by Hammurabi, king of Babylonia. It revolted after his death, but his son, Samsuiluna, reconquered the city, captured its king and subdued the land by building at least one garrison fort. During the next century the city fell into decline and may have been abandoned. An unsuccessful effort to resettle Eshnunna was made by the Kassite King Agumkakrime (*c.* 1650 B.C.).

The important "Laws of Eshnunna" are inscribed on two broken tablets found in Tall Abu Harmal, a mound on the outskirts of Baghdad, during an excavation conducted by Taha Baqir for the directorate general of antiquities of the government of Iraq. The site appears to have been a rural town and an outpost of Eshnunna during a period of strength under a king Naramsin about a generation before Hammurabi.

The two tablets are not duplicates but separate copies of some older source. Since they contain a number of errors and omissions, both tablets were probably private copies made by scribes for their own use, rather than official copies. Because the section which held the date of promulgation is damaged in one tablet and is destroyed in the other, it is impossible to date the laws exactly. In general it is thought that the laws are some two generations older than the code of Hammurabi. Both the laws and the code are written in the same style, but with many differences in content. These differences help illuminate the development of ancient law. *See* also BABYLONIA AND ASSYRIA.

BIBLIOGRAPHY.—For the historical periods *see* the chapter "Historical Data" by Thorkild Jacobsen in *The Gimilsin Temple,* published as Oriental Institute Publication XLIII of The University of Chicago (1940). For the laws *see* Albrecht Goetze, *The Laws of Eshnunna, Annual of the American Schools of Oriental Research,* vol. xxxi (1956). The results of the excavation are being published in the "Oriental Institute Publications Series." (J. W. B. H.)

ESKER, a name for long and often winding ridges of gravel and sand in regions of former glaciation, notably in the state of Maine, U.S., in Ireland and in Sweden (where they are called *åsar*). Their origin has given rise to much controversy, but they are generally considered to represent the material formerly filling the channels of streams under, within or even upon the ice. This material is left behind as the ice melts. *See* also GLACIER; PLEISTOCENE EPOCH. (R. H. RA.)

ESKIL (*c.* 1100–1182), Danish archbishop, was a nephew of Asser, the first archbishop of Lund. After being educated abroad, mainly at Hildesheim, he became a dean at Lund in 1131, bishop of Roskilde in 1134 and archbishop of Lund in 1138. He strongly asserted the northern church's independence from the archbishop of Hamburg-Bremen and persuaded the pope to rescind the bull of 1133 which had confirmed Hamburg's claims to superiority. Eskil sought radical reform of the Danish church on the Gregorian model, and in 1139 celebrated its direct affiliation to Rome, at a synod in Lund presided over by a papal legate. In 1145 the new cathedral at Lund was consecrated, and Eskil generously endowed the cathedral school; he also founded new abbeys as daughter foundations of French houses. His championship of the church's independence often meant conflict with the monarchy, as did his consistent rejection of German claims to supremacy over the Danish church and state whenever kings or pretenders allied themselves with the German crown. During the 1150s Eskil had to accept the pope's division of Norway and Sweden into separate ecclesiastical provinces though he retained his primacy over Uppsala and was designated legate. After Valdemar I became sole Danish king in 1157, king and archbishop co-operated until, with the papal schism in 1159, the king recognized the imperial antipope Victor IV and Eskil the pope, Alexander III, who was also his friend. In 1161 he went into exile. In the mid-1160s however, Valdemar turned to Alexander and became reconciled with Eskil, who at Ringsted in 1170 canonized the king's father Knud Lavard and annointed his son Knud VI as joint king. In 1177, possibly under pressure, Eskil designated as his successor Absalon, the champion of Valdemar's policies. Immediately after, he resigned and went to Clairvaux in France, where he died in Sept. 1182.

See H. Koch, *Den danske Kirkes Historie,* vol i (1950); L. Weibull, *Skånes kyrka till 1274* (1946). (A. E. CN.)

ESKILSTUNA, a town in southeast Sweden, in the *län* of Södermanland, lies on the Eskilstuna river which drains Lake Hjälmaren into Lake Mälaren, and is 73 mi. W. of Stockholm. Pop. (1960) 58,793. The centre of the town is Free City square. There are two open-air museums, an art gallery and a 12th-century church, as well as parks and amusement places. Eskilstuna is one of the chief centres of the Swedish steel and metal industry. Cutlery and precision instruments are of particular importance.

Originally called Tuna, the town was renamed Eskilstuna after Eskil, an 11th-century English missionary and martyr who was the first bishop of the province. It was mentioned as a town in the 13th century and in 1659 became a free city and smithery town. (C.-F. BJ.)

ESKIMO, a people occupying nearly all the coast line from Greenland and Labrador in the east to the Bering sea in the west, together with a short stretch of the Siberian shore in the vicinity of Bering strait. Excluding the Aleut (*q.v.*) of the Aleutian Islands, who are usually listed as a separate people, they number approximately 50,000, of whom 22,890 were registered in Greenland, 11,500 in Canada, 15,882 in Alaska and 1,100 in the U.S.S.R. Most of the Greenland and Labrador Eskimo carry a considerable amount of white blood in their veins: indeed, many Greenlanders, with their blue eyes and fair hair, would readily pass as Scandinavians were it not for their Eskimo tongue. Elsewhere pure-blood Eskimo still predominate, but intermixture with whites was steadily increasing during the 20th century and doubtless will continue to increase until the Eskimo vanish completely as a separate people, for in frontier regions like the arctic there is little prejudice against racial admixture.

Despite the enormous extent of their territory, it seems probable that the Eskimo have never numbered more than at mid-20th century, and they may have quite well numbered less. Europeans, it is true, introduced diseases, notably smallpox, in the earliest times, later measles and, after World War I, influenza, which wiped out whole groups; but the losses were repaired fairly quickly, it would seem, since in nearly every region the population had adjusted itself rather closely to the available food supply until Europeans disturbed the economic balance. Judging from the number and size of their abandoned settlements, there were fewer Eskimo in prehistoric Canada than in either Alaska or Greenland, though its arctic and subarctic coast lines are much longer. In earlier days the Eskimo lived exclusively on meat and fish, most of which they derived from the sea; and the coastal waters of Alaska and Greenland have always been richer in sea mammals than those of arctic Canada. Increasing contact with the civilized world has greatly diminished this dependence on the local food resources.

More important than the number and distribution of the population in the 1950s, perhaps, was the fact that it was no longer stationary or declining, but rapidly increasing. The rate was fast-

est in Greenland, 2% as compared with, roughly, 1% in Canada and .75% in Alaska. There was every reason to believe that this upward trend would continue, since famines, infanticide, brood feuds and other factors that kept down the populations in past centuries no longer prevailed. In some districts, however, any considerable increase would strain the local resources, a situation that had already occurred at Ammassalik in east Greenland. That settlement reacted by colonizing Scoresby sound, a district farther north that had been occupied in pre-European times and later abandoned; but Greenland possesses very few areas suitable for colonization, and already the population of the island was showing signs of needing living space.

Eskimo are not a homogeneous people physically, not even those of pure blood. The average stature of the men, for example, ranges from 62.2 in. in some groups to 66.1 in. in others, and the cephalic index from 74 to 82. Nevertheless, they can generally be distinguished from Indians by their more Mongoloid features, by the smallness of their hands and feet and by a few less obvious traits such as shovel-shaped incisor teeth and an exceedingly narrow nose. Another anomaly is the appreciable percentage among them of the B blood type, which seems to be totally absent from their Indian neighbours. Since this is a very stable hereditary trait, one suspects that at least a part of the Eskimo population had a different origin from the Indians, and that they are not, as some writers have thought, merely Indians transformed. Their language supports this theory. It is highly inflecting with a large number of suffixes that can be combined in an almost infinite variety of ways so that tens, and even hundreds, of thousands of words, all with different meanings, can be derived from a single root. The sound system is relatively easy for Europeans, differing from German mainly in a distinction between a palatal and a uvular *k*; and there are no tones such as are found in the neighbouring Athapascan Indian dialects, and also, of course, in the Chinese. Indeed, the Eskimo language appears to bear no resemblance to any Indian tongue, and attempts to link it with the Ural-Altaic languages of northern Eurasia have not met with general acceptance. More promising, perhaps, is the theory of the Danish philologists W. Thalbitzer and L. L. Hammerich, who suggested that it may be a remote offshoot of the Indo-European linguistic stock.

Because of the language's complexity, few adult foreigners have ever learned to speak it fluently; and in northern Alaska the natives themselves are beginning to simplify it by dropping a few of the suffixes. In that region, toward the end of the 19th century, intercourse between Eskimos and whites was facilitated by the development of a special jargon possessing a vocabulary of from 300 to 600 words, most of them derived from Eskimo, but some from English, Hawaiian and perhaps other tongues. It was this jargon, not the real tongue, that one or two incautious travelers designated the Eskimo language. Although the Eskimo maintained their language fairly intact, their daily life and thoughts were greatly modified by contact with whites, especially in Greenland and Labrador, where the contact dates back many centuries. Certain groups along the arctic coast of Canada, however, remained practically untouched by the outside world until the 20th century, and even in the 1960s a few Eskimo there still were maintaining the life of their forefathers almost unaltered.

Early Life and Customs.—Only a people of great ingenuity and endurance could have survived in a region that lay rigid under snow and ice for six to nine months of the year, a region where vegetable foods were unprocurable, trees existed only in one or two marginal districts and large areas lacked wood of any kind, even driftwood, from which to make houses, tools and weapons. The Eskimo had to rely for their food solely on fish, on the land animals (principally caribou, but also musk oxen, bears, hares and marmots) and on the mammals of the sea (four or five species of seals and, in certain regions, beluga, walruses and whales). In winter, accordingly, a dozen or more families congregated in a small community from which the men sallied out each morning to harpoon such seals as exposed their heads in lanes of open water, or rose to breathe at the small holes they kept open for themselves in the sea ice. During the long days of spring these communities broke up. Some families pursued the seals over the open water in kayaks, others wandered off to hunt the mighty bowhead whale, while still others fished through the ice of the still-frozen lakes or trapped the salmon trout that migrated inland from the sea at the breakup of the streams and rivers. Later, in midsummer, most families hunted the caribou when the fur of that animal was in its prime, and some carried on this hunt into the late autumn when the herds were migrating to new pastures, often several hundred miles away. Then, with the darkening days of winter, the families drew together again, bringing their yearly cycle to a close. It was a cycle, of course, that was never absolutely uniform, but varied from year to year and from region to region in accordance with the local conditions.

Against land animals the Eskimo used the bow and arrow, the former a composite weapon of wood, horn or antler backed with twisted sinew to give the necessary tension. (The rifle later replaced it everywhere.) Against sea mammals, on the other hand, their weapon was the retrieving harpoon, supplemented, when employed against whales, by three large floats of inflated sealskin which, attached to the long line leading from the harpoon head, gradually exhausted the wounded animal. Some groups of Eskimo also used rawhide nets for capturing seals but, strangely enough, never adapted them for fishing, being content with the universal hook and trident and with stone weirs and traps in shallow streams. The skin-covered, one-man kayak from which they harpooned the seals in the open sea and speared the swimming caribou in the lakes and rivers was so efficient a craft that Europeans later imitated it in their fold boat. It was not practical, however, for whaling. For that the Eskimo employed, in summer, the umiak, a larger boat, also skin-covered, which resembled the European whaleboat and which served on other occasions to transport the family and its possessions. Winter transport was provided by dog sled, which seems to have been an Eskimo adaptation of the Siberian reindeer sled. More distinctive, perhaps, than the methods of hunting and traveling were the clothing and the dwellings. Only clothing made of fur, fully tailored, offered adequate protection against the rigorous climate, and of the various furs available the Eskimo preferred that of the caribou, although they sometimes used those of other animals such as the seal, polar bear, mountain sheep and hare and, in certain places, even bird skins. The costume itself varied somewhat from region to region, but always comprised a coat, trousers, stockings and either shoes or boots, two of each garment being worn in cold weather, the inner with the fur against the body, the outer with the fur outside. The outer shoe or boot nearly always was made from sealskin, because, unlike caribou fur, it does not spoil with dampness. Women wore garments similar to those of the men, but cut differently. Many ski enthusiasts adopted the coat of the Alaskan Eskimo and even retained its Eskimo name, parka.

In the western arctic, where the Mackenzie and Yukon rivers had strewn the coast with driftwood, and also in Greenland, where driftwood brought by ocean currents was much less scarce than in modern times, the Eskimo built log cabins, which they insulated on the outside with turf and illumined with sky lights made from gut. One entered by a semiunderground passageway through a trap door in the middle of the floor. When the drifting snow of winter piled over and around these houses, they could be quite comfortable; but in spring many of them became flooded with water and uninhabitable, and the occupants moved into tents of seal or caribou skin. Where wood was scarce some families constructed houses of stone, which they roofed with whale ribs and turf. From Coronation gulf to Labrador, however, the Eskimo spent the winter in round huts built from snow blocks, the well-known igloos, which seem to have been invented by the Canadian Eskimo, since they were unknown in Alaska and little used in Greenland. Given snow of the proper consistency, and a long-bladed knife of bone, ivory or metal, one man could erect within an hour a snowhouse large enough to shelter his entire family. The inmates slept on a low platform of snow, which they covered with a mattress of willow twigs and overlaid with caribou furs; and they illumined their home with a skylight, not of gut, but of translucent fresh-water ice cut expressly for this purpose at the beginning of winter. Like their kinsmen who lived in log cabins, however,

when the warm sun of spring began to melt the roof over their heads, they too abandoned their vanishing huts and established themselves in skin tents. The Eskimo solved the problem of heat and light in their dwellings by burning animal fat (nearly always seal blubber) in shallow saucer-shaped lamps, made from pottery throughout most of Alaska and from stone elsewhere. They boiled their meat and fish over these lamps unless, as frequently happened, they ate their food frozen or dried. Raw fish or meat that was neither dried nor frozen they disliked almost as much as do Europeans. They are rather maligned, therefore, when they are called Eskimo, a name derived apparently from a Cree Indian word meaning "eaters of raw meat." Their own name for themselves is simply *inuit* ("people").

In a single-roomed hut of logs or snow there can be no privacy, nor can a man accumulate many possessions. Except for his dogs —which were less numerous formerly than in modern times, when firearms and European foods simplified the problem of feeding them—an Eskimo could claim as his private possessions only the things he made for himself or derived, either in trade or as gifts, from their makers: such things, for example, as his tools and weapons, his sled and his kayak. Even the house he constructed with his own hands was his only so long as he occupied it. He could neither sell it nor prevent another family from taking possession after he himself had ceased to require it. The land belonged to all and food had to be shared in common, at least theoretically, since fishing and the chase are precarious pursuits and no family dared deliberately allow its neighbour to starve when it might itself face starvation a few days or weeks later. There were no written laws, of course, and no chiefs or police to enforce the traditional rules handed down from one century to another. Eskimo society was amorphous; its only clearly defined unit was the family. Yet two powerful factors tended to hold it together: public opinion and fear of a blood feud. In the blood feud not only the malefactor was pursued but also his kin, and since it often continued for several generations, it played no insignificant part in keeping the population stationary. The Eskimos themselves recognized its banefulness, and in Greenland they tried in vain to curb it with the song duel, a public gathering at which two enemies loudly voiced their quarrel by lampooning each other in song.

There was a fairly sharp division of labour between the sexes. Men hunted and built the homes; women cooked the food, dressed the skins and made the clothing. Each sex was therefore so dependent on the other that celibacy was unknown and widows and widowers remarried as soon as they were able; those too old to remarry found shelter with kindred. A man usually had only one wife, but the more active and ambitious occasionally took a second and even a third, especially if the first wife was childless; and a woman occasionally had two husbands, though she seldom retained both more than a few weeks or months. Orphans invariably found homes, preferably with near relatives, and in most cases were well-treated; but old folk tales relate that they were sometimes degraded to menials and shamefully abused until they became old enough to fend for themselves. Most Eskimo were very fond of children, even though they did not hesitate to destroy newborn babes, especially girls. In a land where there were no vegetable foods and no roads a mother had to nurse her child and carry it everywhere on her back until it was at least three years old; and the care of a second baby during that period would have been beyond her strength.

Though all, or nearly all, Eskimo were at least nominal Christians by mid-20th century, they continued to interpret their universe spiritually; that is, they saw some spiritual force in every phenomenon of nature, in sticks and stones, in the wind and the snow, in every beast and bird; and they endowed all these with human characteristics. In earlier times most of the forces seemed to them neither friendly nor hostile; they existed, just as the Eskimo existed. Some, however, could become dangerous if not treated with due respect, and a few were always hostile. The shades of Eskimo dead went to join this spirit world, and they too might be hostile, even to their surviving kin. It was extremely important never to offend the spirits of the game animals, since they might punish the Eskimo with sickness or with famine. Hence, there arose countless regulations and tabus, those relating directly to food being the most numerous and complicated. They differed somewhat from region to region but everywhere had the same purpose, to avoid antagonizing the innumerable spirits and thus preserve harmony between man and his environment.

Near the magnetic pole, and again in the extreme north of Alaska, K. Rasmussen was told of a sky-god, Sila, who seemed to have taken on some of the attributes of a supreme being; but one suspects this was an echo of Christianity. The Eskimo of the eastern arctic elevated a female supernatural being, Sedna, to the rank of a real deity because they believed that she controlled the supply of seals; yet even to her they did not pray or offer worship but tried to control her through incantations and through the magic power of their shamans. The latter were men and women claiming familiar spirits that gave them power to travel invisibly to distant places, to see and to know what was beyond the sight and ken of ordinary humans, to cure sickness and to mediate between the lay Eskimo and the supernatural world. They were thus both the physicians and the priests of their communities. But they might also be the sorcerers, and for that reason were blamed for many calamities that seemed to have no physical cause.

In spite of these rather gloomy beliefs and the severity of their environment, the Eskimo were, and still are, a most cheerful people, contrasting in this respect with the adjacent Indians. It requires only a smile or a joke to induce them to confront the greatest perils without complaint or fear. Unquestionably it was this cheerful temperament which in no small measure enabled them to survive and even find happiness in lands that other peoples have judged uninhabitable.

Their origin is an enigma, but by digging into the ruins of their ancient settlements scattered in considerable numbers from Alaska to Greenland, archaeologists have discovered that they have a long and complex history on the American continent. (See *Eskimo Archaeology,* below.)

BIBLIOGRAPHY.—Franz Boas, in the *Bulletin of the American Museum of Natural History* and *The Central Eskimo* (1888); L. M. Turner, *Ethnology of the Ungava District* (1894); E. W. Hawkes, "The Labrador Eskimo," *Memoir 91* of the Geological Survey of Canada (1916); K. Rasmussen, *Across Arctic America* (1927), a greatly abridged translation from the Danish; V. Stefansson, *My Life With the Eskimo* (1913), *The Friendly Arctic* (1921), *Hunters of the Great North* (1922); D. Jenness, *The Life of the Copper Eskimos* and several articles in *Report of the Canadian Arctic Expedition of 1913–18* (1922); J. Murdoch, *Ethnological Results of the Point Barrow Expedition* (1892); E. W. Nelson, *The Eskimo About Bering Strait* (1899); E. M. Weyer, *Eskimos: Their Environment and Folkways* (1932); C. E. Whittaker, *Arctic Eskimo* (1937); K. Burket-Smith, *The Eskimos,* rev. ed. by C. Daryll Forde (1959); P. Freuchen, *Book of the Eskimos* (1961). (D. J.)

ESKIMO ARCHAEOLOGY

Eskimo archaeology had its beginning in 1921–23 with the excavations of Therkel Mathiassen which revealed the existence of a prehistoric Eskimo culture, the Thule, in the areas west and north of Hudson bay. The Thule was primarily a maritime culture, based on the hunting of seals, whales, walrus and caribou. It was closer in many ways to Greenland and Alaskan Eskimo culture than to that of the present central Eskimo. Skeletal remains from Thule sites are almost identical with those from Point Barrow, Alaska, and cultural evidence also indicated this as the region from which the Thule Eskimo came. E. Holtved's excavations in Inglefield land showed that the Thule people had reached northwest Greenland in the 12th century A.D. Excavations in west, south and east Greenland traced a cultural continuity from the 13th–14th century Inugsuk, a modified form of Thule culture, to that of the 19th century. Meanwhile, in Alaska, older forms of Eskimo culture were discovered, including the Birnirk, a stage immediately ancestral to Thule.

On the basis of radiocarbon dating, the oldest known Eskimo culture is that from a site excavated by W. S. Laughlin on Umnak Island in the Aleutians, for which an age of 3,018, plus or minus 230 years was recorded. Frederica de Laguna's excavations at Cook inlet and Prince William sound in southern Alaska, provided a detailed record of culture development in the southernmost part

of the Eskimo territory. Prehistoric south Alaskan and Aleutian culture was basically Eskimo in character but specifically different in many respects from that of western and northern Alaska. It was somewhat simpler technologically and its art, particularly carving in the round, was less highly developed. (*See* IVORY CARVING: *North American.*) Though the Aleutian and south Alaskan region had the greatest population density of any part of the Eskimo territory, its culture was remarkably stable, in contrast to the Bering strait region where culture changes were frequent and pronounced. Influences from south Alaska seem to have extended to southern British Columbia in Indian territory, where the earliest culture exhibits Eskimolike features. The little information available for the Bering seacoast indicates that Eskimo culture in this region rested on an Ipiutaklike foundation, to which Punuk features were added.

In western and northern Alaska—Norton sound to Point Barrow—the principal prehistoric Eskimo culture stages are Old Bering sea, Ipiutak, Birnirk and Punuk. The first traces of Old Bering sea culture were discovered by Diamond Jenness on Little Diomede Island in 1926. Later excavations on St. Lawrence Island brought full information on this old but highly developed Eskimo culture and on a later phase, the Punuk, which was contemporaneous with the Canadian Thule culture. Radiocarbon dating gives the age of the earliest phase of Old Bering sea culture, called Okvik, as 2,258, plus or minus 230 years. Radiocarbon and tree-ring analysis give ages ranging from 700 to 1,350 years for Punuk. The Old Bering sea Eskimo were a typical maritime people who lived in permanent, semisubterranean houses and who depended for their livelihood mainly on seals, walrus, birds and fish. Though it was the oldest Eskimo culture known from northern Alaska, the Old Bering sea was more highly developed, particularly in its elaborate art style, than any later Eskimo culture with the exception of Ipiutak. The Punuk culture, which followed it, was in part an outgrowth of Old Bering sea and in part the result of new influences from Siberia. The freehand circles and graceful flowing lines of Old Bering sea art were replaced by simplified Punuk designs consisting of straight, evenly incised lines and circles made with engraving tools tipped with Siberian iron. Okvik has been found on the St. Lawrence and Diomede Islands, on the Siberian coast opposite Bering strait and on the Alaskan mainland where it occurs as a component of the Ipiutak culture at Point Hope, Kotzebue sound and Seward peninsula. The later phases of Old Bering sea, as represented by its art styles II and III, flourished on the Siberian coast and the nearby St. Lawrence and Diomede Islands; on the Alaskan mainland it has been found only at the Ipiutak site at Point Hope. Punuk culture centred on the St. Lawrence and Diomede Islands and northeastern Sibera. It also occurred at Cape Prince of Wales and a few other localities on the Arctic coast of Alaska.

The Ipiutak culture, discovered by H. Larsen, F. Rainey and J. L. Giddings, Jr., at Point Hope in 1939, presents a number of problems. The two radiocarbon dates that have been obtained for this culture—912 and 973 plus or minus 170 years—are probably too recent. In view of the many and close resemblances between Old Bering sea and Ipiutak, an age of around 1,600 to 1,700 years for the latter would seem probable. Ipiutak implements of ivory, antler and bone are more elaborate, both in form and decoration, than those of any other Eskimo culture. Connections with the Siberian Iron Age are indicated by similarities in burial practices and ceremonialism, by animal carvings and designs that recall Scytho-Siberian art and by the presence of a few fragments of iron of Siberian origin. The chipped stone implements from Ipiutak, on the other hand, are more archaic in character, resembling those from Neolithic sites in Siberia as well as early Eskimo and pre-Eskimo sites in the American arctic. The Ipiutak culture lacked such typical Eskimo features as sleds, harpoon floats, bow drills, pottery, lamps and rubbed slate implements. The last three, however, were present in a closely related stage of culture called Near Ipiutak. Ipiutak proper is known only from Point Hope, Kotzebue sound and Seward peninsula, but a related form of culture was found by Larsen in the Bristol bay-Kuskokwim area and by Giddings at Norton sound, the latter dated by radiocarbon at 1460 plus or minus 200 years and 2016 plus or minus 250 years.

Before the Thule Eskimo moved eastward into Canada and Greenland some 700 or 800 years ago, these regions had been occupied for centuries by another group of Eskimo, the Dorset. The Dorset culture was first described by D. Jenness in 1925 on the basis of material that had been excavated by Eskimos at Cape Dorset on the south coast of Baffin Island and on Coats and Mansel islands in Hudson bay. Later excavations showed that the Dorset culture had been widely distributed through the eastern arctic from Newfoundland to east Greenland. It differed in many ways from the Thule culture. Dorset harpoon heads and foreshafts, knives, lamps, chipped stone implements and ivory carvings were distinctive in form and smaller and more delicately made than those of the Thule people. The Dorsets also lacked the dog sled, bow drill, bone arrowheads, ulus and other typical Thule features. As no skeletal remains have been found, the physical type of the Dorset people is unknown. It is questionable to what extent the Dorset may have influenced historic Canadian and Greenland Eskimo culture. However, the dominant sealing harpoon head in these regions was demonstrably derived from a Dorset prototype. The same is true of the side-bladed lance of the Sadlermiut Eskimo who lived on Southampton Island until their extinction in 1903, and it is even possible that these Eskimo themselves were descendants of the old Dorset people. It is not certain when the Dorset culture disappeared but it was after the arrival of the Thule migrants from Alaska, for there are some indications of contact between the two groups. A long-time span for the Dorset culture is thus indicated, as shown by a radiocarbon date of 2632, plus or minus 128 years, for an early or proto-Dorset phase of culture at Native point, Southampton Island. A stratified midden at Sermimiut, Disko bay, west Greenland, yielded a radiocarbon date of A.D. 30 for a Dorset occupation level underlying Inugsuk and of 600 B.C. for a still lower level representing an older form of culture called Sarqaq. Traces of the latter have also been found in the Canadian arctic but, as it has not been described in full, its relationship to Dorset is uncertain. In another direction there are strong affinities between Dorset and the Denbigh Flint complex and other pre-Eskimo flint sites in Alaska and western Canada which suggest a western—Alaskan—origin for the Dorset culture. The Denbigh Flint complex was discovered in 1948 by Giddings at Cape Denbigh on the Bering seacoast of Alaska. It is dated by radiocarbon as from 3,500 to 5,000 years old, though it may be somewhat older. The Denbigh complex was characterized by microblades (lamellar flakes), tiny blades made from such flakes, finely chipped side and end blades, burins and burin spalls. All of these types occur in the Dorset culture as well as in the Mesolithic and early Neolithic of Siberia. It also seems likely that some of the Denbigh end and side blades were ancestral to those found at Ipiutak. The Denbigh complex is represented at several sites in interior Alaska and western Canada; in the latter region R. S. MacNeish has found Denbigh overlying earlier implements resembling those from early man sites in the western United States. The discovery of the Denbigh Flint complex and its relationships both to Mesolithic culture in Eurasia and to early stages of Eskimo culture in the American arctic reinforce previous indications that Eskimo culture as a whole was of Mesolithic origin. *See* also references under "Eskimo" in the Index volume.

BIBLIOGRAPHY.—*For Canada and Greenland:* T. Mathiassen, *Archaeology of the Central Eskimos* (1927); E. Holtved, *Archaeological Investigations in the Thule District* (1944); D. Jenness, "A New Eskimo Culture in Hudson Bay," *Geogr. Rev.*, 15:428–37 (July 1925); J. Meldgaard, "A Paleo-Eskimo Culture in West Greenland," *Amer. Antiq.*, 17:222–30 (Jan. 1952); E. Knuth, "The Paleo-Eskimo Culture of Northeast Greenland Elucidated by Three New Sites," *Amer. Antiq.*, 19:367–81 (April 1954); R. S. MacNeish, "The Pointed Mountain Site Near Fort Liard, Northwest Territories, Canada," *Amer. Antiq.*, 19:234–53 (Jan. 1954); H. B. Collins, "Recent Developments in the Dorset Culture Area," *Amer. Antiq.*, 18:32–39 (Jan. 1953), "The T 1 Site at Native Point, Southampton Island, N.W.T.," *Anthropological Papers University of Alaska*, 4:63–89 (May 1956). *For Alaska:* F. de Laguna, *The Archaeology of Cook Inlet, Alaska* (1934), *The Prehistory of Northern North America as Seen From the Yukon* (1947), *Chugach Prehistory: the Archaeology of Prince William Sound, Alaska* (1956); W. S. Laughlin, "The Aleut-Eskimo Community," *Anthropological*

Papers University of Alaska, 1:24–48 (Dec. 1952); H. B. Collins, *Archaeology of St. Lawrence Island, Alaska* (1937); H. Larsen and F. Rainey, *Ipiutak and the Arctic Whale Hunting Culture* (1948); H. Larsen, "Archaeological Investigations in Southwestern Alaska," *Amer. Antiq.,* 15:177–186 (Jan. 1950); J. L. Giddings, "The Denbigh Flint Complex," *Amer. Antiq.,* 16:193–203 (Jan. 1951); *The Arctic Woodland Culture of the Kobuk River* (1952). *For Siberia:* S. I. Rudenko, *The Ancient Culture of the Bering Sea Area: the Eskimo Problem,* in Russian (1947). *General:* D. Jenness, "Prehistoric Culture Waves from Asia to America," *Smithsonian Report for 1940* (1941); H. B. Collins, "Outline of Eskimo Prehistory," *Smithsonian misc. Coll.* (1940), "The Origin and Antiquity of the Eskimo," *Smithsonian Report for 1950, Arctic Area* (1951), *Arctic Area,* Program of the History of America, Comisión de Historia, Mexico City (1954), "Archaeological Research in the North American Arctic," *Arctic,* 7:296–306 (1954).

(H. B. Cs.)

ESKIMO-ALEUT LANGUAGES. As early as 1814, the Danish linguist Rasmus Rask observed similarities in basic vocabulary and in the inflections of Eskimo and Aleut that led him to the inference that they had come from a single earlier language, but some time later scholars doubted the theory. About the mid-20th century several additional studies greatly strengthened Rask's idea. The degree of relationship is perhaps comparable to that between English and Russian; it is estimated that the two may have been a single language up to about 4,000 years ago.

In the 1950s, there were about 5,000 to 6,000 speakers of Aleut, several hundred on the Commander Islands off the coast of Siberia and the rest on the Aleutian chain and on the Alaskan mainland. Two principal dialects are recognized: eastern, or Unalaskan, and western or Attuan; Atkan is a variant of the latter.

Eskimo speakers include about 1,000 in Siberia, 16,000 in Alaska, 10,000 in Canada and 25,000 in Greenland. Though the variations in the language are small considering the large territorial spread, it is not one language but at least two. As the Danish-Eskimo explorer Knud Rasmussen discovered, a speaker of Greenland Eskimo can make himself understood, despite some dialect differences, through Labrador, the arctic coast of Canada and northern Alaska. In the rest of Alaska and the eastern tip of Siberia, the language is about as different from the first type as English is from Dutch. Within this second kind of Eskimo the variations are greater than in the first, suggesting that the language has been in use there a much longer time and that there has been a fairly rapid spread east from Alaska during the last several centuries.

Different local forms of Eskimo are referred to by villages or regions, as Greenland, Labrador, Mackenzie, Barrow, East Cape (Siberia), Kuskokwim, Nunivak. There are no established names to distinguish the main types, but some scholars have spoken of Inuk and Yuk Eskimo, *juk* (*j* equivalent to English *y*) being a simplified form of the word for "human being" as used in southern Alaska and in Siberia, and *inuk* being the fuller form used elsewhere. The people are sometimes called Innuit and Yuit, based on the plural of the same roots.

Relationships.—Rask recognized a further relationship of Eskimo-Aleut with the Ural-Altaic family. Other scholars attempted to link it with Indo-European. It was long thought that Eskimo-Aleut had no relationship with American Indian languages, despite a few points of similarity (such as the dual suffix *-k* of Eskimoan and the *-ki* plural of Algonkian). Evidence was brought forward in the late 1950s, however, to show systematic agreements pointing toward common origin, particularly with the Wakashan stock of the northwest coast of North America. Furthermore it was shown that Eskimo-Aleut has affinities with the Chukchian languages of Siberia. The theory proposed, but still inadequately established, was that Wakashan, Eskimo-Aleut, Chukchian, Yukagir, Ket, Uralian and Altaian form a phylum, "Altai-Wakashan," whose common period dates back about 9,000 years. Relationships with Indo-European, Koreo-Ainuan, Penutian, Salishan and others in both Eurasia and America must be attributed to a still older time.

Alphabet.—Until recent centuries, the only writing known to Eskimos and Aleuts was that employing a sequence of pictographs. From 1721 onward Greenland Eskimo has been written both in printed publications and in the schools in an alphabet based on Latin. In Canada the Cree syllabary, consisting of a set of geometric signs drawn in different positions according to the vowel, was adapted to Eskimo around 1885. In Soviet Siberia, Eskimo was written in Latin letters beginning around 1939, but later the Russian (Cyrillic) characters were used. In other Eskimo areas there has been sporadic use of alphabetic writing.

For Aleutian the Russian missionary I. Venyaminoff introduced a Cyrillic alphabet in 1829, which has continued to be used in Alaska into the middle of the 20th century. In the Soviet Union, the written form of Aleut had the same history as Eskimo, using first Latin and later Cyrillic characters. Soviet publications of 1934 reported that the Aleuts were 99% literate.

Scientific works on Eskimo-Aleut generally attempt to regularize the spelling and use simple signs to replace digraphs in the practical alphabets, especially ŋ instead of *ng* for the velar nasal (as in *sing*) (in this article the alphabet of the International Phonetic association is used, with modifications).

Phonology.—Eskimo and Aleut have simple systems of basic sounds, or phonemes, some of which vary according to the sounds that precede and follow. The variability accounts for the complicated appearance of some written records and for the differences in the representation of the same word by different authors. There are three basic vowels, usually written *i a u* (as in "machine," "fa," "Lulu," but the first and third vary toward the sounds of "pet" and "go") and from 13 to about 20 consonants, depending on dialect. Inuk consonants include the stops *p t k q;* the light fricatives *v j l g* ʁ; a strong fricative *s,* replaced by *h* in parts of Canada, and an additional tongue-blade sibilant *ś* in Greenlandic; nasals m n ŋ. Frequent and typical are the back velars *q* and ʁ corresponding to *k* and *g* (fricative, like Spanish intervocalic *g* in *vago,* not stopped as in English) but produced at the deepest part of the mouth by friction between the posterior tongue and the soft palate. The light fricatives in Inuk vary between voiced and voiceless according to position. In Yuk there is a separate set of voiceless fricatives, *f ʃ ɬ x* χ, and an *r* produced with a tap of the tongue; *c* (sound of ch) regularly replaces *s.*

Aleut has the same phonemes as Inuk Eskimo with the addition of *c, d* and *h,* and the omission of *ś* and *ŋ;* although voiceless fricatives are sometimes written, they may represent positional variants of the light fricatives. Aleut *h* regularly corresponds to *p* of Eskimo, as Al. *hujuq,* Esk. *pujuq* "smoke," and the latter must be the archaic sound. Aleut *d* probably represents an original distinct phoneme, which fell together with *t* in Eskimo; *e.g.,* Al. *adaq,* Esk, *ata* "father." The Aleut contrast of *c* against *s* likewise must be archaic. For proto Eskimo-Aleut, a fourth vowel, **e,* must be inferred to account for correspondences of Eskimo *i* to Aleut *a, e.g.,* Al. *akajuq,* Esk. *ikiq* "strait."

The Eskimo-Aleut languages are characterized by notable restrictions in the positions in which phonemes occur. In Inuk a word may begin or end only in a vowel or in a single consonant, and there are never more than two consonants together within a word. Light fricatives and also ŋ and *ś* never begin a word; the only consonants at the end of the word are *q, k* and *t,* except for some dialects in which ʁ, *g* and *n* are found instead. In Yuk the rules are somewhat different, because of the loss of **e* when followed by a single consonant. The situation in Aleut is similar to that of Yuk. As for the vowels, in most dialects there are groups only of two, including double vowels (pronounced long) *ii uu aa* and combinations like *ai au ui iu;* eastern Inuk is exceptional with groups up to five resulting from the loss of certain light fricatives between vowels. Syllables ending in a consonant or in a vowel group receive a heavier stress, but the accent never has independent value, being in this respect as in French or Polish rather than English. The simple structure of syllables and the alternation between heavy and light stress, give Eskimo an effect which has been compared to the constantly changing rhythm of a drum.

Morphology.—Eskimo-Aleut languages are highly inflectional, going beyond classical Greek in the number of its noun cases and verbal modes and having more complex pronominal endings. Everything is expressed by suffixes, with changes in the stem only for the plural of some nouns. In Eskimo, which has more inflections than Aleut, there are 312 inflectional combinations for the noun, and almost 1,000 for the verb. Since the endings are mostly

regular, however, they probably represent less complication than the formations of Greek.

Besides singular and plural there is a dual, as Esk, *iglu* "house," *iglu-t* "houses," *iglu-k* "a pair of houses." There are six adverbial cases expressing ideas like "at," "from," "to," "through," "with," "like." There are two syntactic cases: the relative or superordinate marks the subject of a transitive verb or the possessor of an object; the absolute is used for an intransitive subject or for a direct object. Besides number and case, the noun takes a possessive inflection, as *iglu-ga* "my house," *iglu-ɢsi* "your house," *iglu-a* "his (or her) house," *iglu-ni* "his own house," and so on through the dual and plural of both the underlying noun and the pronominal possessor. The pronoun system includes, besides the usual three persons, a repeating or reflexive third person, referring to one already in the context as against a third person newly referred to.

Verbal endings show mode and number; tense is left unexpressed or is given by a separate suffix apart from the inflectional ending. The modes include indicative, interrogative, optative-imperative, subjunctive, subject-participial and object-participial. The conjunctive expresses the idea of *when* or *if it is*, the subjunctive *if it should be*. The two participles express accompanying actions respectively of the subject and the object of the main verb. Pronoun endings are similar to those of the noun, *e.g.*, *taku-va-ga* "I saw (or see) it," *taku-va-ɢsi* "you saw it," *taku-va-a* "he saw it." These examples are of double pronominal reference, indicating both subject and object; with different endings, only the subject may be shown, *e.g.*, *taku-vu-ŋa* "I saw."

Eskimo-Aleut languages use virtually no prefixes and no internal vowel change, but make extensive use of suffixes. These often convey ideas expressed by separate words or by phrases in English, such as "inhabitant," "instrument," "place," "not," "intend to," "find," "completely," "repeatedly," and many more. Most suffixes are added before the inflection; certain conjunctive and emphatic elements follow it. Words are sometimes quite long, as *kalaalli-suu-ŋuɢ-tik-kum-i-naa-ssa-ŝu-t* "that they would not easily let themselves become like Greenlanders," whose successive parts are Greenlander-like-become-allow-easily-not-done-would-being-they." As may be seen in the example, the combination of meanings proceeds in a direction opposite to English syntax, from the last element toward the first, being comparable to "water repellent" rather than "repels water."

A number of interesting differences may be noted in the expression of ideas in Eskimo and English. The Eskimo sometimes avoids the use of "I" by saying "one wishes to go home." Some simple words of English are given by the negative of the opposite and vice versa, as "not bad" used as the normal expression of "good" or "to be not unknowing" for "to know." There are simple words for "falling snow," "snow on the ground," "drift snow," "crusted snow"; one might compare English "rain," "puddle," "flood," "ice," which give the chemical compound H_2O in different forms. Eskimo counting has various references to the use of the fingers and toes as: 6 is "one on the second hand," 11 is "one on the first foot," 20 is "the whole person."

BIBLIOGRAPHY.—K. Bergslund, *A Grammatical Outline of the Eskimo Language of West Greenland* (1955), "Aleut Demonstratives and the Aleut-Eskimo Relationship," *Int. Journ. of Am. Linguistics*, 17:167–179 (1951); Schulz-Lorentzen, *Dictionary of the Greenland Language* (1927); M. Swadesh, "South Greenlandic (Eskimo)," in H. Hoijer *et al., Linguistic Structures of Native America* (1946), "Unaalik and Proto Eskimo," *Int. Journ. of Am. Linguistics*, four articles in vol. 17–18 (1951–52), *Linguistic Relations Across Bering Strait* (1960); John Hinz, *Grammar and Vocabulary of the Eskimo Language as Spoken by the Kuskokwim and Southwest Coast Eskimos of Alaska* (1944); Gordon Marsh and Morris Swadesh, "Eskimo-Aleut Correspondences," *Int. Journ. of Am. Linguistics* (1951); R. H. Geoghegan, *The Aleut Language* (1944). (M. Sw.)

ESKIŞEHIR, a city in northwestern Turkey and capital of the *il* (province) of Eskişehir, is situated 125 mi. W. of Ankara, on the right bank of the Porsuk river, a tributary of the Sakarya. Pop. (1960) 153,190. It is divided by a small stream into a commercial quarter on low ground which contains the bazaars, khans and hot sulfur springs and a residential quarter on higher ground. The foundation of the present city most likely began with settlements around the hot springs in Byzantine times. Eskişehir developed rapidly after its connection with the Anatolia-Baghdad railway in 1894. Its economic life expanded after the immigration of Turks from European Turkey and especially after the war with Greece in 1922. The town has become one of the largest industrial centres of Turkey. Main industries include sugar refineries, cement and brick factories, railway equipment, aircraft and agricultural implement workshops. The town is also a centre for cotton research. The once famous meerschaum industry has lost its importance.

Eskişehir ("Old Town") is one of the most ancient urban centres of Anatolia. The Phrygian city of Dorylaeum, whose ruins are about 5 mi. S.W. of the modern town, is mentioned in connection with the wars of Lysimachus and Antigonus in about 302 B.C. It was the scene of the defeat of the Turks by the crusaders in 1097.

ESKIŞEHIR IL has a population (1960) of 389,129 and an area of 5,082 sq.mi. As a result of its favourable position between the Porsuk and the Ulu mountains, it receives relatively more rain than the rest of inner Anatolia. The region has extensive fertile plains along the Porsuk and in the south and is a granary of Anatolia. The chief products are wheat, barley, oats and sugar beets. Grapes, cotton and vegetables are grown in the deep Sakarya valley and stock raising is also of importance.

(N. TU.; S. ER.; E. TU.)

ESMARCH, (JOHANNES) FRIEDRICH (AUGUST) VON (1823–1908), German surgeon who is best known for his contributions to military surgery, including introduction of the use of the first-aid bandage on the battlefield. Born at Tönning in Schleswig-Holstein on Jan. 9, 1823, he studied medicine at Kiel and Göttingen. He graduated in 1848 and in the same year was called up as army surgeon in the war with Denmark, during which he was taken prisoner by the Danes. After the war he returned to Kiel, where he became professor of surgery (1857). He saw further military service against Denmark (1864), Austria (1866) and France (1870), when he held the rank of surgeon general.

Esmarch wrote a handbook of military surgical technique and instituted training in first aid for civilian and military personnel. His manuals on first aid were the best in their field and were widely used. A patent of nobility was conferred upon him in 1877 by his nephew, Emperor William II, to whom Esmarch became related by marrying a royal princess. He died at Kiel on Feb. 23, 1908.

(L. M. Z.)

ESMERALDAS, the northernmost province of Ecuador, bounded northeast by Colombia and Carchi, southeast by Imbabura, southwest by Manabí, and northwest by the Pacific ocean. Area 6,134 sq.mi.; pop. (1960 est.) 111,500. The province lies almost wholly on the Pacific coastal lowland and is covered by tropical rain forest. Settlement has spread inland along the valleys of the Río Santiago and of the Río Esmeraldas. The provincial capital and main seaport is Esmeraldas city (pop. [1959 est.] 20,700). An all-weather highway has been built to connect the port with settlements along the western base of the Andes. The forests yield balsa wood, tagua nuts and timber. Clearings along the rivers and the highway are used to produce bananas, cacao, coconuts and rice. (P. E. J.)

ESOPHAGUS, DISEASES OF. The function of the esophagus is to convey food from the mouth to the stomach (*see* GASTROINTESTINAL TRACT). This is accomplished by peristalsis or waves of constriction of the circular muscle of the esophagus as a part of the reflex act of deglutition or swallowing. As a consequence, the important symptom of abnormality or disease of the esophagus is difficulty in swallowing (dysphagia), which may result from an esophageal diverticulum or saclike outpouching that usually occurs in the upper part of the esophagus near its junction with the pharynx. Such pouches may become quite large, holding up to a pint of food, and may cause swelling, chiefly on the left side of the neck. These diverticula can be safely removed by a surgical operation. A gradually increasing difficulty in swallowing, especially of solid food, is the main symptom of cancer of the esophagus, which begins in the mucous membrane or inner lining and, as it grows, surrounds the esophagus and produces obstruction. It is most common in males over 50 years of age.

Portions of the esophagus or even the entire organ may be removed by a surgical operation in which the stomach is displaced upward into the chest and united to the upper end of the transected esophagus. Surgical excision of cancer by this method has been most successful when the tumour is present in the lower third of the esophagus. Carcinomas in the upper part rapidly become incurable because of extension to neighbouring tissues that cannot be removed.

Persistent difficulty of swallowing both liquid and solid foods occurs in a condition known as cardiospasm. This is essentially a functional disorder in the motility of the lower end of the esophagus, which may become enormously dilated and elongated so that it pursues a tortuous course in the chest. The difficulty in swallowing may cause marked emaciation. In mild cases repeated dilation of the lower end of the esophagus may give relief, but severe cases are best treated by a surgical operation in which the circular muscles of the lower esophagus are divided.

Varicose veins in the lower third of the esophagus are frequently present in patients with cirrhosis of the liver or in other diseases that obstruct the portal vein. These varicosities may rupture, causing severe or even fatal hemorrhage. Complete obstruction to the esophagus may occur in the newborn as the result of a congenital defect or later in life as a result of damage to the esophagus from the accidental swallowing of corrosive liquids. These strictures can usually be corrected by a surgical operation, although in some patents feeding may have to be accomplished through a direct opening into the stomach. (L. R. Dt.)

ESOTERIC, having an inner or secret meaning. This term and its correlative exoteric were first applied in the ancient Greek mysteries to those who were initiated (*eso,* "within") and to those who were not (*exo,* "outside"), respectively. They were then transferred to denote the distinction supposedly drawn by certain philosophers between the teaching given to the whole circle of their pupils and that containing a higher and secret philosophy which was reserved for a select number of privileged disciples. This distinction was probably adopted by the Pythagoreans and was also attributed to Plato and, by some late writers, to Aristotle. Esoteric in the sense of mystic is also used of a secret doctrine of theosophy, supposed to have been traditional among certain disciples of Buddhism.

ESPAGNOLS SUR MER, BATTLE OF LES, a naval engagement fought in the English channel off Winchelsea on Aug. 29, 1350, between an English fleet under Edward III and about 40 armed Castilian merchantmen under Don Carlos de La Cerda, who on his way to Flanders had captured English ships and was then returning with merchandise from Sluys. The battle was the consequence of bitter antagonism between Castilian sailors from the Biscay coast and sailors from England and the English-held port of Bayonne in Gascony: despite the inclusion of Castile in the current truce (from 1347) between England and France and attempts to achieve peace by independent negotiations, there was constant unofficial fighting at sea. In retaliation for what appear to have been heavy English losses, Edward III decided to intercept the Spanish fleet on its homeward voyage, collecting perhaps 50 ships and a strong force of soldiers, led by Edward, prince of Wales, and other notable captains. Edward was at Winchelsea on Aug. 28, and on the following afternoon engaged the enemy near the English coast. The Spaniards, whose ships were larger than those of the English, were eager for a fight. As the fleets approached, English archers and enemy crossbowmen exchanged shots; when the ships met, they grappled, and fierce fighting began. By nightfall the English had captured a substantial number of enemy ships (17 according to the best-informed contemporary), but the remainder escaped. At least two English ships were sunk, and there were many wounded. Edward III, it seems, owed his success primarily to his soldiers' courage and ability, the archers playing a decisive part. This engagement, following the battle of Sluys (*q.v.*) in 1340, contributed to Edward III's exaggerated reputation as "lord of the Sea."

See Sir N. Harris Nicolas, *History of the Royal Navy,* vol. ii (1847); and C. Fernández Duro, *La Marina de Castilla* (no date).

(J. W. Se.)

ESPAILLAT, a province in northern Dominican Republic. Area 376 sq.mi.; pop. (1960) 117,126. It is crossed by the low Cordillera Septentrional and occupies part of the northern coastal plain as well as part of the central section of La Vega Real. Though small in area, it is an important agricultural producer and is one of the most densely populated regions of the nation. It is the country's leading coffee area and also produces cacao, bananas, tobacco and corn. The province was created in 1885 from La Vega province. Its capital, Moca (pop. [1960] 13,829), was founded in 1780. (D. R. D.)

ESPARTERO, BALDOMERO (1793–1879), Spanish general and statesman, entered politics after his military successes in the first Carlist War. He was twice head of government (1841–43, 1854–56), but he showed political incompetence typical of the Spanish soldier in politics during the 19th century. He was born at Granátula (Ciudad Real) of working-class parents. His real name was Joaquín Baldomero Fernández Alvarez Espartero. At an early age he entered a seminary, but he abandoned his ecclesiastical career at 15 in order to join the forces fighting the French. When the army was reorganized at Cadiz in 1810 he studied at the Colegio de Ingenieros, but although he rose to the rank of sublieutenant in this branch of the army, he remained with the infantry. The end of the war found him at Valladolid whence he left for America in Gen. Pablo Morillo's army to put down the insurrections.

Espartero returned to Spain when he was 32, holding the rank of colonel. On the death of Ferdinand VII he showed himself a strong supporter of the queen regent María Cristina and enthusiastically joined the forces opposed to Don Carlos. He soon provided fresh evidence of his military ability and was appointed commander in chief. As reward for his victory at the battle of Luchana (Dec. 1836), which led to the raising of the siege of Bilbao, he was created conde de Luchana. Later he opened up the negotiations with Gen. Rafael Maroto which led to the convention of Vergara (1839) and ended the civil war. This success earned Espartero the popular soubriquet *el pacificador* and the title duque de la Victoria. He had begun to dabble in politics in 1836; on his return to Madrid (1840) he became head of the government and selected a cabinet of ministers who agreed with his progressive ideas. María Cristina preferred to resign the regency (Oct. 1840) rather than accept his program of reforms. Espartero was then himself appointed regent by the *Cortes* (May 1841).

Espartero's regency revealed his little understanding of politics. The progressive party was not united and when Agustín Argüelles was appointed tutor to Isabella II by the *Cortes,* María Cristina's protests from Paris gained the support of the moderates. Generals Concha and Diego de León attempted to seize the queen in Sept. 1841 and the severity with which their rebellion was crushed made the government unpopular. A republican revolt in 1842 was put down with equal harshness. In 1843 Generals Ramón Narváez and Francisco Serrano rose against Espartero and obliged him to flee to England where he lived until 1849 when he returned to Spain and lived in retirement at Logroño.

Espartero made his reappearance in politics in 1854 to share control of the government with Gen. Leopoldo O'Donnell during the so-called *bienio progresista.* He finally retired in 1856 and did not even intervene when Isabella was dethroned (1868). He refused the suggestion that he should be a candidate for the Spanish crown, but in spite of this received eight votes in the *Cortes* (1870). Subsequently he was awarded the title príncipe de Vergara, together with the style of royal highness, by King Amadeo. Espartero died at Logroño on Jan. 8, 1879.

(R. S. Ll.)

ESPARTO (Spanish Grass), *Stipa tenacissima,* a feathery-panicled grass processed to make ropes, wickerwork and fine paper. It is indigenous to south Spain and north Africa (where it is known as halfa or alfa), and is especially abundant in the sterile and rugged parts of Murcia and Valencia, and in Algeria, flourishing best in sandy, ferruginous soils, in dry, sunny situations on the sea coast. It attains a height of three or four feet. The stems are cylindrical, and clothed with short hair, and grow in clusters of from two to ten feet in circumference; when young they serve as

food for cattle, but after a few years acquire great toughness. The leaves vary from six inches to three feet in length, and are gray-green in colour; on account of their tenacity of fibre and flexibility they have for centuries been used for making ropes, sandals, baskets, mats and other articles. Ships' cables of esparto, being light, float on water, and were early adopted for use in the Spanish navy.

Esparto leaves contain 56% by weight of fibre, or about 10% more than straw, and came into use for the manufacture of paper. For that purpose they were first utilized by the French, and in 1857 were introduced into Great Britain. Esparto grass may be raised from seed, but cannot be harvested for 12 or 15 years after sowing.

Another grass, *Lygeum spartium,* with stiff, rushlike leaves, growing in rocky soil on the high plains of the Mediterranean countries, especially Spain and Algeria, is also a source of esparto.

For the manufacture of paper esparto is used dry and without cutting; roots, flowers and stray weeds are removed, and the material is then boiled with caustic soda, washed and bleached.

ESPERANTO, the most successful of many constructed languages, was created by the Russian physician and linguist Lazarus Ludwig Zamenhof (1859–1917), who believed that national and racial hatreds were exacerbated by language barriers and who sought to enable people of all races to communicate easily. His first brochure, "An International Language, by Dr. Esperanto" (in Russian), was published in 1887. This pseudonym became the name of the language, as it spread, first in Russia, then in Germany and the United States. A reliable 1952 estimate gave 280,000 Esperanto users in 32 countries. In the 1960s the total of persons speaking and understanding Esperanto may have approached 500,000. It was recognized in 1924 as a "clear language" for telegraphy and is used at international congresses, in commerce and on the radio (there were regular broadcasts from 20 stations after World War II).

The library of the British Esperanto association listed 30,000 items of original and translated literature, including seven or eight scientific journals. Esperanto thus has had remarkable success considering the apathy which faces all constructed languages.

Its basic principles include: (1) grammatical regularity; (2) use of roots selected from major European languages; (3) derivation of a large vocabulary from these, by affixation.

Phonology and Orthography.—Esperanto orthography is phonemic and represents 21 consonants, 2 semivowels and 5 vowels, as follows:

Consonants.—Stops: *p, b, t, d, k, g;* affricates: *c* (English cats); *ĉ* (catch); *ĝ* (cadge); fricatives: *f, v, s, z, ŝ* (shop); *ĵ* (leisure); *ĥ* (loch—rare, replaceable by *h*); *h;* liquids and nasals: *l, r, m, n.* Consonant clusters: word-initial; 31 groups of two; 7 groups of three (*e.g., sp, sc, ŝp, ŝm, spr, ŝpr*); many medial clusters.

Semivowels.—These are *j* (prevocalic as *y*es, postvocalic as boy); *ŭ* (postvocalic as how).

Vowels.—The vowels are *i, e, a, o, u* (roughly as in Spanish, with some latitude in pronunciation). Diphthongs (vowel + semivowel in one syllable) are *ej, aj, oj, uj, eŭ, aŭ.* In addition, all possible dissyllabic vowel-sequences occur (*a-a* and *a-u* only across word-boundaries).

Esperanto is accentless and has no intonation rules, although speakers tend toward a "generalized European" intonation. There is a fixed, noncontrastive, penultimate stress.

Grammar and Word Formation.—Esperanto sentence, clause and phrase structure follows common western Indo-European patterns.

Word Structure.—An Esperanto word is either a primary word or a composition of two or more bound morphemes, *i.e.,* one root (more than one in compound words) plus at least one affix.

Primary Words.—There is a closed list of 158 free forms which can be used as words or as roots (and in some cases as affixes): *e.g., mi,* "I"; *tri,* "three"; *kun,* "with"; *kaj,* "and"; *ĉu* (interrogative particle). Included among primary words are 45 "correlatives" systematically formed from five initial and nine final bound morphemes such as *i-* (indefinite); *ki-* (interrogative); *ti-* (demonstrative); *-am* (time); *-u* (person or thing); *-om* (quantity); hence, *iam,* "at some time"; *kiam,* "when?"; *tiu,* "that one"; *kiom,* "how much?"; etc.

Roots.—The *Complete Dictionary* of 1954 gives an open list of 7,866 bound morphemes derived chiefly from Romance and Germanic sources, *e.g., -frat-,* "brother"; *-hund-,* "dog"; *-bon-,* "good"; *-mort-,* "die."

Affixes.—There is a closed list of bound morphemes of two types, *lexical* prefixes and suffixes and *grammatical* suffixes. The 8 lexical prefixes and 31 lexical suffixes modify the lexical meaning of the root as follows: *bo-,* "relative by marriage," gives *bofrato,* "brother-in-law"; *mal-,* "opposite," gives *malbona,* "bad"; *-ac* (pejorative), gives *hundaĉo,* "cur"; *-ig* (causative), gives *mortigi,* "kill"; etc. The grammatical suffixes mark the form-class of words as follows: *-o* (noun); *-a* (adjective); *-e* (adverb); *-i* (verb, *i.e.,* infinitive); etc. They also mark other grammatical categories and relations; *e.g.,* with nouns and adjectives: *-j* (plural); *-n* (accusative); with verbs: *-u* (imperative); *-t-* (passive participle); *-nt-* (active participle); *-s* (finite); *-a-, -i-, -o-* (+ *s* or *(n)ta*) and *-u-* (+ *-s*), which show finite or participial tense and mood—present, past, future and conditional. Hence, from *ten-,* "hold," comes *tenas,* "hold(s)"; *tenis,* "held"; *tenus,* "would hold"; *tenanta,* "holding"; *tenata,* "being held"; *teninta,* "having held"; etc.

Ido and Nov-Esperanto.—In 1907, in Paris a project for a reformed Esperanto was submitted. This language, Ido (from the initials of International Delegation +*o*, the substantival suffix), sponsored by the Marquis Louis de Beaufront, made the process of derivation rigorously unambiguous by means of the principles of *monosignificance* (a single meaning for each root, however modified by affixes) and *reversibility* (the recovery of the root-meanings of derived words by removal of their affixes). These principles removed some of the arbitrariness of Esperanto derivation but, though logically satisfying, were not linguistically realistic. R. de Saussure therefore advocated a mildly reformed Esperanto (Nov-Esperanto). Ido, however, has had some success and embodies other reforms, such as the abolition of accented letters, of the compulsory accusative case, of number-agreement of adjectives and of the correlative words. *See* also UNIVERSAL LANGUAGE.

See M. C. Butler, *Step by Step in Esperanto,* 6th ed. (1943); United Nations Educational, Scientific and Cultural Organization, *Scientific and Technical Translating* (1957). (Jo. C. C.)

ESPINEL, VICENTE MARTÍNEZ (1550–1624), Spanish poet and novelist, who also enjoyed some success as a musician, was born at Ronda, Málaga, and was baptized on Dec. 28, 1550. Expelled from the University of Salamanca in 1572, he served as a soldier abroad, returning to Spain about 1584. In 1587 he took orders and later became choirmaster at Plasencia. His collection of verse *Diversas Rimas* (1591) shows technical accomplishment and caustic wit. He is, however, chiefly remembered as the author of a partly autobiographical picaresque novel *Relaciones de la vida del Escudero Marcos de Obregón* (1618; modern ed. by S. Gili Gaya, 1922), which Le Sage used as a source for *Gil Blas.* Espinel revived the *décima,* a stanza of ten octosyllabic lines, in a form since known as the *espinela.* He is also supposed to have added a fifth string to the guitar. He died in Madrid on Feb. 4, 1624.

ESPINHAÇO, SERRA DO, a range of low mountains in eastern Brazil, contains a wealth of minerals. It runs south-north from central Minas Gerais into South Bahia, forming the divide between the tributaries of the Rio São Francisco on the west and the streams that descend directly to the Atlantic on the east. Its conspicuously even summit stands between 3,600 and 5,000 ft. above sea level. It rises sharply from the hilly upland on which it stands. The Serra do Espinhaço is continued northward in the Chapada Diamantino, but in the south it ends abruptly near Counselhiero Lafaiete and is not connected with the Serra da Mantiqueira.

Since the early 18th century it has produced gold, diamonds and semiprecious stones, but it is important chiefly for its vast store of high-grade iron ore and manganese. Iron ore is mined at Itabira and Lafaiete. The range also contains known sources of chromium, nickel, tungsten, molybdenum and zirconium, and is the

world's chief source of quartz crystals. (P. E. J.)

ESPINOSA, PEDRO (1578–1650), Spanish erotic and devotional poet, was born at Antequera. When young he associated with talented groups of Andalusian writers who contributed to his anthology *Flores de poetas ilustres de España* (1605). His own early poetry, erotic and descriptive, reflects prevailing baroque influences, especially in the long *Fábula del Genil* with its double, ornamental planes of reality and transferred reality and with its jeweled colours subtly concretized. After a religious crisis, in 1606 he became a hermit, then chaplain to the duke of Medina Sidonia, and wrote religious poems, usually of less artistic ability. He died at Sanlúcar de Barrameda on Oct. 21, 1650.

See F. Rodríguez Marín, *Pedro Espinosa* (1907), *Obras de Pedro Espinosa* (1909); *Spanish Golden Age Poetry and Drama*, ed. by E. Allison Peers (1946). (I. L. Mc C.)

ESPIONAGE, as practised between states, may be defined as the attempt to obtain secretly or under false pretenses information about one government for transmission to another government. It is as old as recorded history. Five thousand years ago the Egyptians had a well-organized secret service, and the study of espionage was considered one of the many secret sciences. It is reported that during the reign of King Thutmose III a Captain Thute, with the aid of his spies, successfully smuggled 200 heavily armed soldiers into the city of Jaffa by having them sewn in flour sacks and carried into the city as a shipment of flour. In the *Iliad* Homer sings of the Trojan horse, perhaps the greatest story of espionage. In the middle ages Joan of Arc was betrayed by Bishop Pierre Cauchon of Beauvais, a spy in the pay of the English king, and in World Wars I and II the women spies Martha Cnockaert of Belgium and Anne Marie Walters of France rendered invaluable aid to their countries during the German occupations.

Spies must have certain basic characteristics, among them self-discipline, courage, unusual common sense, long-enduring patience, ability to obey orders, capability of making responsive reports and capacity for sound judgment. The individual spy is usually motivated by one of three reasons: love of his own country, hatred of a certain ideology or of a certain country or people, or need for money. He is frequently a well-trained specialist in the particular branch of espionage in which he acts. He may be expert in either political affairs or in military matters, the two main categories into which espionage is generally divided.

Political Espionage.—Political espionage is directed toward securing important information concerning politics, industry, commerce, agriculture, labour, transportation and similar matters. It is practised in both the national and international spheres. Many governments employ spies, informers and secret agents to keep them informed of the activities, opinions and relations of their own citizens and to stifle opposition. The Gestapo (*Geheime Staatspolizei,* or secret state police) in Germany during the Nazi period and the M.G.B. (ministry of state security) in Stalin's Russia are examples of organizations so used. Their agents penetrated into all strata of society and attempted to gain the confidence of all classes of people—peasants as well as prime ministers—in order to secure information.

In political espionage the spy frequently acts as an *agent provocateur*, particularly during periods of crisis such as civil wars, revolutions and religious persecutions. Skilful use of espionage was made in England by Lord Burghley and Sir Francis Walsingham in the reign of Elizabeth I; in France by Cardinal Richelieu, whose complete system of political espionage was headed by Père Joseph de Tremblay, and by Louis XV, whose master spy was Gabriel de Sartines, one-time Paris police officer who became secretary of the navy; and in the United States by both sides during the Civil War, 1861–65.

Joseph Fouché, French minister of police under Napoleon and the real creator of modern political espionage, succeeded in foiling the numerous plots against the French government during the consulate and the empire. In Russia the Okhrana, or defensive police, was founded in 1881 to combat terrorism; its activities reached their height in the period 1906–17. Beginning in 1917, after the Bolshevists seized the government of Russia, the Cheka (the Extraordinary Commissions to Combat Counterrevolution, Sabotage and Speculation) was established by Lenin and continued to employ the methods of the Okhrana until 1922 when it was replaced by the O.G.P.U. (the United General Political administration), an even more fearsome organization which was later replaced (1934) by the notorious N.K.V.D. (people's commissariat of home affairs), with agents covering the world. The N.K.V.D. was changed to the M.V.D. (ministry of home affairs), which, after the execution of its chief, Lavrenti Beria, in 1953, was renamed the K.V.D. (Committee for State Safety). In addition to the N.K.V.D. and the M.V.D., the Russian Communist party encircled the globe with the agents of the Comintern and Cominform, the international agencies of the Communist party. It was established that this Communist spy ring successfully secured secret atomic data from Canada. An increasing number of self-confessed former Communists testified to the existence of a smoothly functioning spy network permeating the various governmental agencies of the United States, the United Kingdom and France. Among these former Communists whose stories were widely circulated were Louis Budenz, once editor of the New York *Daily Worker;* Victor Kravchenko, author of *I Chose Freedom* (1946); Gen. Walter Krivitsky, a former staff officer in the Soviet intelligence; and Whittaker Chambers, *Witness* (1952).

Political espionage is often practised in the international sphere. In the 19th century after the Napoleonic Wars, the Holy Alliance covered Europe with a network of spies. It had informers, *agents provocateurs* and spies among the various groups of political *émigrés* in Paris and London. In the 20th century prior to the outbreak of the Spanish civil war, Germany, Italy and the U.S.S.R. sent many spies and agents into Spain to stir up dissension.

In peacetime the duties of diplomats include in addition to their diplomatic tasks the observation and reporting of what goes on in the countries to which they are sent. They are required to obtain accurate information on current events and on all matters that, directly or indirectly, affect the interest of the country they represent, even if such matters would normally appear to fall outside their sphere of action. The modern ambassador has a technical staff of military, naval, aeronautical, commercial and legal attachés, whose duty it is to collect accurate information on various branches of national activity in foreign countries. So long as a diplomatic agent maintains a correct attitude and does not use unfair means, such as bribing officials or stealing documents, his activities are quite distinct from those that come under the definition of espionage. It has often happened, however, that diplomatic agents, such as secretaries of legation, consuls and attachés, have failed to observe a correct attitude. In 1915, for example, the activities of Capt. Franz von Papen, German military attaché and one of the secret heads of the German spy ring in the United States, were uncovered; he was handed his passports and required to leave the United States as *persona non grata.*

Military Espionage.—Military espionage is directed toward securing in time of war or peace all possible secret information concerning the armed forces of one country by another country. The methods employed to secure such information often differ in time of peace from those employed in time of war.

Frederick the Great of Prussia has been called the father of modern military espionage. He established an extremely effective organization of spies, whom he divided into four classifications: common spies, such as peasants or people of the poor classes who spy in order to earn small amounts of money; double spies, such as agents who work for both sides and whose chief value is in conveying false reports to the enemy; spies of high rank, such as officers and ambassadors, courtiers and noblemen, who demand a great deal of money as a bribe; and persons who are forced to engage in spying against their will, such as local burghers of captured towns. Frederick issued regulations governing the recruiting of spies and made successful use of all four kinds. It was Frederick who said, "Marshal de Soubise is always followed by a hundred cooks; I am always preceded by a hundred spies." He also stated that it was easier to train a peasant as a spy than to train high-ranking army officers to disguise themselves as peasants in order to become spies.

During the Napoleonic Wars military espionage was an extremely effective weapon used by all belligerents. Napoleon had

his Karl Schülmeister, the shopkeeper and ironmonger who became the protégé and intelligence staff officer of the Austrian marshal Baron Karl Mack von Leiberich. His false reports to Mack concerning internal conditions in France and in the French army caused Mack's defeat by Napoleon at Ulm and Austerlitz. Tsar Alexander I had his Col. Alexander Tchernitchev, who secured from the French traitor Michel the complete plans for the proposed invasion of Russia by the Grande Armée. Alexander also had Colonel Figner, who, posing as a French officer, won the confidence of the French Gen. Jean Rapp and was entrusted by Rapp with carrying to Napoleon confidential dispatches, which, of course, were transmitted to the Russian headquarters. The duke of Wellington had his Col. Henry Hardinge, whose agents in France transmitted true accounts of Napoleon's plans prior to Waterloo. Wellington's successful campaign was in large measure the result of the accurate information of his intelligence officer.

Upon the outbreak of the American Civil War in April 1861, neither the Federal nor the Confederate governments had organized systems of military espionage. Following the Confederate victory at Bull Run in July 1861, which was largely the result of the accurate reports sent from Washington, D.C., by Mrs. Rose Greenhow to the adjutant general of the Confederate forces at Manassas, Va., the Federal government made use of the services of the noted detective Allan Pinkerton. Two Federal agents, Timothy Webster and Elizabeth Van Lew, moved in the highest Confederate circles. Miss Van Lew was not paid for her services but spent her own fortune in securing and transmitting information, which Gen. U. S. Grant is reported to have told her was "the most valuable information received from Richmond during the war." The Federal government apparently forgot her in later years and she died in abject poverty. The Confederate government established such an efficient system of military espionage that during the first year of the war the Confederate leaders in Richmond received within 48 hours complete reports of every meeting of Pres. Abraham Lincoln's cabinet, as well as most other Federal secrets of importance to the Confederacy. The famous southern beauty Belle Boyd made her way under fire through the Federal lines to deliver to Gen. Thomas ("Stonewall") Jackson data concerning the movements and strength of three Federal forces. Jackson quickly put her information to good use and thereby defeated the three Federal forces, whereupon he wrote to Miss Boyd: "I thank you, for myself and for the Army, for the immense service that you have rendered your country today."

In 1870 before the outbreak of the Franco-German War the Germans were reported to have had 30,000 spies in France. After the surrender of Napoleon III at Sedan the Germans organized such an immense system of espionage throughout France that the German staff had accurate knowledge of everything that took place in Paris during the siege; references to these conditions may be found in Alphonse Daudet's *Contes du lundi*.

Germany took as careful precautions for World War II as for World War I. Espionage was facilitated by the large numbers of German nationals who had settled in foreign countries. The number of espionage cases tried during World War I indicates the extent of this espionage system. The most famous trial was that of Mata Hari, tried by a French court and executed by a firing squad. She was reportedly betrayed by a German agent, Walter Wilhelm Canaris, who later became an admiral in the German navy and chief of the foreign office intelligence under Joachim von Ribbentrop. During World War I Canaris worked under Col. Walther Nicolai, head of the German high command's intelligence service. In World War II Nicolai was again chief of military intelligence and Canaris became the head of the German secret service. Their combined efforts, however, were not successful in preventing the head of the Soviet service and chief of the M.V.D., Lavrenti Beria, from securing during the course of the war reports on the latest submarine improvements used on German U-boats, and from his Communist underground information that Germany was building robot planes. The sinking of the British battleship "Royal Oak" in Scapa Flow in Oct. 1939 was successfully accomplished through one of Canaris' agents, a former captain in the imperial German navy, Alfred Wehring. Under the name of Albert Ortel, he lived for several years in Kirkwall, Scot., and, by a short-wave transmitter, sent to Canaris the information concerning the inadequate defenses of Scapa Flow.

The British secret service was the keystone of Anglo-U.S. intelligence. British agents reported fully on all important events that happened in the German-occupied countries as well as in Germany. Early in World War II the United States established the office of strategic services (OSS), which parachuted many trained agents and saboteurs into France where they aided the leaders of the *Résistance* and, according to Gen. Dwight D. Eisenhower, kept five German divisions immobilized prior to and during the Normandy invasion.

Soviet intelligence, through Richard Sorge and his unit in Japan, was kept fully informed of all major German moves both before and after Germany attacked Russia. On May 12, 1941, five weeks before the invasion of Russia, Sorge reported that 170 German divisions were concentrated along the Soviet frontier and would attack along the entire line on June 20 in the direction of Moscow. Sorge was arrested by the Japanese and executed late in 1944.

Espionage After World War II.—After the end of hostilities in 1945, military espionage and political espionage continued. Trials in Canada, England and the United States revealed how thoroughly the Soviet Union had covered its allies with a network of spies. The U.S.S.R. centred its espionage operations on foreign soil in its embassies and legations, both the M.V.D. and Soviet military intelligence being customarily represented in the embassies by, respectively, one of the secretaries and by the military attachés. In 1947 the United States established the Central Intelligence agency (CIA) to provide a focal point for the interpretation and dissemination of political and military intelligence.

A sensational espionage case opened in the United States in 1948 when Whittaker Chambers, a former Communist, charged that Alger Hiss, a former official in the department of state, had been a member of the Communist spy ring in Washington, D.C., during the 1930s. Hiss was indicted for perjury by a federal grand jury and in 1950 was sentenced to five years in prison. Another widely publicized espionage case in the United States in the early 1950s involved Judith Coplon, an employee of the department of state, and Valentin Gubichev, a Soviet consular official. In England a German-born physicist, Klaus Fuchs, was found guilty in 1950 of espionage for the Soviet Union and sentenced to 14 years in prison. Fuchs' collaborator in the United States, Harry Gold, was sentenced to 30 years imprisonment, as was Morton Sobell, who was also involved in espionage for the Soviet Union. Julius and Ethel Rosenberg were charged with having turned over atomic secrets to the Soviet government. They were tried in New York city in 1951, found guilty under the Espionage act of 1917, and after a long series of appeals were executed in 1953. The Rosenbergs were the first civilians in U.S. history to suffer the death penalty for espionage. Col. Rudolf Ivanovich Abel was found guilty of espionage in the United States for the U.S.S.R. and in Nov. 1957 was sentenced to 30 years in prison. Early in 1962 he was released to the Soviet Union in exchange for Francis Gary Powers, pilot of an American U-2 high-altitude reconnaissance plane shot down over the U.S.S.R. in 1960.

Counterintelligence.—Espionage is combated by counterintelligence, the principal tasks of which are to prevent espionage and the commission of treacherous acts against the state, and, when such acts have been committed, to search out and arrest the espionage agents responsible. According to Oreste Pinto, who operated during World War II under the cover name of Lieut. Col. Frank Jackson, and who was once called by General Eisenhower "the greatest living authority on security," the task of a counterintelligence agent is a thankless one.

Law.—International law recognizes the well-established right of belligerents to employ spies and other secret agents for obtaining information of the enemy. Resort to the practice of espionage does not involve an offense against the laws of war. Spies are punished not because they have violated international law, but only to render such methods of obtaining information as dangerous, difficult and ineffective as possible. Article 29 of The Hague regulations provides that a person can be considered a spy only when,

acting clandestinely or on false pretenses, he obtains or endeavours to obtain information in the zone of operations of a belligerent. The military codes or articles of war of most nations provide that any person who in time of war shall be found lurking or acting as a spy in or about any of the fortifications, posts or encampments of any of its armies shall be tried by a general court-martial or by a military commission, and shall, upon conviction, suffer death. During the American Revolution Maj. John André of the British army and Capt. Nathan Hale of the United States army were convicted of spying and hanged by their respective captors. Persons of all classes, whether military or civilian, without regard to citizenship or sex, are included within the meaning of The Hague regulations and such articles of war.

The Hague regulations extend a certain amount of protection by providing that a spy who is captured by the enemy after he has rejoined the army to which he belongs incurs no responsibility for his previous acts of espionage and must be treated as a prisoner of war.

See also INTELLIGENCE, MILITARY AND POLITICAL; SABOTAGE; FIFTH COLUMN.

BIBLIOGRAPHY.—Kurt Singer, *Three Thousand Years of Espionage* (1948), and *The World's Best Spy Stories* (1954); Ronald Seth, *Spies at Work* (1954), and *The Art of Spying* (1957); David J. Dallin, *Soviet Espionage* (1955); Oreste Pinto, *Spy-Catcher* (1952); Madelaine Duke, *Top Secret Mission* (1955); Francis Noel-Baker, *The Spy Web* (1954).
(W. G. D.; X.)

ESPÍRITO SANTO, a maritime state of Brazil, bounded north by Bahia, east by the Atlantic ocean, south by Rio de Janeiro and west by Minas Gerais. Pop. (1960) 1,188,665; area, 15,281 sq.mi. The western border of the state is traversed by low ranges of mountains forming a northward continuation of the Serra do Mar. The eastern part of the state belongs to the coastal plain, in great part low and swampy, with large areas of sand barrens, and broken by isolated groups and ranges of hills. With the exception of these sandy plains the country is heavily forested, even the mountain sides being covered with vegetation to their summits. The northern and southern parts are fertile, but the central districts generally are comparatively poor. The inland valleys and slopes are very fertile and heavily forested, and much of the Brazilian export of rosewood and other cabinet woods is drawn from this state. There is only one good bay on the coast, that of Espírito Santo, on which the port and capital city of Vitória (*q.v.*) is situated. The river mouths are obstructed by sand bars and admit small vessels only. The Doce, São Mateus and Itapemirim rivers rise in Minas Gerais and flow entirely across Espirito Santo. The lower courses of these rivers are generally navigable, that of the Rio Doce for a distance of 90 mi. The climate of the coastal zone and deeper valleys is hot, humid and unhealthful. In the higher country the temperature is lower and the climate is salubrious. Espírito Santo is almost exclusively agricultural, sugar cane, coffee, rice, cotton, tobacco, cassava and tropical fruits being the principal products.

There are three railway lines in operation in the state: one running from Vitória to Cachoeiro de Itapemirim (50 mi.), thence inland to Manhuassú and Carangola; another line, reaching Santo Eduardo in Rio de Janeiro state (58 mi.), where connection is made with the system running into Rio de Janeiro city; and a third railway running northwesterly from Vitória into Minas Gerais, which moves iron ore from interior mining districts to the coast for export.

The chief cities and towns of the state, with populations in 1950, were: Vitória (82,748); the port of São Mateus (6,075), on the river of the same name 16 mi. from the sea; and Cachoeiro de Itapemirim (39,470), an important commercial centre in the south.

Espírito Santo formed part of one of the original captaincies which were given to Vasco Fernandes Coutinho by the Portuguese crown. The first settlement (1535) was at the entrance to the bay of Espírito Santo, and its name was afterward given to the bay and captaincy. It once included the municipality of Campos, now belonging to the state of Rio de Janeiro. In 1799 it became an independent captaincy, in 1824 a province of the Brazilian empire, and in 1889 a state of the federal republic.

The islands of Trindade and Martim Vaz, which lie about 715 mi. E. of Vitória, belong politically to this state. They are uninhabited, but some importance was once attached to the island of Trindade because Great Britain twice attempted to take possession of it. It rises to more than 2,000 ft. above sea level and is about 6 mi. in circumference, but it has no value other than that of an ocean cable station. (R. D'E.; Js. J. P.)

ESPOZ Y MINA, FRANCISCO (1781–1836), Spanish general, was an outstanding guerrilla leader during the Peninsular War who later became noted for the harshness of his anti-Carlist reprisals during the first Carlist War. Born at Idocín (Navarre) on June 17, 1781, he enlisted in 1808 and succeeded in 1810 to the command of the guerrilla band led by his nephew Francisco Javier Mina ("Mina el Mozo") after the latter had been taken prisoner by the French. His prowess with this band led to his promotion to colonel and gained notable victories at Sangüesa (1812) and Mendívil (1813). In 1813–14 he served under the duke of Wellington. After the restoration of Ferdinand VII, Espoz y Mina was involved in an unsuccessful attempt to proclaim a constitution and had to go into exile. He returned to Spain in 1820 and was appointed captain general of Navarre and Galicia, but his support of the Liberals in Catalonia in 1822 obliged him to flee to England in 1823. He entered Spain at the head of a small force in 1830, but lack of support made him retire again to France. He returned to Spain under the amnesty of 1833 and was appointed in 1834 general in command of the northern army. Defeated in several actions against the Carlists, he resigned in April 1835. He was later given the command in Catalonia, where he carried out cruel reprisals, including the shooting of the mother of the Carlist general Ramón Cabrera. In 1836 he resigned once more and went to France where he died on Dec. 13. (R. S. LL.)

ESPRONCEDA, JOSÉ DE (1808–1842), Spanish poet, an exponent of extrovert romanticism in the Byronic manner, was born near Almendralejo de los Barros on March 25, 1808. At the age of 14 he joined a secret society and, sentenced to five years' seclusion in the Franciscan convent at Guadalajara, began an epic poem entitled *El Pelayo*. He escaped to Portugal and thence to England, where he found the famous Teresa of his poems, whom he had met at Lisbon. In 1830 he eloped with her to Paris, took part in the July revolution and soon after joined the raid of Chapalangarra on Navarre. In 1833, at the amnesty, he obtained a commission in the queen's guards; but this he soon forfeited by a political song. Banished to Cuéllar, he wrote a poor novel entitled *Sancho Saldaña ó el Castellano de Cuéllar* (1834). He took an active part in the revolutionary risings of 1835 and 1836 and in 1840 was appointed secretary of legation at The Hague. In 1842 he was elected deputy for Almería but his parliamentary career was cut short by his death, in Madria on May 23, 1842. His poems (1840) at once gained for him a reputation which continues undiminished.

The influence of Byron pervades Espronceda's life and work. It is present in an ambitious variant on the Don Juan legend, *El Estudiante de Salamanca*, Elvira's letter being obviously modeled on Julia's letter in *Don Juan;* the *Canción del Pirata* is suggested by *The Corsair;* and the Byronic inspiration is not wanting even in the noble fragment entitled *El Diablo Mundo*, based on the story of Faust. But in *El Mendigo*, in *El Reo de Muerte*, in *El Verdugo* and in the sombre vehement lines, *A Jarifa en una orgía*, Espronceda proves himself one of the most potent and original lyrical poets produced by Spain during the 19th century.

See E. Rodríguez Solís, *Espronceda: su tiempo, su vida, y sus obras* (1883); J. Cascales Muñoz, *D. José de Espronceda* (1914).

ESQUIRE (O.F. *esquier;* Lat. *scutarius*), originally a shield-bearer, a personal attendant on a knight (often his son, as in Chaucer's *Canterbury Tales*) who also carried his lance and helm. In tournaments he preceded his master bearing the knight's tilting lance in his right hand and his helmet in his left. He ranked immediately below the knight and his was in fact an apprenticeship which often led to knighthood. His office, however, was one of function and not of birth and there was nothing hereditary about his position. In the later middle ages esquire (armiger) was the customary description of holders of knight's fees who had not

taken up their knighthood—whence the surviving custom of entitling the principal landowner in the parish "the squire."

With the decay of the feudal system the title lost its original meaning and from the 16th to the 18th centuries a series of lawyers and antiquaries such as William Camden, Sir Edward Coke, John Selden and Sir William Blackstone, drew up lists of those they thought were entitled to be styled esquires. It is largely on the basis of their classification that the present position is defined. Camden was Clarenceaux King of Arms and it is the College of Arms which has issued (unofficially) the list to be found in most reference books on the subject. According to these the title is now properly given to: (1) The eldest sons of baronets and knights and their eldest sons in perpetual succession; (2) the eldest sons of the younger sons of peers and their eldest sons in like perpetual succession. For in England all the sons of peers rank as commoners at law despite their courtesy titles; (3) esquires of Knights of the Bath—each knight had the right to appoint two esquires to attend him at his installation and at coronations. Nowadays the title is extended to all commanders, companions, officers and members of any order of knighthood and to Companions of the Distinguished Service and Imperial Service Order; (4) those who qualify by virtue of their office, *i.e.*, because when appointed they are described as esquires in patents or commissions issued by the sovereign. These include Kings and Heralds of Arms, serjeants-at-law, commissioners of the court of bankruptcy, masters of the supreme court, serjeants-at-arms, justices of the peace (when in commission) mayors (when in office) sheriffs (for life), deputy-lieutenants, the principal officers of the royal household, service officers above the rank of captain in the army, lieutenant in the navy and flight lieutenant in the air force, and Royal Academicians (because George III inserted the privilege in their charter of foundation in 1768).

The list is a curious one because the crown has been arbitrary in its use of the term esquire, employing and withholding it at will. By courtesy it is now extended to all professional men and has in fact become as meaningless as the word gentleman, being frequently abbreviated *Esq.* and appended to surnames replacing prefixed titles such as Dr. and Mr. The crown did at one time create esquires by letters patent or by investiture, placing a silver collar round the candidate's neck and presenting him with a pair of silver spurs, but this procedure has long since been discontinued.

The title has been confined to the English-speaking world, but it has never been popular in the self-governing dominions and has become largely obsolete in the United States. It was at one time also accorded by courtesy to foreign noblemen but these, though their titles give them no rank in England, are now usually addressed as they would be in their own country.

In the table of social precedence esquires rank in England after the younger sons of knights and before gentlemen of coat armour, in Scotland after Queen's counsel and in Northern Ireland after the inspector-general of the Royal Ulster constabulary.

In heraldry the helmet of an esquire is always depicted in profile dexter with the vizor closed, and in formal modes of address he is styled Sir and his wife Madam. (S. B.-R. P.)

ESQUIROL, JEAN ETIENNE DOMINIQUE (1772–1840), French psychiatrist, who first sketched out the main forms of insanity, and was the founder of a great school of French alienists who developed the work he had begun, was born at Toulouse on Feb. 3, 1772. In 1794 he became a pupil at the military hospital of Narbonne, and subsequently studied in Paris at the Salpêtrière under Philippe Pinel (*q.v.*), whose assistant he became. In 1811 he was chosen physician to the Salpêtrière, and in 1817 he began a course of lectures on the treatment of the insane, which led to the appointment of a commission to inquire into the subject.

The asylums of Rouen, Nantes and Montpellier were built in accordance with Esquirol's plans. In 1823 he became inspector-general of the university of Paris for the faculties of medicine and in 1826 chief physician of the asylum at Charenton. He died in Paris on Dec. 13, 1840.

Esquirol's principal work is *Des maladies mentales, considérées sous les rapports médicaux, hygiéniques, et médico-légaux* (1838), which provides the first rational, scientific treatment of the subject.

ESQUIROS, HENRI ALPHONSE (1812–1876), French author and politician, whose writings reveal his socialist convictions, was born in Paris, May 23, 1812. He published a collection of poems *Les Hirondelles* (1834) and two novels, *Le Magicien* (1837) and *Charlotte Corday* (1840). His *Évangile du peuple* (1840), in which he put forward his political theories, earned him a term of imprisonment.

Elected to the legislative assembly in 1850, Esquiros was exiled the following year because of his opposition to the Second Empire. Returning to France in 1869 he again became a member of the assembly. Among his other works on social subjects are *Histoire des Montagnards* (2 vol., 1847) and *Histoire des martyrs de la liberté* (1851). Esquiros died at Versailles, May 12, 1876.

ESSAD PASHA (1864–1920), Albanian political leader who played a prominent part in Albanian affairs in the early years of the 20th century, came from the rich Albanian family of the Toptani in Tirana. Born at Tirana in 1864, he entered the political service of the sultan Abdul-Hamid II and when the Turkish revolution broke out in 1908 joinied the Young Turk movement and became deputy for Albania in the Turkish parliament. His influence over the Albanian population and the anxiety of the Constantinople government not to antagonize so skilful a politician secured for him in 1912 command of the forces at Shkodër (Scutari) then besieged by the Montenegrins, though he knew almost nothing of military matters. To have a free hand for secret dealings with the Montenegrins, Essad had Hasan Riza, the town commandant, assassinated on Jan. 13, 1913, and, acting as a traitor to his country, he then handed over Shkodër to Montenegro which released him after a cynical take-over ceremony (April 25, 1913). He next went to Durrës (Durazzo) and took part as minister of the interior and of war in the government formed under the newly arrived prince of Wied, but he intrigued against the government with the Italians and had to be rescued by them and taken to Italy. During World War I, he gained the support of France and was recognized by the French authorities in Salonika as head of an Albanian government in exile. His ambition was to become prince of central Albania and to hand the rest of the country over to Serbia and Greece, but on June 13, 1920, he was killed in front of the Hotel Continental in Paris by Avni Rustem, an Albanian. (T. Z.)

ESSAY. The essay, as a literary form, is usually spoken of as a peculiarly English thing, and as one of the glories of English literature. But in its origin it must always be associated with the great name of Montaigne (*q.v.*) who published the first two books of his *Essais* in 1580 when he was 47 years old. Nine years earlier he had retired to his country home and his Périgord estates to devote himself, in such tranquility as those troubled times in France afforded, to a new form of writing on which he bestowed the new title of *essais* to explain the experimental character of his work. He wrote slowly, not systematically, and it was not until 1588 that the final text of the *Essais* was published in Paris.

It has always been thought that Montaigne's genius had some affinity with the English, and when in 1603 John Florio (*q.v.*), the son of an Italian refugee living in London, published the first English translation of the *Essais* the fame of Montaigne spread rapidly through England. Dr. Johnson defined the essay as "an irregular, undigested piece," but the word "undigested" is singularly out of place to describe the work of Montaigne (or indeed that of any other essayist of the first rank such as Addison, Hazlitt, Leigh Hunt or Charles Lamb). The *Essais* are the result of profound meditation and prolonged reflection, though Montaigne was at pains to try and disguise these things. He declared that had he been minded to win the world's opinion and favour he might well have written in another vein, but he desired, he said, to delineate himself in simple and ordinary fashion. He wrote of himself with startling frankness and captured his wayward fancies about the intimate things of life with superb skill, and set them down in his own vivid and arresting way. His disarming and calculated casualness has won the approval of succeeding generations. Sir Edmund Gosse, perhaps, came nearest to a pithy summing-up of the work of Montaigne when he wrote, in his article on Essay for the 11th edition of the *Encyclopædia Britannica*

> It was in the delightful chapters of his new strange book
> that Montaigne introduced the fashion of writing briefly,
> irregularly, with constant digressions and interruptions,
> about the world as it appears to the individual who writes.

The meaning of the word "essay" has been much extended since Montaigne's day, and has been used indiscriminately in describing such different works as John Locke's *An Essay Concerning Human Understanding* and Charles Lamb's "A Dissertation Upon Roast Pig." But to the English mind the true essay is a composition of moderate length, usually in prose, which deals in an easy, cursory way with the chosen subject, and with the relation of that subject to the writer. Abraham Cowley (*q.v.*) has been called the father of the English essay of Montaigne's type, although he was not quite the earliest English essayist (*see* further ENGLISH LITERATURE: *Early 17th-Century: Prose and Poetry*). But he was the first English writer of distinction to follow the example of Montaigne and to conform to the English conception of what the true essay ought to be. His more ambitious works in prose and verse are well-nigh forgotten by the general reader, but his essays, published in 1668 (in *Essays in Prose and Verse*) are still read with pleasure. His debt to Montaigne is obvious, and in the essays "Of Greatness" and "Of Solitude" it is expressed; but the essay "Of Myself" owes almost everything to the influence of Montaigne, and indeed but echoes the latter's famous words—"it is my selfe I pourtray."

Francis Bacon (*q.v.*) is the English name worthy to stand by the side of Montaigne as an essayist, though his essays are utterly different in style and content. Three enlarged editions of the essays were published in Bacon's lifetime: the first in 1597 containing 10 essays, the second (1612) with 38 essays and the third (1625) with 58. The topics were grave and weighty—Of Ambition, Of Judicature, Of Truth, Of Great Place, and the like—and the style is lofty and at times austere. Bacon called his book "essays," he explained, because his writings were "brief notes" and not treatises; and while in the dedication to the 1625 edition he could say with pride of the essays that "they come home to men's business and bosoms," he never attained the freedom and ease, the seeming formlessness held in by an invisible chain, which are the glory of Montaigne and distinguish the typical essayist.

It would seem that at first both in England and France no lesser writer was willing to adopt a title which belonged to those two names and they do indeed stand alone in their greatness. But in the 18th century the essay had a great vogue. In 1709 the first number of Steele's *Tatler* appeared. It was a penny newspaper, and its purpose was to amuse and instruct fashionable readers by a series of short papers dealing with ordinary human affairs. When Addison joined Steele in the *Tatler* and later in the *Spectator* and discoursed of everything under the sun, grave or gay, as the fancy took him, he made the essay a notable thing to his own age, and to all succeeding ages. Many of the essays of Addison and Steele (*qq.v.*) and their imitators were a mixture of the manner of Montaigne with that of La Bruyère (*q.v.*), combining the form of the essay with that of the character study based on the *Characters* of Theophrastus, the pupil of Aristotle. Theophrastus may be said to have influenced the form of the English essay, for his *Characters* formed the model for Sir Thomas Overbury (*q.v.*), and for *Microcosmographie* (1628) by John Earle (*q.v.*), which, although written when he was only 27, shows the worldly wisdom of a much older man. Outstanding 18th-century essayists besides Addison and Steele were Henry Fielding, Dr. Johnson and Oliver Goldsmith, but their fame rests on other foundations; and, as essayists, they were all eclipsed in the first quarter of the 19th century by the appearance of one of the greatest of all essayists, Charles Lamb. *The Essays of Elia*, which began to appear in 1820, are universally recognized as the work of a man of genius. They fulfil the general reader's idea of what an essay should be to perfection. They are written with keen insight and knowledge of human nature, with humour of a rare kind and with true pathos. They are discursive in the most enchanting fashion and are full of grace and tenderness and exquisite fancy.

Leigh Hunt (*q.v.*) was a most versatile writer but of all his claims to remembrance, his light, sparkling essays in *The Indicator* (1819) and *The Companion* (1828) are perhaps the chief. The strange figure of Thomas De Quincey (*q.v.*) must not be passed over, for he was first and foremost an essayist of a very distinctive kind. Although De Quincey is best remembered by famous pieces such as "On Murder Considered as One of the Fine Arts" or the *Confessions of an English Opium Eater* or "The English Mail-Coach," he was primarily a writer of magazine articles. Practically all his writing appeared originally in some periodical and out of his larger works there are to be collected by the discerning reader some of the most delightful essays in the language. In a sketch, however rapid, of the essay in the 19th century in England, while it is not possible to do more than mention a few of the greater names, the important figure of Thomas Babington Macaulay (*q.v.*) must be given prominence. His vigorous and trenchant essays are known all over the world and are a permanent enrichment to English literature. Macaulay's earliest essay, on Milton, appeared in 1825, not long after the first *Essays of Elia*. No two products cast in the same mold could be more unlike in substance; Lamb, light and charming and autobiographical, and Macaulay, strictly impersonal, full of controversy and replete with history and general knowledge.

The essay as a form of writing has remained popular with English-speaking writers and readers. Great names abound—Matthew Arnold, Walter Pater, Robert Louis Stevenson, Coleridge, Walter Bagehot, Ruskin, Carlyle, Max Beerbohm, E. V. Lucas, Hilaire Belloc, G. K. Chesterton, Virginia Woolf, Alice Meynell, Llewellyn Powys, Robert Lynd (better known as Y. Y., and a master of the modern periodical and newspaper essay), Augustine Birrell, J. B. Priestley—and hundreds more. English literature has reason to be proud of its essayists when a list of this kind, selected almost at random, could be duplicated over and over again. Almost all these essayists are worthy of a much fuller discussion than is possible here; but the essays of Powys, in particular, call for comment, because they are not, perhaps, so widely known as they deserve to be. In the *Dorset Essays* (1935) and *Swiss Essays* (1947) there are passages of beauty, and his bright, enquiring mind played upon all sorts of subjects, to which he brought his great gifts of humour and pathos. His little essay on "The Grave of William Barnes" is an illustration of the master touch that led many to call him a man of genius. Augustine Birrell, too, was a man of letters who possessed a delicious sense of humour and in all his essays he displays it, whatever the nature of the subject. His volume of essays on Carlyle, Milton, Pope, Johnson and Gibbon is filled not only with learning and scholarship and all the fruits of wide reading but with a gay humour.

In the United States the use made of the essay has in general followed the English pattern, although the form has been developed more as a vehicle for literary criticism than in the manner of Montaigne and Lamb. It served admirably as a means of expression for the Transcendental movement, particularly in the work of two great essayists, Ralph Waldo Emerson and Henry David Thoreau (*qq.v.*); the differences in their writings indicate something of the latitude of the form. Emerson's essays, on such subjects as nature, self-reliance and compensation, set forth his philosophy in a style notable for its pithy, stimulating sentences. Thoreau's *Walden* (1854) is a masterpiece of the narrative essay and has had many imitators. The freedom allowed by the essay was happily suited to the talents of another New England writer, Oliver Wendell Holmes (*q.v.*), whose *Autocrat of the Breakfast Table* (1858) has delighted generations of readers. Among other U.S. essayists widely known outside their own country are Edgar Allan Poe (*q.v.*), for his essays on his theory of poetic composition; Washington Irving and James Russell Lowell (*qq.v.*), to name only a few. Among 20th-century essayists special mention may be made of Clarence Day, James Thurber and the notable philosophic essayist George Santayana (*qq.v.*), Agnes Repplier, Christopher Morley and E. B. White.

The history of the form in France has taken a different course, but the essays of C. A. Sainte-Beuve (*q.v.*), known as *Causeries du Lundi*, which began to appear in the *Constitutionnel* in 1849, greatly influenced French literary opinion, and were followed by a brilliant succession of true essay-writers, including Théophile Gautier, Anatole France, Jules Lemaitre, Ferdinand Brunetière

and Émile Faguet (*qq.v.*). Other national literatures also have their essayists. In fact, it may be said that the essay as a literary form is widely used in most countries and is welcomed by a host of readers.

BIBLIOGRAPHY.—H. Walker, *The English Essay and Essayists* (1915); *Essayists Past and Present*, ed. by J. B. Priestley (1925); O. Williams, *The Essay* (1915); *Selected English Essays*, ed. by W. Peacock (1903); *Selected Modern English Essays*, ed. by H. S. Milford (1925), 2nd series (1932); W. H. Prescott, "Essay Writing," *The North American Review*, vol. xiv, pp. 319–350 (April 1882). (B.)

ESSEN, a German industrial town which after partition of the nation following World War II was in the *Land* (state) of North Rhine-Westphalia, Federal Republic of Germany. The largest town in the Ruhr coal field and a port on the Rhine-Herne canal Essen is situated on the Ruhr river, 60 km. (37 mi.) N. of Cologne. It is the seat of a Roman Catholic bishop (from Jan. 1, 1958). Pop. (1959 est.) 727,929. Essen was originally the seat of an aristocratic convent founded in 852 and still represented by the cathedral which stands in the centre of the town. The building was extended in the 11th and 13th centuries and completed in the 15th century. In the suburb of Werden is the abbey church, founded in 796 as a monastic foundation. The convent and the monastery were dissolved in 1803. Essen claimed the standing of an imperial state from the 13th century until 1802, when it was occupied by Prussia. In 1807, under Napoleon, it came into the possession of the grand dukes of Berg, but was restored to Prussia in 1814. The development of the ironworks and coal mines during the 19th century was responsible for the growth of Essen from a small town of about 3,000 inhabitants in 1802 to an important industrial centre. After World War I Essen was taken over by the French who from 1923 to 1925 occupied the Ruhr. More than half the buildings of the town were destroyed during World War II, especially during 1943–45. Essen has large administrative and office buildings, hospitals, schools, an opera house and concert hall, two technical trade schools and many housing estates of which the Margarethenhöhe, laid out like a village within the town, is an example. The Ruhr is dammed to form a lake known as the Baldeney-see; near it are extensive woods and parks and the Villa Hügel, originally the home of the Krupp family and from 1953 used for meetings and cultural events. In Essen also is the Gruga park near the botanical gardens. The principal museum contains the Folkwang collection of paintings and there is also a museum of local history.

Essen is a railway centre. Besides the 22 coal mines, mainly to the north, there are the great Krupp works and a new industrial region, to the west of the old town. Essen has heavy, medium and light industries including locomotive and truck construction, chemical and glassworks and factories for textiles, furniture and instruments. It is a market and retail trade centre. (HT. S.)

ESSENES, a Jewish sect or brotherhood of about the 2nd century B.C. to the end of the 1st century A.D. The Essenes lived in monastic communities in the region of the Dead sea and elsewhere, under strict rule, practising asceticism and supporting themselves by manual labour. The discovery of the Dead sea scrolls has led to much debate about the Essenic character of the Qumran sect. *See* JEWISH SECTS DURING THE SECOND COMMONWEALTH; DEAD SEA SCROLLS.

ESSENTIAL OIL is the name given to any volatile substance of vegetable origin that gives a plant, flower or fruit its distinctive odour or flavour. Essential oils are also known as volatile or ethereal oils, the three terms being used interchangeably. Despite their designation as oils, these complex materials of diversified composition are not related to the nonvolatile glyceride oils of either animal or vegetable origin or to the hydrocarbon oils derived from petroleum.

Uses.—Essential oils are used for a wide and varied group of purposes in many industries for the scenting and flavouring of numerous products, such as perfumes, cosmetics and soaps. Some essential oils are more or less powerful external or internal antiseptics, some possess an analgesic, hemolytic or antienzymic action, and some act as sedatives, stimulants and stomachics. The anthelmintic properties of certain oils, especially oil of chenopodium (American wormseed), are made use of in medicine. The essential oils of spices have been used as flavouring materials since time immemorial. Naturally occurring oils, their isolates and synthetic essential oils are also used to mask objectionable odours in working areas and in industrial products such as artificial leather, paint and rubber.

Occurrence.—Essential oils are found in special cells, glands or ducts that are located sometimes in several parts of a plant—leaves, barks, roots, flowers and fruits—and sometimes in just one or two parts. As a rule the oils are present in very small amounts and comprise only a tiny fraction of the entire plant material—oil (attar) of rose comprises about 0.2% of rose petals. The quantity and composition of an oil vary with the type of plant and also with growing conditions; *e.g.*, light, climate, soil and altitude. Essential oils occur in about 60 plant families, such as the laurel, myrtle and mint, but only about 150 plant species are grown commercially for the production of essential oil.

The function of the essential oil in the living plant is not completely understood. Odours of flowers may attract or repel certain insects, thereby aiding natural selection. Root oils, wood oils and leaf oils may provide protection against plant parasites or against animal depredation. Oleoresinous exudations formed on the trunk of a wounded tree appear to act as protective seals against disease, parasites and loss of sap. In any case, the oils are produced during some metabolic process of the plant and are secreted or excreted as odoriferous by-products.

The fragrant oils are not necessarily present as such in the living plant but may occur as odourless compounds termed glycosides; *e.g.*, amygdalin in bitter almond. When the plant tissues are macerated, an enzymatic reaction occurs and causes the glycoside to undergo a chemical change. This action, in turn, liberates the distinctive essential oil.

Production.—Essential oil production involves extremes of primitive culture and modern processing methods. In most instances aromatic plants grow wild or are cultivated by primitive methods in small quantities as garden or patch crops by growers who produce only one oil. Skill and long experience are important in cultivation because the odoriferous crop must be gathered at the proper stage of development and processed properly. If an error in timing or processing occurs, the essential oil may have an unsatisfactory odour or flavour. The oil is bought by field brokers, then by village buyers and finally by the exporters in the shipping ports. Only a few oils are produced by methods based upon modern principles of plant breeding, mechanized agriculture, engineering and mass production that involve concentration of effort, specialized and highly skilled labour, heavy capital investment and accessibility of a variety of year-round crops. In this group are the flower oils grown in the Grasse region of southern France and the citrus oils produced in California and Florida. The essential oil industry, however, is no longer limited to producing and distributing naturally derived essential oils; it now is also concerned with the development, production and testing of synthetic essential oils and mixtures. In many cases the synthetic products are the same chemically as the natural essential oil.

Essential oils are obtained from plants by three methods:

1. Steam distillation, the most important method, is applicable to the widest range of substances; *e.g.*, peppermint, camphor and cloves. It is favoured especially for the production of the more stable oils such as oils of cedar and of clove.

2. Expression, by hand (*écuelle* and *éponge*) or by machinery, is used almost entirely for the production of oils contained in the rinds of fruits such as lemon and bergamot.

3. Extraction is carried out with volatile solvents such as petroleum ether and alcohol; by cold neutral fats (enfleurage); or by hot oils or fats (maceration).

Properties.—Essential oils are almost completely insoluble in water but are highly soluble in alcohol, ether, fatty oils and mineral oils. As a rule they are not oily to the touch and leave no permanent grease spot. They are usually liquid at ordinary temperatures; some, such as oil of camphor and rose oil, deposit, on standing, solid matter (stearoptenes) that crystallizes from the liquid remainder (elaeoptenes). The natural essential oils consist usually of mixtures of several chemical compounds. The most

widely studied essential oil is the oil distilled from the wood of the camphor tree. Exceptionally large amounts are commercially fractionated, making it possible to isolate components present in minute amounts. More than 75 individual compounds have been identified in this oil. (*See* CAMPHOR.)

Classification of Constituents.—The components of essential oils are, as a rule, compounds of hydrogen, carbon and oxygen only. However, hydrocarbons (hydrogen and carbon only) and compounds containing nitrogen and sulfur are also found. The constituents of essential oils may be classified according to their chemical nature into the following seven groups:

1. Hydrocarbons, such as paraffin hydrocarbons (alkanes), olefins (alkenes), aromatic (*i.e.*, cyclic rather than chain structure) hydrocarbons and terpenes. Of these, the terpenes and sesquiterpenes are most prominent and characteristic of essential oils. The fundamental building block of a terpene is the isoprene unit, C_5H_8. Two isoprene units form a terpene ($C_{10}H_{16}$), three isoprene units comprise a sesquiterpene and four isoprenes make up a diterpene; examples of terpenes are pinene in oil of turpentine and limonene in lemon and orange oils. (*See* TERPENES.)

2. Alcohols, such as (a) straight chain varieties, *e.g.*, linalool in oils of lavender, orange flowers and linaloe; geraniol in oil of rose and citronella; (b) terpene alcohols, *e.g.*, borneol in camphor and terpineol in oil of linaloe; and (c) phenols, *e.g.*, menthol in oil of peppermint, thymol in thyme oil and eugenol in oil of clove.

3. Esters, *e.g.*, isoamyl acetate in synthetic pear flavour, methyl salicylate in oil of wintergreen, and geranyl acetate (main constituent of geraniol) in oils of palmarosa and lemon grass. Acids occasionally occur in the free state.

4. Aldehydes, *e.g.*, citral in lemon oil, citronellal in oil of lemon, benzaldehyde in oil of bitter almond and vanillin in vanilla.

5. Ketones, *e.g.*, menthone in oil of peppermint, camphor in oil of camphor and ionone synthesized from citral for violet odours.

6. Oxides and lactones, *e.g.*, safrole in oils of sassafras and camphor, coumarin in tonka bean or synthesized from salicylaldehyde for "new-mown hay" odours.

7. Nitrogen and sulfur compounds, *e.g.*, hydrocyanic (prussic) acid in oil of bitter almond, nitrobenzene (oil of mirbane) synthesized from benzene for almond odour, allyl isothiocyanate in oil of mustard and organic disulfides in garlic and onion oils.

Many other types of compounds exist in essential oils and most are found in a variety of natural oils.

Terpenes and other components bearing isoprene units oxidize and polymerize on standing and thereby produce undesirable characteristics. Elimination of the terpenic portion usually results in better alcohol solubility, better stability against oxidation and resin formation, and greater strength of odour and flavour. The terpene hydrocarbons are usually removed from several natural oils by fractional distillation under vacuum; natural oils that have been treated in this way are known as terpeneless oils. Members of another variety, known as concentrated oils, are usually prepared by distilling off the lower boiling part of the terpene fraction.

The so-called synthetic oils may be grouped into two classes: (1) isolates, or artificial oils, each of which is a more or less pure chemical isolated from a natural essential oil; *e.g.*, menthol and eucalyptol, isolated by freezing from oils of peppermint and eucalyptus respectively; (2) synthetics, oils synthesized by chemical means from isolates or other raw materials such as coal tar derivatives (*e.g.*, methyl salicylate from salicylic acid). The synthetic essences may be identical with the odoriferous bodies present in a natural oil; *e.g.*, vanillin synthesized from eugenol in oil of clove or from guaiacol is the same as the vanillin that occurs in vanilla. Other synthetic oils, however, have no counterparts among natural essential oils. Such synthetics include ionone, which is derived from citral isolated from oil of lemon grass and is the basis of all violet scents and perhaps the most important of all synthetic perfumes; and nitrobenzene, which possesses a coarse almond odour although it is chemically distinct from the benzaldehyde of natural almond oil.

It is not possible to make a strict correlation of odour and chemical constitution. The hydrocarbons have, on the whole, no pronounced odour; scent appears when oxygen, nitrogen or sulfur in certain groups such as hydroxyl (—OH), aldehyde (—CHO), nitrile (—CN), isothiocyanate (—SCN), etc., are added to the molecule, but particular groupings cannot be associated with particular types of odour.

History.—The history, art and science of essential oils and perfumes are inextricably related. (*See* PERFUME.) The development of the essential oil industry is based upon and paralleled the art of distillation, which was developed in Egypt, Persia and India. Early data on essential oils are vague, but it is known that extensive trade in odoriferous oils and ointments was carried on in the ancient countries of the Orient and in ancient Greece and Rome. The Greek historian Herodotus (5th century B.C.), the Roman Pliny the Elder (A.D. 23 or 24–79) and his contemporary Pedanius Dioscorides (author of the *Materia Medica* that dominated medicine for 1,500 years) mention oil of turpentine but describe no other volatile oil. The first authentic description of the distillation of essential oils is generally ascribed to the Catalan physician Arnaldo de Vilanova (1240–1311) who, by mentioning products of distillation other than oil of turpentine, may be said to have introduced the art of distillation to European practitioners.

Production and use of essential oil apparently did not become general until the latter half of the 16th century. *Liber de arte Distillandi de simplicubus*, a book on distillation by Hieronymus Brunschwygk (*c.* 1450–1513), covered the entire field of distillation techniques and products then known. It mentions only four essential oils—those of turpentine, juniper wood, rosemary and spike lavender. The systematic study of essential oils may be said to have started with the beginnings of modern organic chemistry, particularly with J. B. A. Dumas (1800–84). At first, attention was focused on their use in medicinal drugs but eventually their potentialities for perfumes, beverages and foodstuffs assumed far greater importance.

See also SCENTED PLANTS; FLAVOURINGS.

BIBLIOGRAPHY.—E. S. Guenther, *The Essential Oils*, vol. i–vii (1948–1952); J. L. Simonsen, *The Terpenes*, vol. i–iii, 2nd ed. (1947–1952); P. Z. Bedoukian, *Perfumery Synthetics and Isolates* (1951). (P. HE.)

ESSEQUIBO (ESSEQUEBO), the largest river between the Amazon and the Orinoco, South America, rises in the Akaraí mountains (Serra Acaraí) on the Brazilian border and flows about 630 mi. through the savanna and forest of British Guiana to the Atlantic. Its estuary, 20 mi. wide, is obstructed by islands and silt. The chief tributaries are the Rupununi, Mazaruni and Cuyuni, all left bank, and the system drains more than half of British Guiana. The Essequibo is navigable by small ocean vessels to Bartica (50 mi.), and by boats for long reaches between rapids.

The old Dutch colony of Essequibo was settled from 1616, with the seat of government first at Kyk-over-al Island, near Bartica, and later on at Fort Island. The colony lost its predominance when the population moved to the coast and the superiority of the Demerara estuary as a port was felt. The modern county, or region, of Essequibo comprises three-quarters of British Guiana. The estuarine islands are Hog, Fort, Leguan and Wakenaam.

(G. LN.)

ESSEX, EARLS OF. The title of earl of Essex was first held by Geoffrey de Mandeville (*q.v.*), to whom Stephen granted it in 1140, and afterward by his childless sons, Geoffrey (d. 1166) and William (d. 1189). In 1199 the earldom was successfully claimed by King John's justiciar, Geoffrey FitzPeter (d. 1213), whose wife, Beatrice, was descended from a sister of the 1st earl. When his sons Geoffrey (d. 1216) and William (d. 1230), both known by the name of "de Mandeville," died without issue, the title passed in 1236, through their sister Matilda, to her son Humphrey de Bohun, earl of Hereford, and remained in the family of Bohun (*q.v.*) until their male line became extinct in 1373. The earldom of Essex was one of the dignities of Thomas of Woodstock, duke of Gloucester, who married Eleanor, the elder Bohun coheiress, and it then passed to his grandson Henry Bourchier (d. 1483), Edward IV's treasurer, who was created earl of Essex in 1461. The second and last Bourchier earl was Henry's grandson Henry, who died early in 1540. A few weeks before his execution in 1540 Thomas Cromwell was created earl of Essex; then, in 1543, William Parr, afterward marquess of Northampton, obtained the earldom by right of his wife Anne, daughter of the last Bourchier earl. Northampton lost the earldom when he was attainted in 1553; and afterward it passed to the famous family of Devereux, Walter Devereux, who was created earl of Essex in 1572, being

related to the Bourchiers. Robert, the 3rd and last Devereux earl, died in 1646.

In 1661 Arthur Capel was created earl of Essex, and the earldom is still held by his descendants. (C. D. R.)

ESSEX, ARTHUR CAPEL, 1ST EARL OF (1632–1683), English statesman who distinguished himself by the rectitude and moderation of his political and administrative career under Charles II, was the son of Arthur Capel, 1st Baron Capel. His youth was darkened by the Civil War and by the capture and execution of his father in 1649. Earlier, in June 1648, although ill, Arthur had been taken to the siege of Colchester by soldiers of Lord Fairfax and paraded round the defenses in the hope of inducing Lord Capel to surrender the town.

At the Restoration Capel was created Viscount Malden and earl of Essex (April 20, 1661), but he showed little desire to enter fully into court life and his antagonism led Charles II to couple him with Lord Holles as "stiff and sullen men." His public career began with an embassy to Denmark in 1669 which was so successful that upon his return he was promised any preferment he might desire. He was appointed lord lieutenant of Ireland in 1672 and he displayed great integrity during his five-year tenure, purging the financial administration and repelling further attempts to exploit the Irish land market. His unaccommodating honesty brought him much enmity at court and he was recalled in 1677. He had not forfeited the king's favour, however, and in 1679 he became a lord of the treasury. With the earls of Halifax and Sunderland he formed the "Triumvirate" which dominated policy for a short period in the precarious conditions created by the "popish plot." At first he supported the scheme of Halifax for imposing statutory "limitations" in the event of the succession of the Roman Catholic duke of York (afterward James II), but disgust with the court led him to resign the treasury in Nov. 1679 and join the earl of Shaftesbury in voting for the complete exclusion of James in Nov. 1680. After the defeat of the exclusion bill he carried a motion for an association to execute the scheme of expedients proposed by Halifax. In Jan. 1681 he led 15 peers in petitioning the king to abandon his intention of summoning a parliament at Oxford.

Capel's changed attitude to political life also showed itself in his zealous prosecution of Roman Catholics involved in the "popish plot." However he interceded for the life of Oliver Plunkett, primate of Ireland, whom he had known and respected. He was involved in Shaftesbury's incitements of the "Whig" London mob, but he protested at his scheme to seize the Tower of London in 1682. After Shaftesbury's flight his continued association with the duke of Monmouth's supporters, who included Lord Russell, Algernon Sidney and Lord Howard of Escrick, was highly dangerous. He seems to have taken no active part with them in planning the Rye House plot of June 1683, but after the capture of the leaders he was sent for and imprisoned in the Tower. There, in the apartments from which his father had gone to execution, he showed signs of depression. On July 13 he was found with his throat cut. The remarkable violence with which the wound had been inflicted encouraged political pamphleteers to dispute the probability that it was suicide. The evidence is not conclusive but it is possible that he desired to prevent an attainder and preserve his estate for his family. His death was much deplored and Charles is alleged to have intended to pardon him in requital for his father's sacrifice.

Essex did not leave a strong mark upon the political life of his age. He was unambitious, but his competence as an administrator and the purity of his public and private life makes him an outstanding figure. John Evelyn describes him as "a sober, wise, judicious and pondering person . . . very well versed in English history and affairs, industrious, frugal, methodical, and every way accomplished." (H. G. Ro.)

ESSEX, ROBERT DEVEREUX, 2ND EARL OF (1566?–1601), English soldier and courtier, famous for his relationship with Elizabeth I, was born at Netherwood, Herefordshire, on Nov. 10, 1566 or 1567, and educated at Trinity college, Cambridge. A delicate, precocious boy, with an inheritance impoverished by the Irish adventures of his father (*see* ESSEX, WALTER DEVEREUX), he lived quietly in the country until he accompanied his stepfather, the earl of Leicester, to the Netherlands in 1585–86. There he fought bravely at the Zutphen skirmish. On his return he was made master of the horse in 1587 and in 1588 a knight of the Garter and general of the horse in the army gathered at Tilbury against the Armada. He became a prime favourite of Elizabeth I and received numerous grants, among them the lucrative sweet wine farm, but he soon found court life irksome. His irritation found vent in quarrels with his rivals Sir Walter Raleigh and Sir Christopher Blount and in 1589 he ran away to join the expedition to Portugal under Sir Francis Drake and Sir John Norris. Peremptorily recalled by the queen, he was soon high in her favour again, though in 1590 he incurred her anger by his secret marriage to Frances Walsingham, widow of Sir Philip Sidney. In 1591 the queen reluctantly allowed him to indulge his taste for soldiering as commander of a force sent to help Henry IV of France. In this enterprise he displayed more recklessness and knight errantry than generalship, though the failure to attain the objective by capturing Rouen from the Catholic league was less his fault than Henry's.

The next few years offered Essex few opportunities for soldiering and for a time his energies found other outlets. With the help of Anthony Bacon he built up a network of foreign correspondences which rivaled that of the Cecils to whom, after Sir Francis Walsingham's death in 1590, the queen entrusted most of the work (though not the title) of the principal secretary. In 1593 he was made privy councilor and in 1594 discovered the alleged plot of the Portuguese exile, Roderigo Lopez, to poison the queen. In these fields, however, his greatness was circumscribed by the long-established ascendancy of William Cecil, Lord Burghley, and the growing influence of Burghley's son Robert Cecil. Nor was the queen prepared to yield him a monopoly of her favour, as she showed by her rejection of his insistent demands (1593–94) for the appointment of his protégé Francis Bacon as attorney general or solicitor general. Then in 1596 the revival of offensive operations against Spain opened new opportunities for military adventure and in that summer Essex was one of the commanders of the great expedition which seized and sacked Cádiz. This spectacular but indecisive success did not greatly impress the queen, but it increased the earl's popularity, especially among military men. He became the leading advocate of a more vigorous strategy. He worked out a plan for cutting off Spain's supplies of treasure from the new world and of naval stores from the Baltic by seizing and holding a Spanish port as a base for an English blockading fleet. Moreover, the threat of a new Spanish armada in the autumn gave him the chance to play a leading part in organizing England's military defenses and in March 1597 he was appointed master of the ordnance. On his return from Cádiz Francis Bacon had warned him against the danger of alienating the queen by pressing his ambition too far, by seeking popularity and relying upon a "military dependency"; instead, he should seek a more limited and "domestical greatness." But it was not easy to turn back and in 1597, although his more ambitious plan was abandoned, Essex shared with Raleigh the command of the Islands voyage, which narrowly failed to intercept the Spanish treasure ships at the Azores. Next year the possibility of peace with Spain sharpened his rivalry with the Cecils, while the growing seriousness of Tyrone's rebellion in Ireland led to fresh differences about appointments and strategy. During one of these disputes Essex turned his back upon the queen, who promptly slapped his face.

By now, indeed, the queen, as Francis Bacon had foretold, was finding Essex of "a nature not to be ruled" and was growing alarmed by his importunate ambition. Nevertheless, in 1599 he was sent to Ireland as lord lieutenant. His campaign was unsuccessful and he laid himself open to serious charges by opposing the commands of the queen and council, concluding a truce with Tyrone and suddenly deserting his post and coming over to vindicate himself privately to the queen. In June 1600 he was deprived of his offices and ordered to live as a prisoner in his own house. His liberty was soon restored but he was still forbidden the court and in September the queen refused to renew his farm of the sweet

wines, a mainstay of his finances. In desperation he toyed, as he had done in Ireland, with schemes to seize control of the court and regain his ascendancy by force, and on Feb. 8, 1601, he and 300 of his followers paraded through London shouting "For the queen! a plot is laid for my life." Finding no support from the city, he returned to Essex house and soon afterward surrendered. After a trial at which Francis Bacon was one of his prosecutors, he was condemned to death and executed at the Tower of London on Feb. 25, 1601.

Bibliography.—W. B. Devereux, *Lives and Letters of the Devereux, Earls of Essex, 1540–1646*, 2 vol. (1853); E. A. Abbott, *Bacon and Essex* (1877); L. H. Cadwallader, *The Career of the Earl of Essex From . . . 1597 to . . . 1601* (1923); G. B. Harrison, *Life and Death of Robert Devereux, Earl of Essex* (1937). (R. B. Wm.)

ESSEX, ROBERT DEVEREUX, 3rd Earl of (1591–1646), English soldier, who commanded the parliamentarian army at the beginning of the Civil War, was born in London in Jan. 1591, the son of Robert Devereux, 2nd earl of Essex. He was restored to his family titles and estates in 1604 and was taken under the care of James I, who arranged his marriage to Frances Howard, daughter of the earl of Suffolk, on Jan. 5, 1606. Essex, a shy, awkward and inexperienced youth, was then sent on a foreign tour and returned to find that his wife had fallen in love with the king's Scottish favourite, Robert Carr. James appointed a divorce commission which annulled the marriage (1613), not on the ground that Essex was impotent, but that he was incapable of cohabiting with his wife. His second marriage, in 1631, with Elizabeth, daughter of Sir William Paulet, was no happier.

Essex first learned the art of war when he commanded a volunteer regiment under Sir Horace Vere in 1620. He was vice-admiral in the unsuccessful expedition against Cádiz in 1625. In 1626, like Sir John Eliot, another vice-admiral, he refused to contribute to a forced loan. Charles I appointed him second-in-command in the first campaign of the Bishops' Wars (1639) and made him a privy councilor in 1641, but Essex advised the king to summon parliament in 1640, and later favoured the execution of the king's chief minister, the earl of Strafford.

Essex was trusted by the leaders of the house of commons and was appointed to command the parliamentarian army in July 1642. After his virtual defeat at Edgehill (Oct. 1642) he ordered his officers to train their men in the "necessary rudiments of war" rather than "ceremonious forms of military discipline." In 1643 Essex advanced on Windsor, captured Reading, relieved Gloucester and fought the first battle of Newbury. He attempted to resign his commission but this was refused. The next year he planned to assault Oxford, but, after the king escaped, went to the relief of Lyme in Dorset, when he allowed his army to outrun its communications, be cut off and surrounded (Aug. 1644) in Cornwall. He himself escaped and was then taken ill. He disliked the formation of armies other than his own, quarreled with other generals, and opposed the establishment of the "New Model army" in 1645. He resigned his command the day before the Self-Denying ordinance, providing that no members of parliament should hold public office, was finally passed (April 1645). After further illness Essex died unexpectedly in London on Sept. 14, 1646, and was buried in Westminster abbey. The earldom then became extinct. Essex had a retiring disposition, lethargic habits, and a profound sense of public duty. He proved himself a not incompetent soldier, but he was always highly sensitive and subject to conflicting loyalties and emotions. *See* also Civil War, English.

Bibliography.—Robert Codrington, "The Life and Death of Robert, Earl of Essex" in *Harleian Miscellany*, vol. i (1808); W. B. Devereux, *Lives and Letters of the Devereux, Earls of Essex*, vol. ii (1833); C. V. Wedgwood, *The King's War, 1641–1647* (1958). For an account of the divorce, *see* William McElwee, *The Murder of Sir Thomas Overbury* (1952). (M.P.A.)

ESSEX, WALTER DEVEREUX, 1st Earl of (1539–1576), English soldier who led an unsuccessful and sanguinary colonizing expedition to Ulster in 1573–75, was born at Carmarthen castle on Sept. 16, 1539. He was the eldest son of Sir Richard Devereux (d. 1547) and the grandson of Walter Devereux, 3rd Baron Ferrers and Viscount Hereford (created 1550), to whose titles he succeeded in 1558. His grandfather was a nephew, through his mother, of Henry Bourchier (d. 1539), the last of the earls of Essex in that line. Walter married, between 1560 and 1565, Lettice, daughter of Sir Francis Knollys. He served as "high marshal of the field" under the earl of Warwick and Lord Clinton in the suppression of the northern insurrection in 1569, and in 1572 he received the Garter and was created earl of Essex. In 1573 Essex offered on certain conditions to subdue and colonize, at his own expense, a portion of the Irish province of Ulster, at that time completely under the dominion of the rebel O'Neills led by Sir Brian MacPhelim and Tirlogh Luineach, supported by the Scots on the north coast of Antrim under Sorley Boy MacDonnell. He left for Ireland in July 1573, but a storm dispersed his fleet and his forces did not assemble till late in the autumn. His troops were diminished in winter quarters at Belfast to little more than 200 men. Intrigues of various sorts and fighting of a guerrilla type followed, and Essex had difficulties both with the lord deputy, Sir William Fitzwilliam, and with the queen. His movements in Ulster took the form of raids and brutal massacres among the O'Neills; he treacherously captured MacPhelim at a conference in Belfast in Oct. 1574 and, after slaughtering his attendants, had him and his wife and brother executed at Dublin.

Elizabeth, instigated apparently by the earl of Leicester, suddenly commanded Essex to "break off his enterprise"; but, as she left him a certain discretionary power, he took advantage of it to defeat Tirlogh Luineach, chastise Antrim and massacre several hundred of Sorley Boy's following, chiefly women and children, discovered hiding in the caves at Rathlin. He returned to England in the end of 1575. He went to Ireland again the next year, but he died of dysentery on Sept. 22, 1576, soon after his arrival in Dublin. After his death the earl of Leicester married his widow.

See W. B. Devereux, *Lives and Letters of the Devereux, Earls of Essex, 1540–1646*, 2 vol. (1853). (R. B. Wm.)

ESSEX, an eastern county of England, bounded north by Cambridgeshire and Suffolk, east by the North sea, south by the Thames and west by London, Middlesex and Hertfordshire. Under the London Government Act 1963 the southwestern corner of Essex (Chingford, Walthamstow, Leyton, a small part of Chigwell, Wanstead and Woodford, West Ham, East Ham, Ilford, Barking, Dagenham, Romford, Hornchurch) was to become part of Greater London on April 1, 1965, and on the same date Middlesex would cease to exist (*see* London). The present article, however, relates to Essex before implementation of the act. A mainly rural county, it has several seaside resorts—Southend-on-Sea, Burnham-on-Crouch, Clacton, Frinton, Walton-on-the-Naze.

Physical Features.—Although inland the surface is undulating, there are no hills of any great altitude (Danbury hill, Laindon hill and High Beech are all below 400 ft.) and the county is generally flat. The chalk highlands of the northwest slope gradually south and east toward the low alluvial coastline, and, following this configuration, the rivers flow outward from the northwest; the Lea and Roding into the Thames, and the Crouch, Blackwater and Stour into the North sea. The Lea forms part of the western and the Stour part of the northern boundary.

Geologically, Essex forms part of the London basin, a series of clay and drift strata contained in a chalk hollow, the southern rim of which rises in Kent and sinks below ground in most of south Essex. It is near the surface at Grays and Purfleet, where it is used in the manufacture of cement. The northern rim of the basin emerges in the chalk highlands around Saffron Walden. Overlying the chalk is a series of clay strata, the most important being the London clay, which covers a large part of the surface of the county; in the southeast at Southend-on-Sea, Walton-on-the-Naze and Harwich, and in the west at Hainault and Epping forest. The attractive fossiliferous Red Crag is exposed at Walton-on-the-Naze. Glacial drift, providing a variety of rich agricultural soils, also covers much of the surface, and there are remarkable instances of glacial erosion, notably in the Cam valley, where a deep gorge is filled with glacial drift, which soundings at Newport have revealed to be 340 ft. thick.

The coast is marshy and alluvial, with many inlets and coastal islands, such as Canvey, Foulness and Mersea. Much marshland has been reclaimed, and dykes and embankments are necessary to protect it from the sea. The sea level in the south and east has

gradually risen in historic times, and disastrous floods like the one in 1953, occur occasionally. In the north, the low cliffs in the Frinton and Walton area are being eroded away by the sea.

Inland, the soil is remarkably varied, and many kinds of agriculture are found, from heavy crops on the rich alluvial soil of the southeast to pasturage on the chalk highlands of the northwest. The clay soil of the Epping area supports Epping and Hainault forests, the remains of the great Waltham forest which in medieval times covered much of southwest Essex. Epping forest was preserved for the public by an Act of 1878, after the London County council had fought on its behalf in a notable controversy, and has since become a noted observation field for naturalists.

History.—There are few pre-Roman remains in Essex. Gryme's Dyke, near Colchester, is thought to have protected the important British town on that site. Loughton camp and Ambresbury banks, both in Epping forest, are also thought to be pre-Roman, and remains of lake-dwellings found near Braintree may be of the Bronze Age.

In Roman times, Colchester became one of the few *coloniae*, or municipalities, in Britain. It has many Roman remains, including the Balkerne Gate, and much of the town wall. There were also Romano-British towns at Great Chesterford, Chelmsford and Rivenhall, and many villa sites have been identified. The Bartlow hills at Ashdon are Roman burial-mounds, and there are many "red hills" (saltworks) along the coast. The Saxon invaders of the 5th century gave their name to Essex (the land of the East Saxons), and their settlements, indicated by place-names ending in *-ing*, *-ham* and *-ton*. Saxon remains have been found at, among other places, Rainham and Prittlewell, where they are continuous with Roman remains. Some village names, such as Thorpe-le-Soken and Kirby-le-Soken, are of Danish origin, the Danes having won the battle of Maldon in 991 (*see* MALDON). Norman landowners also added their surnames to many village names, such as Tolleshunt d'Arcy and Beauchamp Roding.

After the Norman Conquest the chief landowners in Essex were, in addition to the king, the bishop of London, the canons of St. Paul's, the bishop of Bayeux, Eustace II count of Boulogne, Geoffrey de Mandeville (1st earl of Essex) and Robert Gernon. Prominent among those who acquired land in Essex after the Dissolution were Thomas Lord Audley, Richard Lord Rich, and Sir William Petre. From the 16th to the 19th centuries Essex, especially the southwest of the county, became a fashionable residential area for the gentry.

Essex was closely associated with many great events in English political history, partly because of its proximity to London. During the reign of Stephen, Geoffrey de Mandeville extorted privileges from both the king and the empress Matilda, and wielded unfettered power over his possessions in Essex and the eastern counties. Among the barons foremost in the struggle for Magna Carta were Robert FitzWalter of Dunmow; Robert de Vere, 17th earl of Oxford, of Castle Hedingham; Richard de Muntfichet of Stansted; John FitzRobert of Clavering castle; and William Lanvallei of Colchester. Essex men, including the former Colchester priest John Ball, were prominent in the Peasants' Revolt of 1381. It was at Stratford in Essex that in 1397 Richard II treacherously caused his uncle, the duke of Gloucester, to be kidnapped and hurried off to be murdered at Calais. During the Wars of the Roses the de Veres of Hedingham supported the Lancastrians, while the Bourchiers and Lord FitzWalter of Woodham Walter were Yorkists. In the Civil War Essex, with the exception of a very few influential families, such as the Lucas family of Colchester, supported parliament, but the royalists took and held Colchester during a long siege in 1648.

Essex is mainly an agricultural county, having in its favour a variety of fertile soil and a mild climate. In 1086 much of the county was covered with forest. Although through clearance the only large area of woodland now remaining is Epping forest, the county is still fairly well wooded. The principal tree is the oak: the famous Fairlop oak at Hainault, which was blown down in 1820, is said to have measured 45 ft. in girth. Land has also been reclaimed from the marshes in the southeast, and the rich

A. F. KERSTING

COLCHESTER CASTLE, ESSEX: 11TH CENTURY

alluvial soil now produces heavy yields of crops. In many parishes enclosure took place early in the middle ages, though in some, open-field farming continued until the 19th century. In the early enclosed south and east the pattern of settlement consists of isolated farms and hamlets, but in the north and west, where enclosure came later, nucleated villages are more common. Agriculture continued to flourish until the depression of the late 19th century, which had disastrous economic effects, especially in the southeast.

Essex is poor in building stone but rich in timber, a fact reflected in its architecture. The Romans used a particularly hard form of cement mixed with rubble to build the wall at Colchester, and manufactured quantities of bricks and building tiles. Timber has always been the chief domestic building material and many fine examples of timber-framed houses of the 13th to the 15th centuries still survive, particularly in the rural north and east. Many are plastered and colour-washed; some, like Paycockes at Great Coggeshall, are paneled with brick-nogging. In the 16th century brick was used for the building of large houses such as Audley End, Hill hall in Theydon Mount, and Ingatestone hall. Faulkbourne hall is a brick house of the 15th century. Two Norman castles survive: Hedingham castle, built of Barnack stone, and Colchester, said to be the largest keep in existence and built by the Normans of Roman material. Moated mounds at Pleshey and Ongar mark the sites of castles, and many old manor and farm houses are surrounded by moats.

The oldest church in Essex, St. Peter's-on-the-wall at Bradwell, was built by Bishop Cedd in about 654 from stone from the nearby Roman fort. Many Saxon churches were built of timber: the timber nave at Greensted, near Ongar, is the only surviving example in England. After the Norman Conquest many churches were built of flint and rubble; some incorporate Roman brick. They are unimposing in comparison with the churches of stone-building areas in the west country, but the enforced use of timber has produced some fine timber towers, notably Blackmore and Navestock. Some churches along the Thames estuary, such as Hornchurch, are built of Kentish ragstone. Dedham, Saffron Walden and Thaxted have fine 15th-century churches. At Little Maplestead is one of the five circular churches in England, built by the Knights Hospitallers in about 1350. From the 15th century onward a few churches were built of brick: *e.g.*, Ingatestone.

For most of its history Essex was in the diocese of London. In 1836 it was transferred to that of Rochester, with the exception of nine parishes in the southwest which remained in London. In 1863 these nine parishes were also transferred to Rochester. In 1875 the whole county was transferred to the new diocese of St. Albans, and in 1914 it became, by itself, the diocese of Chelmsford. There were a number of peculiars, or exceptional jurisdictions, belonging to the archbishop of Canterbury, the canons of St. Paul's, New college, Oxford, Waltham abbey and Westminster abbey. The parishes in Essex are manorial in origin, churches

being built by the lords of manors for their own use. Hence many churches are situated near the manor house rather than near the village, and at Willingale Spain and Willingale Doe, two churches, originating from two manors, stand in the same churchyard.

The areas around the Thames estuary sustained considerable air raid damage during World War II.

Population and Administration.—The population of the geographical county in 1961 was 2,288,058 and that of the administrative county 1,859,916, of which the rural districts accounted for only 254,676. Much of the population is concentrated in the southwest, the two county boroughs of East Ham and West Ham having a population of 105,682 and 157,367 respectively. Ilford, Romford, Wanstead and Woodford and Chigwell are other towns in the southwest with large and growing populations. Southend-on-sea, the only other county borough, has a population of 165,093, Colchester of 65,080 and Chelmsford 48,104. London county council housing estates have been built at Loughton and Harold Hill, which also has an industrial estate, and complete New Towns at Harlow and Basildon. Because of good communications, southwest Essex is becoming a residential area for workers in London.

Essex is divided for parliamentary purposes into ten county divisions: Epping, Chigwell, Chelmsford, Saffron Walden, Maldon, Colchester, Harwich South East, and Billericay. The following parliamentary boroughs also return one member in each constituency or division: West Ham, East Ham, Southend-on-Sea (two members each), Leyton, Barking, Romford, Hornchurch, Woodford, Dagenham (one member each).

Essex is in the southeastern circuit and assizes are held at the Shire hall in Chelmsford which is the county town. It forms one court of quarter sessions and is divided into 21 petty sessional districts. Colchester, Southend and West Ham have separate courts of quarter sessions. The southwest of the county falls within the area of the metropolitan police district, which includes the municipal boroughs of Barking, Chingford, Dagenham, Ilford, Leyton, Walthamstow and Wanstead and Woodford, together with the urban districts of Chigwell and Waltham Holy Cross. The county is in the diocese of Chelmsford. There is a garrison at Colchester. The seaward limit of the Port of London authority on the Thames is a line from Havengore creek, 3 mi. N.E. of Shoeburyness, to Warden point on the Kent coast.

Industries and Communications.—Essex has always been a predominantly agricultural county. John Norden, in 1594, called it the "English Goshen" and remarked on its fruitfulness and the variety of its produce. Wheat and barley are the chief crops. Sugar beet is also grown, and potatoes and other vegetables for the London market. There is some dairy farming in the Epping and Ongar area, where the London Co-operative society owns large farms, but pasturage is not common in Essex because of the comparative dryness of the climate. There is seed growing in the Kelvedon area. Industry is mainly in the southwest of the county, near the river and London. There are food and chemical works, gunpowder works near Waltham Abbey, and engine shops at Stratford. Automobiles are made at Dagenham, photographic materials at Ilford, and there is a factory at Debden (Loughton) for Bank of England notes. There are busy docks on the Thames at Tilbury and East and West Ham (Silvertown and Custom House), and the East London waterworks by the Lea near Walthamstow. Apart from this area, there is industrial activity in other parts of Essex. Colchester and Chelmsford have engineering works, there are silkworks at Braintree and Bocking, large breweries at Brentwood, Chelmsford and Romford, and lime and cement works at Grays. There are nuclear power stations near Chelmsford and at Bradwell-on-Sea, and an atomic weapons research station on Foulness island. In spite of its long indented coastline, Essex is not important as a maritime county, because of the absence of deep-water anchorages. The Colne oyster beds produce the famous Colchester natives, and there are similar beds in the Crouch and Roach.

London Underground lines run east via East Ham to Upminster, and northeast via Stratford to Woodford, Hainault, Loughton, Epping and Ongar. The other rail services are run by British railways. The main line from London to Norwich runs northeast across the county through Stratford, Shenfield, Chelmsford, Kelvedon, Colchester and Manningtree. There are electric services as far as Chelmsford.

BIBLIOGRAPHY.—G. A. Ward (formerly Thornton), *Essex Local History, a Short Guide to Books and Manuscripts* (1950); J. Norden, *Speculi Britanniae pars; an Historical and Chorographical Description of the County of Essex . . .* [1594], ed. by Sir H. Ellis (1840); P. Morant, *The History and Antiquities of the County of Essex . . .*, 2 vol. (1768); *A New and Complete History of Essex by a Gentleman*, 6 vol. (1770–72); T. Wright, *The History and Topography of the County of Essex*, 2 vol. (1831–85); *Victoria County History of Essex* (1903–); P. H. Reaney, *The Place-names of Essex* (1935); publications of the Essex Record Office, Chelmsford, notably *A Reproduction of a Map of the County of Essex*, 1777, by J. Chapman and P. André (1950); *Catalogue of Maps . . .* 1566–1855, ed. by F. G. Emmison (1947; 1952); *English History From Essex Sources*, 2 vol. (1952), 1550–1750, ed. by A. C. Edwards, and 1750–1900, ed. by A. F. J. Brown. *See* also Royal Commission on Historical Monuments, *An Inventory of the Historical Monuments in Essex*, 4 vol. (1916–23); *Transactions* of the Essex Archaeological Society (1858); *Transactions* of the Epping Forest and County of Essex Naturalists' Field Club (1881–92) succeeded by *The Essex Naturalist* (1887 *et seq.*); *Essex Review* (1892 *et seq.*); Kelly's *Directories of Essex* (1845–1937); N. B. L. Pevsner, *The Buildings of England: Essex.* (1954).) (A. V. Wo.)

ESSEX, KINGDOM OF, one of the kingdoms of Anglo-Saxon England, is shown by place-name evidence to be an area of early settlement, though archaeological remains are not plentiful. It probably originally included Middlesex, and London was its chief town. Its royal line differs from all others in tracing its descent from the Saxon god Seaxneat instead of from Woden. In 604 its king was Saeberht, son of Ricule, sister of Aethelberht of Kent. Aethelberht persuaded Saeberht to adopt Christianity, built St. Paul's in London and established Mellitus as bishop. However, when Aethelberht died in 616 and Saeberht soon afterward, the latter's three heathen sons expelled Mellitus, and Essex was not reconverted until soon after 650, when its king, Sigeberht II, was persuaded by Oswiu of Northumbria to be baptized. Bishop Cedd was sent as missionary and built churches at Ythancaestir (Bradwell-on-Sea) and Tilbury. Sigeberht was murdered by kinsmen, but his successor, Swithhelm, was converted to Christianity by King Anna of East Anglia, not later than 653. Essex often had joint kings, and at the time of the pestilence of 664 Swithhelm had been succeeded by Sigehere and his uncle Sebbi, of whom the former with his people renounced Christianity, while Sebbi and his subjects remained faithful. Their overlord, Wulfhere of Mercia, sent Bishop Jaruman to restore the faith. Not long afterward, Wulfhere sold the see of London to the West Saxon Wine, which proves that Mercian power was already strong in this town, and in the 8th century Mercian kings were supreme there. Archbishop Theodore in 675 replaced the simoniacal Wine by Eorcenwald, whose sister Aethelburh founded the one important East Saxon house, the double monastery of Barking. Sebbi retired into a monastery, after a reign of 30 years, in about 694. His two sons Sigeheard and Swaefred were associated with him as kings in a charter of about 690–693. About this time the East Saxons had some control in Kent, for the Swaefheard whom Bede describes as ruling there with Wintred refers in a charter to his father King Sebbi, and an earlier East Saxon king, Sigehere, made a grant of land to Hoo, Kent. Swaefheard's charter is confirmed by King Aethelred of Mercia, and he may have owed his power in Kent to his overlord's gift. In 709 Sigehere's son King Offa of Essex—the only member of the royal family whose name does not begin with *s*—abdicated and went to Rome with Coenred of Mercia. From the Anglo-Saxon Chronicle and from charters and lists in the chronicle of Florence of Worcester we learn of later kings—Selred, killed in 746; Swithred; Sigeric, who went to Rome in 798; and Sigered, who attests charters of Mercian kings up to 823. All were doubtless subject to Mercian overlords.

The East Saxons submitted to Egbert of Wessex in 825, and Essex became part of the subkingdom held by Egbert's son Aethelwulf. It was given by him to his eldest son, Aethelstan, in 839 and, on Aethelwulf's death in 858, to his third son, Aethelberht. When the latter succeeded to Wessex in 860, this subkingdom ceased to have a separate existence. Nothing is known of Essex under West Saxon rule. Essex became part of the kingdom of the Danish leader Guthrum by his treaty with Alfred the Great (878).

It accepted in 902 Edward the Elder's rival, his cousin Aethelwold, who had been taken as king by the Northumbrian Danes; but he was killed later in that year. Edward began its reconquest in 912 by building a borough at Witham, which he followed by others, Maldon in 916 and Colchester in 917. Henceforward Essex was in the charge of English ealdormen, the most famous being Brihtnoth, the hero of the battle of Maldon in 991. Yet, as late as this, there was a conspiracy to receive Sweyn I of Denmark as king in Essex. (D. Wk.)

ESSLINGEN (Esslingen am Neckar), a town of Germany which after the partition of the nation following World War II was included in the *Land* (state) of Baden-Württemberg of the Federal Republic of Germany. It lies in a fertile district on the Neckar river 14 km. (9 mi.) S.E. of Stuttgart by road. Pop. (1961) 83,236. On a commanding height above the town stands the old citadel and in the inner town are three town halls: the old *Rathaus* (about 1430), the new *Rathaus*, formerly a palace, and the baroque *Rathaus* (1705–15) of the imperial city, now the district court. The building of the Church of Our Lady (*Frauenkirche*), with its magnificent tower, built during the 14th to 16th centuries marked the beginning of the special Gothic style of architecture that developed in Swabia (*Hallenkirche*). St. Paul's (1233–68) once belonging to a Dominican monastery, is the oldest existing church of mendicant friars in Germany. The municipal archives contain valuable records on the medieval, Reformation and baroque periods. Esslingen is now an important industrial town with railway, machine and electrical works and spinning mills; cloth, gloves and metal goods are also made there. In the middle ages it was famous for its wines.

Esslingen was mentioned in the records as Cella in 777 and as Hetsilinga in 866. It presumably received its charter before 1219 under Frederick II (1194–1250), who surrounded it with walls. After the Great Interregnum (1254–73) it became a free imperial city. Often threatened by the rulers of Württemberg, it did not become part of that country until 1802. (O. Bo.)

ESSONNE, a *département* of France formed in 1964 from the southeastern part of the former *département* of Seine-et-Oise south of Paris. It is bounded north by the *départements* of Hauts-de-Seine and Val-de-Marne, east by Seine-et-Marne, south by Loiret and west by Eure-et-Loir and Yvelines. When the *département* was formed, its population was estimated at about 600,000.

The limestone platform that lies south of Paris has extensive cappings of sand and sandstone, which form picturesque wooded ridges aligned from northwest to southeast. The valleys of the Orge, the Essonne, and its tributary the Juine, add further diversity to the landscape. Much market gardening is carried on along their marshy floors, where the streams flow sluggishly in braided courses. Beyond the Seine Valley in the east of the *département* the western margins of the limestone platform of Brie are reached, but northeast of Corbeil the extensive forest of Sénart interposes between the large, rich farms of Brie and the Seine Valley. In the southwest of the *département* is the featureless plain of Beauce (*q.v.*). Its rich arable land, producing heavy crops of wheat and sugar beet, preserves a profoundly rural aspect, but nearer Paris the population of the *département* is increasing rapidly. The outer suburbs of Paris have encroached extensively in the north, and prongs of Greater Paris extend up the Seine Valley and farther west toward Arpajon. Trunk roads from Paris to the southeast run on each side of the Seine Valley, and the new southern motorway from Paris to the Rhône Valley has been built farther west across the plateau. The main road from Paris to Orléans runs south by Montlhéry, where the medieval tower provides an extensive view, and Arpajon to Étampes and then across Beauce. The Paris-Orléans railway, leaving Paris by the Seine Valley, pursues a less direct course to Étampes. The *préfecture* is at Corbeil, an old town (with a medieval gateway surviving) and riverport on the Seine, which has provided a nucleus for industrial development. Flourmilling is especially important. Essonnes, on the opposite bank at the confluence of the Essonne tributary, has paper works and other industries. Dourdan and Étampes are old-established market towns on the northeastern margin of Beauce. (Ar. E. S.)

ESTAING, CHARLES HECTOR, Comte d' (1729–1794), French admiral who during the American Revolution commanded the first expedition in support of the colonies, was born at Ruvel in Auvergne, of an ancient family, on Nov. 24, 1729. As an army officer, he went to India in 1757 and was taken prisoner at the siege of Madras (1759). Lieutenant general in the navy at the age of 33, he was governor of the Antilles from 1763 to 1766. In 1777, however, the new appointment of "vice-admiral of the seas of Asia, Africa and America" was created for him; and in 1778 he was instructed to surprise the English squadrons in North America so as to enable the colonists to resume the offensive. He knew something of the sphere of operations and showed energy and courage in his task; but his officers, who were jealous of his promotion out of turn and disapproved of the liberal opinions that he expressed with lordly persistence, failed to carry out his orders. Furthermore he was no sailor and his strategy was based on his experience of fighting on land. Arriving in Delaware bay in July 1778, he blockaded Adm. Richard Howe (*q.v.*) at Sandy Hook but did not dare cross the bar to attack New York. From there, in accordance with an agreement with Washington, he sailed to Newport, R.I., where, however, a storm prevented an engagement with the English fleet. Finally on Nov. 4 he sailed for the Antilles. There, in December, he could have destroyed an English division at St. Lucia but preferred to attempt a landing, which was repulsed. He captured Grenada on July 3, 1779, and fought an indecisive battle with Adm. John Byron on July 6. Though his forces were numerically superior, he did not again engage the English fleet, but launched an unsuccessful attack on Savannah, Ga., during which he was seriously wounded (Sept.–Oct. 1779). He then took his squadron back to France. When peace was signed with England in 1783 he was in command of the fleet before Cádiz. (*See* also American Revolution: *The War at Sea.*)

A man of advanced ideas, the Comte d'Estaing was elected to the assembly of notables in 1787. On the outbreak of the Revolution (1789) he was commandant of the national guard at Versailles. He was guillotined in Paris, during the Terror, on April 28, 1794. (L. Ni.)

ESTATE AND HOUSE AGENTS. In England the term "estate agency" covers most of the work carried out by those engaged in the profession of auctioneer and estate agent, which is mainly concerned with the sale (by auction or private treaty), letting, management and valuation of real estate. Estate agents also undertake to advise clients who wish to buy or to lease property. These functions require a knowledge of local government, the law of real property, the law of contract, the principles of valuation, building construction, compensation for compulsory acquisition and housing and public health and town-planning legislation. The management of property is an important part of an estate agent's work. It includes advising on such matters as land use, rent, rating assessments, property tax, improvements, repairs and dilapidations, as well as on the wide variety of questions which arise between landlord and tenant and between property owners and public authorities. Firms practising as auctioneers and estate agents may also engage in the valuation and sale of chattels, including, in rural areas, livestock.

No licence is required to enable a person to practise as an estate agent, but the leading professional bodies demand a high level of theoretical and practical knowledge from those whom they admit to full membership. These bodies also regulate the scales of professional charges, which vary with the price, valuation or rent roll of the property concerned. (*See* also Auction; Valuation.)

In the United States those dealing mainly with the sale and management of real estate are commonly known as real estate agents or brokers. The term "realtor," a trade-mark of the National Association of Real Estate Boards, denotes a person engaged in the real estate business who is an active member of the association and who subscribes to its ethics. (R. S. Bo.)

ESTATE AND INHERITANCE TAXES. Death taxes are of two kinds: (1) An estate tax imposed upon the privilege of transmitting property at death; this is a levy upon the entire capital value of all property changing hands at death, fixed mainly by reference to the total value of that property. (2) An in-

heritance tax imposed upon the privilege of receiving property from a decedent at death; this is a levy upon the benefits accruing at death to each beneficiary, calculated by reference to the amount of each benefit and the relationship (if any) of the beneficiary to the deceased. In some systems the value of the property already owned by the beneficiary also enters into the inheritance tax calculation. Death taxes are considered transfer taxes; the tax is a levy upon the transfer of the property, measured, in part, by the value of the property, and not a direct tax upon the property itself.

Estate taxes are usually easier to administer than inheritance taxes. Generally, for an estate tax, only the value of the entire estate need be ascertained. For an inheritance tax the value of each interest passing to each beneficiary must be fixed, and this assessment often necessitates complex actuarial calculations for future interests. An estate tax, since it is based upon the size of the total estate, will usually produce more revenue than an inheritance tax, and for the same reason will perhaps be more effective in breaking up large estates. On the other hand, an inheritance tax may be more discriminating and fairer since it takes into account exactly what each beneficiary actually receives and also his relationship to the deceased. An inheritance tax may encourage adequate provisions by testators for their natural dependents, since the rates for bequests to them are lower than for bequests to strangers. Also, it may result in the breakup of large estates by stimulating their division among the largest possible number of beneficiaries in order to minimize the tax, since the rates of the tax are progressive, increasing with the size of the benefit to each beneficiary.

Purpose and Effect.—Death taxes have two main purposes. The first is to provide revenue. They are, however, a relatively minor source of revenue for modern countries. In the United States the federal estate tax has never amounted to more than 7% of the total tax revenues of the federal government, and after 1942 it fell below 2%. In Canada death taxes generally have not exceeded more than 1% of the tax revenues of the dominion government. In Great Britain the estate duty in the 1950s averaged about 5% of the total tax revenues.

The second main purpose of death taxes is to secure objectives of social policy. They are used to break up large estates in order to prevent the transmission from generation to generation of great accumulations of wealth. The adoption of progressive rates in death taxes so that the percentage of the tax imposed increases with the size of the estate or of the bequest emphasizes this aspect. Also, death taxes may be the only effective way to reach the possession of wealth, as distinct from its receipt. An income tax, based on receipts, may impose the same tax on $5,000 of income whether it represents earned income, *e.g.,* salary, or investment income, *e.g.*, dividends, interest or rents. Death taxes to some extent counterbalance the usual advantages of investment income, although, of course, other attempts have been made to do this, as by applying the income tax to capital gains or to the receipt of gifts and inheritances, or by distinguishing between earned and unearned income under the income tax.

History.—In Great Britain the earliest form of death tax, apart from feudal dues, was a stamp duty upon grants of probate and administration in 1694. The estate duty, based upon the Finance act of 1894, is the only death tax in force. There were inheritance taxes in Great Britain from 1794 until 1949, when they were abolished.

In the United States the first death taxes were imposed by the federal government in 1798, but they did not become a permanent part of the federal fiscal system until the federal estate tax—the only federal death tax in effect—was adopted in 1916. All U.S. states except Nevada have death taxes, the nature of which varies widely. Most states have inheritance taxes, but some have estate taxes, and a few have both. After 1926 the states were encouraged to impose death taxes because the federal estate tax, to a limited extent, grants a credit to an estate if, but only if, that estate has actually paid a death tax to the state. There were proposals that the federal government abandon the death tax field entirely to the states. In view of the mobility of wealth, especially intangibles, and of people within the United States, and the attempts of certain states to attract people and wealth by becoming havens for the well-to-do through the lack of death taxes, the withdrawal of the federal government from death taxation might make death taxes ineffective in the United States.

In Canada provincial governments first enacted death taxes in 1892. The dominion government did not enter the field until 1941. The dominion death tax in force in the early 1960s is an inheritance tax which became effective Jan. 1, 1959, when it replaced the succession tax which had been in effect during the period 1941–59. By agreement between the dominion government and various provincial governments, most of the latter suspended all their death taxes and rented out these taxes to the dominion government. Only Quebec and Ontario, which did not enter into any such agreements with the dominion government, impose inheritance taxes. The rates of the dominion tax were doubled when the tax rental agreements with the provincial governments became effective. However, the dominion tax may be reduced up to one-half for death taxes paid to Quebec or Ontario.

Rates.—In Great Britain the progressive rates of the estate duty in 1962 ranged from 1% on estates of a net value of more than £3,000 but not in excess of £4,000 to 80% on estates valued in excess of £1,000,000. This was a flat rate, applying to the entire value of the estate, including the first £3,000 if the estate exceeded £3,000. However, the amount of the duty was, if necessary, reduced so as not to exceed the total of (1) the maximum duty payable at the next lower rate, plus (2) the amount by which the estate exceeded the maximum to which the next lower rate was applicable.

In the United States the rates in 1962 ranged through 25 brackets from 3% on a taxable estate of $5,000 or less to 77% of the amount by which a taxable estate exceeded $10,000,000. The rates were not flat, but applied only to such portion of each taxable estate as fell into the particular rate bracket. Also, the amount of the taxable estate was determined after subtracting a specific exemption (usually $60,000) and various deductions. For example, if the taxable estate after subtracting the exemption and deductions was $20,000, the tax was $1,600 (3% of $5,000 plus 7% of $5,000 plus 11% of $10,000).

In Canada the rates of the dominion inheritance tax in 1962 ranged through 23 brackets, starting with 10% on the first $5,000 in excess of $50,000 in a taxable estate. The rates progressed with a 2% increase in each bracket up to a maximum rate of 54% on the amount by which a taxable estate exceeded $2,000,000. These rates were not flat, but, as in the case of the federal estate tax in the United States, they applied only to such portion of each taxable estate as fell into the particular rate bracket. There was an exemption of $60,000 if the decedent was survived either by a widow or by an infirm widower and a child under 21 or infirm, with an added exemption of $10,000 for each additional such child. In all other cases, the exemption was $40,000, with an increase of $15,000 for each child under 21 or infirm, but in no case was there any tax unless the net estate, after all deductions, exceeded $50,000.

The statutory rates are frequently deceptive. In the United States, substantial amounts of property may entirely escape the federal estate tax because of certain exemptions or deductions. Thus a married person may leave up to $120,000 to a surviving spouse without having his estate pay any tax whatsoever, or up to one-half of an estate of any size may be left tax free to a surviving spouse. In addition, unlimited amounts may be left tax free to family charitable foundations, or unlimited amounts of life insurance on the life of the deceased for which he has paid virtually all premiums may be left free of tax to beneficiaries if the deceased at death had no incidents of ownership in the policies and did not assign the policies within three years of his death. Moreover, if an estate is left in trust for children for their lives, with remainder to grandchildren or to other descendants, the property, after incurring an estate tax upon the death of the settlor, will pass free of tax from one generation to another upon the deaths of the life beneficiaries.

Property Subject to Tax.—Death taxes usually are imposed upon all property passing at death present within the territory of

the taxing authority at that time, and also upon all property (except foreign realty), wherever situated, of persons who die domiciled within that territory. In addition, the United States imposes the federal tax upon the entire estate, no matter where situated (except foreign realty), of its citizens, irrespective of their domicile. The United States and Canada by a tax convention of 1961, Great Britain and the United States by a tax convention of 1944 and Canada and Great Britain by a tax convention of 1946 attempted to eliminate double death taxation so far as their respective death taxes were concerned (applying in the United States only to the federal estate tax). The United States federal estate tax also permits a limited credit for certain foreign death taxes.

In a federal government such as that of Canada or the United States, there is the problem of death taxes imposed on the same estate by both federal and local governments. This may be handled either by a credit mechanism in the federal tax, as in the United States, or by agreements between the governments involved. There is also the problem of duplication of taxation among local governments. In Canada, the provinces of Quebec and Ontario entered into an agreement to prevent overlapping of their death taxes. In the United States the overlapping of the death taxes of the 49 states having them was left by the federal government to adjustment by the states, about half of which had taken some steps to alleviate this difficulty by the early 1960s.

Generally speaking, all property of any kind in which the decedent owned an interest at death is subject to the death tax. This applies to all property which is included in his probate estate and in which he had an inheritable interest—an interest he could transmit at death. If only such property were subject to a death tax, however, avoidance of the tax could easily be accomplished by giving away property shortly before death. Therefore, death taxes usually extend to property which the deceased transferred to others before death in certain specified ways. Commonly it is provided that all transfers made by the deceased within a certain fixed period—three or five years—before death are either subject to tax or are presumed to be transfers made in contemplation of death and therefore taxable unless proved by the estate to have been made for a life rather than a death motive. In Great Britain, for example, estate duty is payable in respect of gifts *inter vivos* (between the living) made within five years before the donor's death; but this somewhat arbitrary rule was mitigated after 1960 by providing for reductions by fixed percentages of the estate duty so payable, according to the amount of time elapsing between the gift and the donor's death. Thus under British law, if a man makes a gift tomorrow, the full estate duty will be payable on it if he dies in the first or second years thereafter; but if he dies in the third year, only 85% of the duty otherwise leviable in respect of the gift will be payable; if in the fourth year, only 70%; if in the fifth year, only 40%; and of course if he survives the fifth year, no duty will be payable.

Certain other *inter vivos* transfers by the deceased, no matter how long before death they are made, may also be subject to a death tax. These usually include transfers where the deceased retained certain rights in connection with the property, such as: (1) power to revoke it or to alter the terms of disposition of it or to control it; (2) the right to receive the income of the property or to use it until his death; (3) the right to recover the property for himself or his estate if certain contingencies occur, or if beneficiaries fail to meet certain conditions. A death tax may even be imposed where the deceased retained no interest in or control over the property except that his death is the event which causes a transfer of the property to the beneficiaries. In Great Britain, when a life tenant dies, the entire property is subject to a death tax even if the life tenant never possessed more than a life interest in the property. In the United States, there is no estate tax in such cases unless the life tenant possessed a general power of appointment over the property or unless he originally owned the property.

Most death taxes also contain special provisions: (1) to tax property subject to a general power of appointment upon the death of the holder of the power; (2) to tax proceeds of life insurance upon the death of the insured even if the proceeds are payable to named beneficiaries at the death of the decedent; (3) to tax dower and curtesy interests; and (4) to tax the entire value of jointly owned property where, upon the death of one owner, the rights of the surviving owners are enlarged (except to the extent that the survivors gave full monetary consideration for their interests). Generally, transfers which do not deplete the deceased's estate—those for which he received full monetary consideration—are not subject to a death tax; transfers for which partial consideration was received by the deceased are taxed only upon the value of the property in excess of the consideration.

Further, to prevent avoidance of death taxes by outright *inter vivos* gifts of property long before death, a gift tax is often imposed on such transfers. The federal gift tax in the United States has been, in the main, ineffective from this standpoint. This result is partly because the rates of the federal gift tax are much less than the comparable estate tax rates. It also results from the fact that transfers under the gift tax are not cumulative with those under the estate tax so that by dividing transfers between the two taxes the full benefit of the exemptions of each are obtained and the transfers under each also fall into the lowest possible brackets of the progressive rate schedules. To cope with the problem, proposals have been made for a single cumulative integrated transfer tax on all transfers, *inter vivos* or testamentary, by each individual, at least if the transfer is from one generation to another.

Deductions and Exemptions.—Death taxes usually are imposed on the net value of the property transferred. Deductions are therefore allowed, to a certain extent, for the expenses incurred in administering the estate, for certain losses during this period, and for the bona fide debts and incumbrances of the deceased, if incurred for full monetary consideration. Usually certain other deductions or exemptions are allowed on grounds of social or public policy. Thus all or part of amounts transferred for public, charitable or religious purposes may be tax exempt. This provision resulted, in the United States, in the widespread use of charitable foundations which may sometimes permit the transmission of the control, if not the ownership or income, of vast fortunes within a family group, without any real breaking up by the estate tax of the family concentrations of wealth and power. In Great Britain, when a gift *inter vivos* is made to a charity, the five-year period of potential liability for estate duty on the gift is reduced to one year. Most death taxes completely exempt estates whose value does not exceed a stipulated amount. In Great Britain estates of £3,000 or less pay no estate duty. In the United States estates not in excess of $60,000, and in Canada estates not in excess of $50,000, are tax free. In Great Britain agricultural land and certain forms of business property pay only 55% of the statutory rates, and there are special provisions for timber, pictures, works of art, books and collections of a national, historic, artistic or scientific interest. Customarily, there are special relief provisions for estates of persons who die while serving in the armed forces.

In the United States the favoured position of community property under the federal estate tax has resulted in the marital deduction, which permits all noncommunity property passing to a surviving spouse not in excess of one-half of the estate to be tax free.

Death taxes, if they operate in conjunction with gift taxes as in Canada and the United States, usually make some provision for taking into account any gift taxes previously paid by the deceased on property also subject to a death tax. Death taxes often contain some provision for relief from the second death tax where the same property is subjected within a given period of time—five or ten years—to two or more death taxes because of two or more deaths within this period. Thus in Great Britain "quick succession" relief is granted on a sliding scale, ranging from a 75% reduction in duty on property which has passed on a death within three months of the relevant death, down to 10% where the first death was between four and five years before the relevant death.

See also INHERITANCE.

BIBLIOGRAPHY.—*Great Britain:* R. Dymond and R. K. Johns, *Death Duties,* 13th ed. (1960); G. M. Green, *Death Duties,* 4th ed. by C. D.

Harding (1958); A. Hanson, *Death Duties,* 10th ed. by H. E. Smith (1956); *Inland Revenue Reports* (annual).

United States: C. L. B. Lowndes and R. Kramer, *Federal Estate and Gift Taxes* (1962); R. Paul, *Federal Estate and Gift Taxation* (1942), Supplement (1946); W. C. Warren and S. S. Surrey (eds.), *Federal Estate and Gift Taxation* (1961); J. Mertens, *The Law of Federal Gift and Estate Taxation* (1959); *Annual Reports of the Secretary of the Treasury.*

Canada: Dominion Bureau of Statistics, *The Canada Year Book* (annual), *Canadian Succession Duties Reporter* (1961); Department of National Revenue, *Taxation Statistics* (annual). (RT. K.)

ESTATES, ADMINISTRATION OF, the proceedings of Anglo-American law in which the estate of a deceased person is wound up by an executor or administrator and distributed by him to the person or persons entitled thereto.

History and Function.—When a person dies the living persons who are to succeed to his property are indicated by his last will and testament or by the rules of the law of intestate succession (*see* WILL; INTESTATE SUCCESSION; INHERITANCE). Where a paper is produced purporting to be the decedent's will, it cannot under Anglo-American law be acted upon until it has been admitted to probate, *i.e.,* until its genuineness and effectiveness have been authoritatively ascertained by the proper court (*see* PROBATE). But there may also be other questions to be settled in the winding up of a decedent's estate. For example, the language used by the testator in his will may be ambiguous and thus require judicial construction (*e.g.,* the will may say, "I leave $10,000 to my niece Anne," and there are two nieces named Anne). Or an unforeseen contingency may have arisen (*e.g.,* the one niece Anne, to whom the testator gave the $10,000 legacy, died before him; what is to be done with the $10,000? Or the total amount of all legacies may be higher than the fund available for their payment; are all legacies to be abated in the same ratio, or are certain legacies to be preferred over others?) In the case of ambiguity the testator's language has to be interpreted; in the case of incompleteness the will may have to be construed. In cases of both testate and intestate succession the takers may have to be located, outstanding claims of the decedent and other assets may have to be collected and, above all, debts that were outstanding against the decedent at the time of his death must be paid and so must death duties, estate taxes and inheritance taxes.

Roman law took care of all these problems by simply leaving it to the person believing himself to be the decedent's principal successor, his heir, to take hold of his assets by his own efforts. If anyone else wished to contest his right, he had to go into court and sue in the regular forms of civil procedure. The payment of debts, legacies and taxes was secured simply by holding the heir liable for them personally, *i.e.,* out of his own pocket, and thus possibly beyond the value of the assets of the estate. This onerous liability could be avoided by the heir's refusing to accept the inheritance, which, of course, meant that he would not receive any surplus that might ultimately emerge. In order to avoid this hardship, there was established in the 6th century A.D. the possibility of accepting the inheritance under the benefit of inventory: if the heir made an inventory of the estate and took care not to commingle its assets with his own, his liability for the debts of the decedent was limited to the assets of the estate. This manner of handling the problems is still basically that of the countries of the civil law; *e.g.,* France, Brazil or Japan.

In the countries of Anglo-American law, a different pattern prevails. It was developed in the ecclesiastical courts of England, which had complete jurisdiction in matters of succession to personal property until the 16th century, and then, in competition with the court of chancery, until 1857. Under this system the estate of a decedent does not immediately pass upon his death to the ultimate takers but to an intermediary, by whom it is to be administered and distributed to those ultimately entitled to it by the will of the decedent or by the law of intestate succession. The intermediary can be appointed by the decedent himself in his will, in which case he is called the executor, or he is appointed by the court, in which case he is called the administrator. In England and the commonwealth the term personal representative is frequently used to refer to both executor and administrator.

Since the ecclesiastical courts had nothing to do with the descent of real estate, their system applied only to the personal property left by a decedent. His real property descended directly, in the case of a will, to his devisee or devisees or, in the case of intestacy, to his heir. The principal merit of the system of administration lies in its providing an orderly way for the payment of the debts of the decedent before the estate is dissipated. Since a man's real property was usually not liable for the payment of his debts, no need existed for the extension of administration beyond the personal property. The notion that land, as constituting a family's basis of existence, ought to be exempt from attachment by creditors unless specifically mortgaged, appears no longer compatible with the needs of modern market economy. Real property has thus been made generally liable for the payment of its owner's debts. This new state of affairs required the powers of the personal representative to be extended to real property. This extension has not been carried out, however, in the same manner in all jurisdictions. In the Administration of Estates act, 1925, England has gone the whole way. Upon the death of a decedent, title to all his property, personal and real, passes to the personal representative, who is to hold and administer it under the statutory trusts set out in the act. His central duty is the payment of the debts and taxes. He must then distribute the surplus to the legatees or next of kin, as the case may be. All differences between personal and real property have been abolished, and all assets are to be used on an equal footing to raise the cash necessary for the payment of debts, taxes and, as a general rule, also of legacies.

In the United States only a few states have gone so far; *e.g.,* North Dakota. In some, title to all assets passes to the personal representative, who, however, is not to use real estate for the payment of debts until the personal estate has been exhausted or unless different directions have been given by the decedent in his will. In the majority of states, title to real estate still does not pass to the personal representative but directly descends from the decedent to his heirs or devisees; however, if the personal estate does not suffice for the payment of claims, the court will order that the remaining portion be raised through the sale or mortgage of real estate. In another group of states, *e.g.,* Texas, title to all assets, both real and personal, passes directly to the ultimate takers but subject to the power of the personal representative, insofar as necessary, to take possession of and to sell it for the payment of claims, regularly using personal property before real.

In Scots law as well as in the Roman-Dutch law of the Republic of South Africa, the civil-law basis of the law of succession has come to be so modified that liability for the debts of the decedent is generally limited with the result that administration is regarded as a regular stage in a case of succession.

As compared with the Roman law system of settlement of the estate by the heir, the Anglo-American system of administration has the advantage of providing an orderly method for the payment of claims and taxes as well as supervised management in the case of litigation or delay in the final distribution. But it can be expensive, especially where, as in the majority of U.S. states, the personal representative is not supposed to sell any assets, compromise any claim or proceed to distribution without express judicial authorization. The cost is considerably less under the English system, where judicial interference is limited to situations of dispute or trouble, or under the civil-law system, where proceedings corresponding to administration do not take place unless the interest of creditors actually requires that the estate be taken out of the hands of an unreliable, insolvent or incompetent heir, or unless temporary measures of conservation are required, as in cases of dispute or of missing heirs.

Appointment of Executors and Administrators.—In England and the commonwealth countries an executor derives his title from the will and may act as soon as the probate registry or court has granted probate of the will containing his appointment. In the United States an executor's powers are regularly regarded as being based not upon the appointment by the testator but upon a special act of the probate court, the issuance of letters testamentary. Such letters will regularly not be issued until the executor has been sworn in and, unless the testator has directed otherwise,

until the executor has posted a bond. In the case of intestacy an administrator has to be appointed by the court. A right to be appointed, unless found unsuitable for serious cause, is given by statute to those persons who are, or are likely to be, the principal takers of the estate. In both England and the United States the administrator is an appointee of the court, whose office begins with the grant to him of letters of administration, an act which in the United States does not usually take place until a bond has been furnished.

Once appointed, the powers of an administrator as well as those of an executor are deemed to have arisen immediately upon the death of the decedent, so that acts performed before the grant of probate or the issuance of letters are retroactively validated.

Where a will has been admitted to probate but no executor has been appointed in it, or the person is unwilling or unable to act, the person appointed by the court as personal representative is called "administrator with the will annexed" (*administrator cum testamento annexo* or, briefly, *administrator c.t.a.*).

A personal representative whose appointment by the court has become necessary because the executor or administrator has died, resigned or been removed before the completion of the administration is called *administrator de bonis non administratis* or, more briefly, *administrator de bonis non* (*administrator d.b.n.*). A person appointed temporarily to take care of assets of a decedent's estate because of such special need as litigation about the validity of the instrument purporting to appoint an executor is called in England *administrator pendente lite* or, in the United States, "administrator to collect."

In appropriate cases several persons may be appointed to act as co-executors or co-administrators. In England their number is not to exceed four.

If assets are situated in more than one jurisdiction (for instance, in the United States in more than one state) it may be necessary, in addition to the domiciliary administration carried on in the state of the decedent's last residence, also to have so-called ancillary administrations in the other state or states.

Duties and Powers of Personal Representatives.—The personal representative is required to settle the estate as expeditiously and economically as possible. His principal tasks are the following: (1) to collect all outstanding claims of the decedent and bring together all assets of the estate; (2) to have the assets appraised and inventoried; (3) to file all returns for estate, inheritance and other taxes and attend to their proper payment; (4) to attend to the payment of the debts of the decedent; and (5) to distribute the balance to those beneficially entitled. For this purpose statutes, especially in the United States, provide that: (1) by public notice the personal representative shall call upon all persons with claims to file them within a certain period of time; (2) all claims thus filed shall be passed on by the personal representative and, if necessary, be submitted to adjudication; (3) cash necessary for the payment of claims found proper shall be raised, if necessary by the sale of assets, and such sales shall be made in the proper order and manner; (4) all claims found proper shall be paid, if necessary in the order of abatement prescribed by the will or the law; and (5) the remaining assets shall be distributed to the persons entitled thereto under the will or the rules of the intestacy law.

In the United States, but not to the same extent in England, the personal representative is supposed to act as an officer of the court, whose approval he ought to seek for every act of management and distribution. At the end of his administration he is supposed to submit an account, upon the approval of which the personal representative and his bondsman are discharged and the estate is closed.

In all his duties a personal representative is held to a high degree of care, the violation of which renders him personally responsible, especially to creditors, tax authorities and distributees. He is entitled to reimbursement for expenses, provided they are held to have been incurred properly, and he is to receive a fee for his activities in general. His task can be arduous. For anyone acting, or considering acting, as executor or administrator, it is advisable to seek expert legal assistance, especially in the United States. *See* also ESTATE AND INHERITANCE TAXES.

BIBLIOGRAPHY.—T. A. Atkinson, *Handbook of the Law of Wills*, 2nd ed. (1953); M. Rheinstein, *Law of Decedents' Estates*, 2nd ed. (1955); L. M. Simes and P. E. Basye, *Problems in Probate Law* (1946); N. E. Mustoe, *Executors and Administrators*, 5th ed. (1952); D. H. Parry, *Law of Succession, Testate and Intestate*, 4th ed. (1961); E. V. Williams, *Law of Executors and Administrators*, 14th ed. by G. W. Keeton and E. H. Scammell, 2 vol. (1960). (M. RN.)

ESTATES-GENERAL. The assemblies of estates which came into being in the majority of European kingdoms and principalities in the course of the later middle ages were the beginnings of organized political representation of the subjects and led with varying degrees of continuity to the modern parliaments (for England *see* PARLIAMENT; for Spain *see* CORTES).

These assemblies, down to the end of the 18th century, consisted of representatives of social classes or estates. The first estate was that of the clergy, the second estate was that of the nobility, the third estate represented the other privileged classes; *i.e.*, mainly the bourgeois of the towns. The name estates was extended from the classes to their representatives' assembly, though in some regions (for instance Flanders and Hainaut) the term parliament was often used for it until *c.* 1400 and in others (Germany and Poland) the term diet was used. Together the three estates came eventually to speak not simply for their social classes but for the whole country, though the voice of the nonprivileged (the peasants and some towns) was rarely heard. Sometimes one estate would not attend a meeting of the other two: the clergy was absent from a number of Dutch assemblies and from many in the German principalities after the Reformation; and the knights sometimes failed to achieve an active role (*e.g.*, in Flanders).

The ruler convened the estates and normally chose the subjects for discussion. The deliberations of the various estates were often secret, but could be combined in various ways. A unanimous decision was often required, even from each estate, though governments, wanting a prompt reply, were in favour of decisions by simple majority. The Polish diet firmly conserved the unanimity principle, the notorious *liberum veto*. Deliberations were often prolonged because members had to go to consult the authorities in their provinces, towns, etc., before voting.

The representative assemblies had to discuss political and administrative matters with the sovereign and, particularly, to grant aids (*q.v.*) and taxes. They thus won some control over state expenditure and could influence government policy if the balance of forces within the state was in their favour, but their legislative role was very limited.

Origins.—From the late 13th century onward the assemblies were fairly frequent: the English Model parliament met in 1295; the first French estates-general met in 1302; the first Flemish *plein parlement* in 1306. Their origins are difficult to trace. In matters of national politics, especially dynastic crises, representatives of social groups had early negotiated with their rulers, sometimes from a position of strength, and the crown had had to negotiate the levy of certain taxes with the towns, the nobility and the church. Under feudalism (*q.v.*) the monarch, as feudal suzerain, consulted the barons, his tenants, to ask them for aids and could not tax them without their consent—an important point in the development of later taxation practice. When the town bourgeois emerged as a privileged estate, the sovereign had to recognize them too, especially in the financial field. With the rise of strong states and the beginning of bureaucratic, centralized government in the 12th century, as the personal, domanial and feudal revenues of the rulers were no longer adequate, consultation with the important sections of the population became indispensable.

Further Developments.—The estates were most prominent where there was an advanced central government, as in France and Aragon (their role was more limited where city-states were the normal political form, as in Italy). Gradually they won political power at the crown's expense, till in the 15th and even in the 16th centuries their powers in many European countries were greater than those of the parliament in England. With the subsequent rise of absolutism, however, conflict arose between the assemblies and the crown and usually ended in victory for the ruler, so that in

the 17th and 18th centuries the estates (with notable exceptions in Württemberg and Poland) were generally convened only to give formal assent to taxes and to listen passively to the government's decisions. In 1789, however, the French estates-general constituted themselves a National Assembly, abolishing the estates, and heralded a revival of European representative institutions (*see* FRANCE: *History;* FRENCH REVOLUTION, THE). The degree of success achieved by the establishment of modern parliaments in Europe in the 19th century is historically related to that previously achieved by the estates in the several countries.

FRANCE

The French monarchy rapidly became absolute. The estates controlled only the crown's use of the exceptional subsidies granted. When Philip IV convened the three estates to meet in Notre Dame in Paris on April 10, 1302, to discuss the conflict between him and Pope Boniface VIII, they stood firmly by the king; their meeting was followed by a nationwide consultation of public opinion. Likewise in 1308 the three estates were assembled at Tours to be consulted on the suppression of the Templars. In 1308, in 1309 and in 1314 the king consulted the towns on monetary and fiscal matters. After Louis X's death (1316), consultations on the royal succession and on fiscal matters were held repeatedly. In 1320 the estates, assembled at Pontoise and at Poitiers, refused a subsidy.

The Hundred Years' War brought representative institutions to the fore. In 1355 the estates-general, held in Paris and dominated by the towns, obtained full control of the subsidies granted. In 1356, during John II's captivity, the estates-general of the *langue d'oïl* (northern France) met in Paris and those of the *langue d'oc* (*see* LANGUEDOC) in Toulouse. The estates of the *langue d'oïl* then obtained control of the monarchy and its organs, proposed reforms and voted high taxes for the country's defense, but the revolt of Paris under Étienne Marcel (*q.v.*) alarmed the provinces and these first attempts at constitutionalism came to an end with Marcel's assassination (1358). After the war, the crown resumed complete control: Louis XI convened the estates-general only once, at Tours in 1468. After his death, however, in reaction against his depotism, the estates-general met at Tours in 1484. This was an important assembly, comprising more than 250 people and including for the first time representatives of the rural areas, but even so it achieved little, and the crown did not keep its promise to assemble them again in two years' time. Louis XII summoned the estates-general once only; Francis I, not at all.

After a temporary revival during the wars of religion (notably the meeting at Blois in 1588) and a meeting in Paris during the regency of Marie de Médicis (1614) the estates-general were allowed to lapse altogether until 1789 at the beginning of the Revolution (*see* above). Till then the French estates-general had never provided more than a temporary and disruptive check to royal absolutism. From the 16th century, moreover, the third estate had been undermined by its members' tendency to enter the nobility through the magistracy or through official functions (*noblesse de robe*). Provincial estates in France had been numerous from the 14th century. Most of them survived until the Revolution. Isolated and weak, they concerned themselves only with the defense of their own interests.

THE LOW COUNTRIES

By the end of the 14th century, estates were meeting regularly in the majority of the principalities of the Netherlands (*q.v.*), chiefly in order to vote the taxes requested by the government. In the country of Flanders, indeed, a "parliament" is recorded as early as 1279, parliaments were held twice a year and there was also, from 1306, a *plein parlement,* or "full parliament," (essentially a representation of the towns) beside the ordinary meetings. Furthermore, the "three members of Flanders" (namely Ghent, Bruges and Ypres) together played an important role as a representative body. In the episcopal principality of Liège the three estates shared in the legislative work of the government after the peace of Fexhe (1316). In the duchy of Brabant the estates reached the height of their power in the 15th century; moreover, thanks to the terms of the *Joyeuse Entrée* (*q.v.*) of 1356, which was renewed continually until 1792, these estates of Brabant were to enjoy more independence than did any others in the southern Netherlands throughout the period of Habsburg rule.

In the 15th century, in fact, during the Burgundian regime in the Netherlands, there was a steady rise in the influence of the estates, and in 1477, on the collapse of the regime at Charles the Bold's death, they assumed extensive political powers. Subsequently, however, they came into conflict with the Habsburg sovereigns. Philip II of Spain severely curtailed the powers of the estates in those territories which remained Spanish after the revolt of the northern provinces. After the southern provinces had passed to Austria, the estates there continued to decline till the revolt against Joseph II in 1788–89.

In the northern Netherlands, however, the estates remained important. In Holland the assembly of estates (*dagvaart*) had become a regular element of the government during the Burgundian period, as Philip the Good had needed to conciliate the people in order to raise money. In this assembly the knights and the six big towns had been represented, but seldom the clergy. In 1572, however, during the rising against Spanish rule, representatives of the dissident towns and a few noblemen had set themselves up as estates (*staten*) in Dordrecht and recognized William the Silent, prince of Orange, as stadholder. Consequently, when Philip II of Spain was finally repudiated in 1581, the "estates of Holland and West Frisia" became the sovereign authority in the country. They finally comprised the nobility and 18 towns, and unanimity was almost always required in their deliberations.

Estates-general emerged in the Netherlands because the dukes of Burgundy found it more convenient to hold common meetings of representatives of all their Netherlandish possessions than to negotiate with each one of the provincial assemblies separately.

Regardless of earlier attempts, the meeting in Bruges in 1463 is usually considered as the first meeting of the estates-general. Thereafter they met frequently but irregularly, mostly in Brussels. In the south their political role was finished in 1585, but in the north they continued to play an important political role (*see* NETHERLANDS).

THE HOLY ROMAN EMPIRE: GERMANY AND AUSTRIA

The *Reichstag* or diet of the Holy Roman empire (*q.v.*) was the representation of the estates before the emperor. It eventually came to consist of the college of electors (*q.v.*), the college of princes (nonelectoral princes, prelates, counts of the empire and free lords of the empire) and the college of the towns (a Rhineland bench of 14 and a Swabian bench of 37 towns). Real power, however, rested not so much with the emperor as with the rulers of the several principalities of the empire. This article, therefore, deals with the estates (*Stände*) or diets of these territories. In most German principalities, estates began meeting in the 14th or in the 15th centuries, taking advantage of dynastic confusion or monarchic weakness. The clergy was normally the first estate, but was not always represented and in some cases it disappeared with the Reformation; the nobility, especially in the south, often gained the status of free knights of the empire and so disappeared likewise from the estates. The townsmen, however, were regularly represented. The main business of the estates was to vote taxes, but they were also concerned to preserve the unity of the country against the ruler's attempts to dispose of it arbitrarily. Often influential during the Reformation and the Thirty Years' War, they found it hard to withstand the progress of absolutism from the middle of the 17th century onward.

Austria.—The Austrian estates consisted of clergy, nobility (subdivided often into *Herren* and *Ritter*) and town burgesses; in Tirol, where the estates were already important in the 14th century, the free peasants formed a separate estate. In 1402 the dukes Albert IV and William of Austria concluded an agreement with the clergy, nobility and towns for the promulgation of a territorial peace; in 1404 the estates of Austria met in Vienna; thenceforth they met irregularly, when the crown convened them. The *Landtage* or regional diets of the various possessions of the Habsburgs met together for the first time in the *Generallandtag* of 1446.

In the 16th–18th centuries the power of the estates declined: during the religious troubles they mostly had Protestant sympathies, and the victory of the Counter-Reformation was fatal for them (*see* AUSTRIA, EMPIRE OF). Finally the emperor Joseph II reduced them to insignificance, abolishing their right to grant taxes. Though Leopold II allowed them to meet, they were no longer consulted. The constitution of 1848 founded a new type of *Reichstag*, with a senate and a chamber of representatives.

Bavaria.—In Bavaria the representatives of the clergy, the nobility and townsmen assembled in the *Landschaft* exercised considerable power throughout the 15th century, lost most of it in the 16th and last met in 1669 (after 57 years of suspension). The separate estates, however, survived with limited powers till the end of the 18th century.

Brandenburg-Prussia.—In the Hohenzollern principalities the estates played a regular role in affairs in the 15th and 16th centuries. Their position in Brandenburg was weakened by the Great Elector, Frederick William, in the 17th century. The estates of Prussia strenuously resisted absolutism till Frederick William intervened with the force of arms at the *Landtag* of Königsberg (1661–63). It was then agreed that he would convene the estates every three years, but after 1671 this promise was ignored.

Mecklenburg.—The duchy of Mecklenburg had an old representative constitution which resisted all attacks, especially during the 18th century. An agreement of 1755 between the duke and the estates rather favoured the duke, but the estates preserved the essentials of their power. These estates outlasted those of any other country, surviving till 1918.

Saxony.—Originating in the 14th century, the Saxon estates came to the fore as arbiters in dynastic conflicts. The first diet to include representatives from all parts of the country met at Leipzig in 1438; counts, knights and towns, were represented, but not the clergy. Though the diet's powers declined notably in the 18th century the rules of procedure adopted by that of 1728 remained in force till 1831, when a new constitution, agreed between the king and the estates, was enforced.

Württemberg.—The first definite date for a meeting of the estates of Württemberg is 1457. Under the treaty of Tübingen (1514) they were granted extensive liberties by Duke Ulrich. The diet eventually came to be dominated by the towns, as the nobility left it on becoming free knights of the empire. The conversion of Duke Charles Alexander (d. 1757) to Catholicism led to conflict with his Lutheran subjects, which strengthened the resistance of the estates over the question of a standing army. Württemberg had a stronger tradition of representative institutions than any other German state and the constitution of 1819 was negotiated between the estates and the king.

ITALY

In Italy, estates developed only in the less urbanized regions. The parliament of Friuli was absorbed by the Venetian republic in 1420. In the papal states, assemblies were never important and ceased to play any role during the Renaissance. In the kingdom of Naples there were occasional parliaments composed of nobility and representatives of the cities but they met for the last time in 1642. The estates of Piedmont, from which the clergy was often absent and in which the representatives of the communes were the most regular element, were very active from 1460 till their last convocation, by Emmanuel Philibert of Savoy, in 1560, though from 1530 the feudal aristocracy and the communes had ceased to co-operate. In Sicily, King Martin I restored the parliament at the end of the 14th century after a period of anarchy, and the three estates, meeting every three years, kept their authority well into the 17th century. In Sardinia, also, the three estates (*stamenti*), meeting every ten years, resisted the growing absolutism of the 16th and 17th centuries.

BIBLIOGRAPHY.—Various authors, *Studies Presented to the International Commission for the History of Representative and Parliamentary Institutions* (1937–); *Anciens pays et assemblées d'États* (1950–). *France:* G. Picot, *Histoire des états-généraux*, 6 vol., 2nd ed. (1888); E. Lousse, *La Société d'ancien régime: organisation et représentation corporatives* (1943); F. Lot and R. Fawtier, *Histoire des institutions françaises au moyen âge*, vol. ii (1958). *Germany:* H. Spangenberg, *Vom Lehnstaat zum Ständestaat: ein Beitrag zur Entstehung der landständischen Verfassung* (1912); F. L. Carsten, *Princes and Parliaments in Germany* (1959); H. Sachsse, *Die landständische Verfassung Mecklenburgs* (1907); W. Grube, *Der Stuttgarter Landtag, 1457–1957* (1957). *Italy:* A. Marongiu, *L'istituto parlamentare in Italia dalle origini al 1500* (1949); E. Croce, *I parlamenti napoletani sotto la dominazione spagnuola* (1936). (R. C. v. C.)

ESTAUNIÉ, EDOUARD (1862–1942), French writer, known for his novels of character, was born in Dijon, Feb. 4, 1862. He was by profession an engineer and ended his career as inspector general of telegraphs. He was elected (1923) to the Académie Française. He died in Paris, April 1, 1942.

A theme recurrent in the 12 novels of Estaunié is expressed by the title of one of them, *La Vie Secrète* (1908): each man's outward life masks another, ill-understood, different and usually much more important life which may break through the mask unexpectedly to take temporary control of him. The novels gain distinction from their profound and detailed psychological analysis. The more important of them are: *L'Empreinte* (1895); *Les Choses voient* (1912); *L'Ascension de Monsieur Baslèvre* (1920); and *L'Appel de la route* (1922).

See J. Charpentier, *Estaunié* (1932); Camille Cé, *Regards sur l'oeuvre d'Edouard Estaunié* (1935); R. C. Hok, *Edouard Estaunié, the Perplexed Positivist* (1949).

ESTE (ancient ATESTE), a small town and episcopal see in the province of Padua, Italy, lies 31 km. (21 mi.) S.S.W. of Padua by road, at the southern extremity of the Euganean hills (Colli Euganei), 43 m. (141 ft.) above sea level. Pop. (1957 est.) 16,186 (commune). In antiquity it occupied a commanding position beside the Adige river and was for a time the capital of the Veneti; but in A.D. 589 the river changed course and now flows 9 mi. to the south. After a period of complete abandonment it was reoccupied in the middle ages, but it never regained its early importance, and the medieval town is best known as the place of origin, and until the 12th century the principal seat, of the Este family. (*See* ESTE, HOUSE OF.)

Early History and Archaeology.—In antiquity the position of Ateste, near the mouth of the Adige, protected from the south by the line of the river Po but readily accessible by land from the north and from the important trade-route that led round the head of the Adriatic, was one that offered many advantages, particularly in the field of commerce, and for many centuries Ateste was the chief centre of civilization in eastern Italy north of the Po. Throughout the 2nd millennium B.C. the Euganean hills were inhabited by Bronze Age peoples, probably anterior to Indo-European stock and possibly to be identified with the historical Euganei of the Alpine foothills. Early in the 1st millennium B.C. these early inhabitants were conquered and assimilated, or displaced, by an Early Iron Age invading people, almost certainly of Danubian origin, who settled at first in the hills (the first traces are on Monte Lozzo), but who soon afterward moved down to the river and founded Ateste, the capital of the historical Veneti.

The history of pre-Roman Ateste is known almost entirely from its material remains, most of which comprise the contents of the cemeteries by which the town was ringed around on the north, west and south, now preserved in the National museum at Este. The distinguishing feature of the new culture, the burial of the cremated dead in flat urnfields, is one that it shared with other contemporary cultures of central Italy, notably with that of Felsina (modern Bologna) and coastal Etruria. Throughout its history Ateste was in close and continuous contact with other centres, Felsina in particular, but it evolved independently, with its own characteristic products and its own consistent and continuous development.

The four periods into which this development is conventionally divided constitute a useful chronological classification; they should not be taken to imply actual breaks in the sequence. Of the first period, corresponding to the Benacci I phase at Felsina (*q.v.*), the remains are scanty and somewhat equivocal. The richest and most characteristic remains are those of the second and third periods, of which the former corresponds approximately to the second Benacci and Arnoaldi phases at Felsina, the latter to the Certosa phase, covering together the period from the early 7th

to the middle of the 4th century B.C. During the fourth period, after the Gaulish invasion of northern Italy, the Veneti remained politically independent; but they had lost a great deal of their individuality and tended to become merged in the general complex of a civilization from which much of the local character was fast disappearing.

The majority of the finds from the Atestine cemeteries consist of pottery and decorative bronze work. Of the former the most striking products are the cinerary urns, which show a continuous development from Villanovan forms into shapes and types of decoration that derive directly from the local metalwork. Peculiar to the Atestine series are vessels decorated with elaborate geometric designs of bronze studs (Period II) and a very handsome group of wares painted with red ochre and graphite in alternate zones of cherry-red and lustrous black. The outstanding metal work products are the buckets (*situlae*) of sheet bronze, riveted into shape and sometimes richly decorated with geometric and figured designs, engraved or in relief; they are closely related to the well-known *situlae* of Felsina, and were widely exported throughout the east Alpine region. Other characteristic bronze objects are ornaments, such as girdles or fibulae, and a lively series of figurines of human figures and animals, mostly from the 4th–2nd century B.C., votive deposits of a shrine to the goddess Reitia. There are few weapons and, in striking contrast to the nearby port of Adria (*q.v.*), very few Greek imports. Horses were evidently in high repute; and a number of poor inhumation burials are probably those of servants or slaves, some found in conditions that suggest ritual sacrifice on the death of their masters. Written inscriptions appear in the 4th century. Though written in north Etruscan characters, the language is Venetic, an Indo-European dialect of which examples are known from most parts of modern Veneto.

BY COURTESY OF NAZIONALE ATESTINO MUSEUM

BENVENUTI SITULA (BUCKET), EARLY EXAMPLE OF THE THIRD PERIOD OF ATESTINE ART, BRONZE; 5TH CENTURY B.C.

Although the Veneti retained their independence for some time after the Gallic invasion of northern Italy, this event really marked the beginning of their decline, and about 200 B.C. they were quietly and peacefully absorbed by Rome. Already in the 3rd century the old capital, Ateste, had begun to be eclipsed by the rising prosperity of Patavium (Padua); and despite the establishment by Augustus of a colony of veteran soldiers after the battle of Actium (31 B.C.), the Roman city was never important and it disappears from history after the 3rd century A.D. Of the actual Roman town, which occupied only a part of the area of the pre-Roman town and which is itself covered by the modern town, very little can now be seen. The principal surviving remains, now in the National museum, are from the cemeteries and from the villas of the surrounding countryside. They include the tombstone of a maker of bronze *situlae* and an inscription of 141 B.C. recording the settlement by a senatorial commissioner of a boundary dispute between Patavium and Este.

BIBLIOGRAPHY.—D. Randall-McIver, *The Iron Age in Italy* (1927), which gives the Italian authorities; G. Fogolari, *Il Museo Nazionale Atestino in Este* (1957). For the *situlae* see G. Ghirardini in *Monumenti Antichi*, vol. II (1893) and P. Ducati, *La situla di Certosa* (1923). For recent bibliography *see* G. Kaschnitz-Weinberg in *Handbuch der Archäologie*, pp. 387–391 (1954); Arte delle Situle dal Po al Danubio (1961). (J. B. W.-P.)

ESTE, HOUSE OF, an Italian princely family which ruled in Ferrara from the 13th to the end of the 16th century and in Modena and Reggio from the later middle ages to the end of the 18th century. They were of Lombard descent, a branch of the great 10th-century dynasty of the Obertenghi, who held power as margraves between the Lunigiana, Genoa and Milan and who also gave rise to the feudal houses of the Malaspina, the Pallavicino, and margraves of Massa and Parodi in Liguria. The family took its name from the township and castle of Este, 17 mi. S.W. of Padua, and its true founder was the margrave Alberto Azzo II, who died a centenarian in 1097. Alberto married Kunegonda, sister of Welf III, duke of Carinthia and margrave of Verona, who, being himself without children, adopted their son Welf as his heir. In 1070 Welf IV was created duke of Bavaria, and through him the Este became connected with the princely houses of Brunswick and Hanover. The Italian titles and estates, however, passed after some dispute to Folco I (d. *c.* 1136), son of Alberto by his second wife Gersende, daughter of Herbert I count of Maine. Neither he nor his successor Obizzo I (d. 1193) achieved any great distinction, beyond the offices and titles which fell naturally to the upper feudal families; but it was during the lifetime of Obizzo I that the Estensi first acquired political importance in Ferrara (*q.v.*), through a marriage contract imposed on the heiress of the Marcheselli family. The Marcheselli inheritance carried with it the leadership of one of the two great contending factions which divided Ferrara at the time. The opposing family were the Torelli, who in the years to follow were the main rivals of the Este for control of the city.

Obizzo I outlived his son Azzo V, so that in 1193 the succession passed to the latter's son Azzo VI (1170–1212), a man of marked ability who came to exercise considerable authority in fact if not in name over Ferrara; this position, and the possession of Este, involved him also in the politics of the neighbouring towns of Padua, Verona and Mantua. His growing influence was recognized by Pope Innocent III, who created him margrave of Ancona; but a premature death cut short Azzo's career and left the family temporarily weakened. One of his sons, Aldobrandino I (d. 1215), was dispossessed of Este by Padua, while the other son, Azzo VII (1205–64) surnamed Novello, was expelled from Ferrara in 1222 by Salinguerra Torelli. Not until 1240 did Azzo return to power in the city, in alliance with the Guelph league formed by Pope Gregory IX; at the end of a four months' siege the town was occupied and Salinguerra taken prisoner. Azzo was shortly afterward appointed *podestà*.

This marked the beginning of Este rule in Ferrara, which was none the less effective for being informal. It remained informal down to the accession of Azzo's heir Obizzo II (1240–93), the illegitimate offspring of Azzo's son Rinaldo, who had died a prisoner of the emperor Frederick II in Apulia. Obizzo II was formally recognized as Azzo's successor and then created perpetual lord by the people of Ferrara. Obizzo was also the first of the Este to be chosen lord of Modena (*q.v.*), in 1288, and of Reggio nell' Emilia (*q.v.*), in 1289. The progress of Este dominion, however, was disastrously halted soon after this by the impolicy and ambition of Obizzo II's son and successor Azzo VIII (d. 1308), who entertained designs on Parma and Bologna, but whose only practical achievement was to drive Modena and Reggio into revolt in 1306 and to compromise Este authority at home by excluding his brothers, Aldobrandino and Francesco, from all share in power, thus making them his enemies. This provoked conflict over the succession between his brothers and his bastard son, Fresco. The brothers were supported by Pope Clement V, who was overlord of Ferrara; Fresco was supported by the Venetians—to whom, however, he was forced to cede his rights. In the event it was the pope who prevailed, with the help of the Angevins of Naples; but in 1317 the Ferrarese rose against the Angevin King Robert's government and called back their former lords. Thereafter the Estensi ruled uninterruptedly for nearly three centuries. In 1329 Pope John XXII was induced to legitimize their power by granting them the title and authority of papal vicars *in temporalibus*, which they then held continuously until Borso d'Este (*see* below) was created duke of Ferrara by Pope Paul II in 1471. The Estensi also recovered control of Modena in 1336 and of Reggio in 1409.

During the 15th century the most notable members of the family were Niccolò III (1383–1441) and three of his sons. It was Niccolò who profited by the political collapse of the Visconti after 1402 to seize Reggio and for a time Parma as well. Later he became an ally of the Visconti and in 1441 was nominated governor of Milan, where he died the same year not without suspicion of

poison. In his lifetime he was celebrated for his large profusion of illegitimate children; he is also remembered for the cruel death that he was driven to inflict on his second wife, Parisina Malatesta, and his bastard son Ugo, who were found guilty of adultery together (1425). Another of his bastards, LEONELLO (1407–50), was chosen by Niccolò as his successor. Leonello had been educated by the humanist Guarino da Verona, called to Ferrara by his father; and from the middle of the 15th century, under him and his brothers, Ferrara was an ardent centre of Italian life and culture, a city of "dukes and poets," with one of the most brilliant courts in Europe. Leonello's brother and successor BORSO (1413–71) moreover was made duke of Modena and Reggio by the emperor Frederick III in 1452 and duke of Ferrara by Pope Paul II in 1471. Both Leonello and Borso managed to avoid war, but their half brother ERCOLE I (1431–1505) succeeded to the duchy in a difficult period. He had early to face a hostile coalition of Venice and Pope Sixtus IV, in which Venice gained possession of the Polesine (1484); and in 1501 he was obliged to contract a matrimonial alliance between his son Alfonso (*see* below) and Lucrezia Borgia in order to save his possessions from the rapacious hands of her father, Pope Alexander VI, and her brother Cesare. Even so, he continued his dynasty's patronage of the arts, taking the poet Boiardo as his minister and extending his favour and support to Ariosto. IPPOLITO D'ESTE (1479–1520), cardinal from 1493, was his third son.

Even more famous patrons of the great Italians of their time were Ercole I's daughters ISABELLA (1474–1539) and BEATRICE (1475–97). The former was married to Francesco Gonzaga, marquis of Mantua (*see* GONZAGA), the latter to Ludovico Sforza, regent and afterward duke of Milan (*see* SFORZA). Both seem to have been exceptionally gifted in taste and intelligence, both were carefully educated in letters and the arts, and both presided over lavish courts to which they attracted the leading artists and writers of their day. Mantegna and Leonardo da Vinci, Raphael and Bramante, Titian and Giulio Romano, Baldassare Castiglione, Matteo Bandello and others all enjoyed their patronage. By all accounts Beatrice was more amiable of character and more openhanded than her sister, but she exercised far less influence politically during her life, cut short by death in childbirth on Jan. 3, 1497. Isabella by contrast guided the policies of Mantua for more than 30 years throughout a difficult period of its history.

ALFONSO I (1476–1534), who succeeded his father Ercole in 1505, was an accomplished prince, famous in Europe for his skill in mechanics and artillery design. During his reign, however, papal ambitions of territorial expansion in Emilia became more threatening. By consistent adherence to the French interest in Italy, Alfonso came into collision with Pope Julius II and was deprived of Modena (1510) and of Reggio (1512). The Medici popes, Leo X and Clement VII, were both determined on the destruction of the Estensi, but the first was frustrated by death, the second by political weakness, and Alfonso was able to recover Reggio in 1523 and Modena in 1527. Alfonso was succeeded by his son, ERCOLE II (1508–59), who married Renée (*q.v.*), daughter of Louis XII of France. Both he and his brother IPPOLITO (1509–72), cardinal from 1539, were active patrons of literature and the arts. It was Ippolito who built the magnificent Villa d'Este at Tivoli. Next to rule was ALFONSO II (1533–97), the fifth and last duke of Ferrara, remembered for his patronage of Tasso. With him the main branch of the family came to an end. Despite three marriages he had no issue, and though at his death he bequeathed the duchy to his cousin CESARE (1562–1628), the pope, Clement VIII, refused to recognize the settlement, declared Cesare to be of illegitimate birth and, by the treaty of Faenza with Alfonso's sister LUCREZIA (1536–98), finally established direct papal rule in Ferrara (1598).

Cesare kept Modena and Reggio, but with him the Estensi ceased to play so important a part in Italian politics. The subsequent heads of the family were: Cesare's son ALFONSO III (1591–1644), who retired to a monastery in 1629; Alfonso III's son FRANCESCO I (1610–58), who commanded the French army in Italy in 1647; Francesco's elder son ALFONSO IV (1634–63), who fought in the French army against the Spaniards, founded the picture gallery of Modena and was the father of Maria Beatrice or Mary (*q.v.*) of Modena, consort of James II of Great Britain; Alfonso's son FRANCESCO II (1660–94), who originated the Este library, also at Modena, and founded the university; Francesco I's younger son RINALDO (1655–1737), through whose marriage with Charlotte Felicitas of Brunswick-Lüneburg the long-separated branches of the house of Este were reunited; and Rinaldo's son FRANCESCO III (1698–1780), who married a daughter of the French regent, Philippe, duc d'Orléans. Francesco III wished to remain neutral between Spain and Austria in the War of the Austrian Succession, but the Austrians having occupied and devastated his duchy, he took the Spanish side and was appointed captain-general, as nominal commander in chief of the Spanish army in Italy. He was re-established in his possessions by the treaty of Aix-la-Chapelle (1748) and, on being reconciled with the empress Maria Theresa, received from her the title of governor of Lombardy in 1754. With his son ERCOLE III (1727–1803), who was dispossessed of his duchy, a fief of the empire, under the Franco-Austrian treaty of Campo-Formio (1797) but was granted Breisgau in compensation after the treaty of Lunéville (1801), the male line of the Estensi came to an end.

Ercole III's only daughter MARIA BEATRICE (d. 1829), had been married in 1771 to the Austrian archduke Ferdinand, the empress Maria Theresa's third son, who was created duke of Breisgau in 1803. At his death in 1806 he was succeeded by his son FRANCIS IV (1779–1846), the first of the line of Austria-Este, to whom the duchy of Modena was given in 1814. Francis IV's son and successor FRANCIS V (1819–75) lost his possessions through the events of 1859–60, when Modena was annexed to Sardinia-Piedmont (*see* ITALY: *History*). On Francis V's death the Este titles passed by his will to the archduke Francis Ferdinand (*q.v.*), later heir to the Austro-Hungarian throne.

The children of Lady Augusta Murray, daughter of the 4th earl of Dunmore, by her marriage (1793) with Augustus Frederick, later duke of Sussex, sixth son of George III of Great Britain, assumed the old name d'Este because of their Welf ancestry and claimed recognition as members of the royal family; but as the marriage was in violation of the Royal Marriages act of 1772, it was declared invalid and their claims were set aside.

BIBLIOGRAPHY.—L. A. Muratori, *Delle antichità estensi ed italiane,* 2 vol. (1717–40); A. Frizzi, *Memorie per la storia di Ferrara,* 2nd ed., 5 vol. (1847–48); H. Bresslau, *Jahrbücher des deutschen Reichs unter Konrad II,* 2 vol. (1879–84); Julia Cartwright, *Beatrice d'Este* (1899) and *Isabella d'Este,* 2 vol. (1903); A. Solerti, *Ferrara e la corte estense nella seconda metà del secolo xvi* (1900); J. E. G. Gardner, *Dukes and Poets in Ferrara* (1904); G. Pardi, *Leonello d'Este* (1904); L. Simeoni, *Ricerche sulle origini della signoria estense a Modena* (1919); G. De Vergottini, "Ricerche sulle origini del vicariato apostolico," *Studi in onore di E. Besta* (1940). (P. J. J.)

ESTÉBANEZ CALDERÓN, SERAFÍN, called EL SOLITARIO (1799–1867), Spanish writer, the author of *Escenas Andaluzas* (1847), a record of local scenes and customs which is vivid despite a somewhat *recherché* style. His career combined politics, Arabic studies, poetry and the collecting of manuscripts. He died in Madrid, Feb. 5, 1867.

See *"El Solitario" y su tiempo,* 2 vol. (1883), by his nephew, A. Cánovas del Castillo.

ESTELÍ, a department in the west central highlands of Nicaragua. Pop. (1959 est.) 57,715; area 772 sq.mi. Despite rugged relief, the department, because of fairly fertile soils and favourable rainfall distribution, is important for the production of livestock, cotton, tobacco, sesame, cheese, vegetables and fruits. The departmental capital, Estelí (pop. 7,784), at 2,500 ft. elevation and once very isolated, is connected by the Inter-American highway with Managua, 94 mi. away. Also a branch highway to the west connects Estelí with El Sauce, the terminus of the railway to León and Corinto. (C. F. J.)

ESTERHÁZY OF GALÁNTHA, a noble Magyar family whose origin is traced to 1238. The first member of the family to emerge into history was FERENCZ ZERHÁZY (1563–94), deputy lord-lieutenant of the county of Pozsony (Bratislava), who took the name Esterházy when he became baron of Galántha, an estate acquired by the family in 1421. With his sons the family was divided into the lines of Fraknó (Forchtenstein; princes from

1687), Csesznek (counts) and Zólyom (Altsohl; counts).

The Line of Fraknó.—Ferencz's son Miklós Esterházy (1582–1645), born at Galántha on April 8, 1582, was the founder of the Fraknó line. With the political ideal of consolidating the power of the Habsburg dynasty as the means to freeing Hungary from the Turkish yoke, he opposed the two great Protestant champions of the period, Gabriel Bethlen and George Rákóczy I. His diplomacy, however, did much to bring about the treaty of Nikolsburg (1622) with the former and that of Linz (1645) with the latter. For his services he received many honours from the Holy Roman emperors Matthias and Ferdinand II. At the diet of Sopron (1625) he had been elected palatine of Hungary. He died at Nagy-Höflany (Gross-Höflein) on Sept. 11, 1645, leaving five sons. His writings were edited by F. Toldi (1852).

The Princes.—Pál Esterházy (1635–1713), the third son of Miklós, founded the princely branch of the family. He was born at Kis-Márton on Sept. 8, 1635. In 1663 he fought, together with Miklós Zrinyi, against the Turks and distinguished himself under Raimund Montecuccoli. He was appointed commander in chief of southern Hungary in 1667 and was elected palatine in 1681. He participated in the deliverance of Vienna from the Turks in 1683 and entered Buda at the head of 20,000 men in 1686. For his devotion to the Habsburgs he was created prince of the empire in 1687 (the right of succession being assured to the first-born sons of his house in 1712). He loved the arts and sciences, wrote several religious works and was one of the chief compilers of the *Trophaeum domus inclytae Estoratianae* for the glorification of his family. He died at Kis-Márton on March 26, 1713.

Prince Pál Antal (d. 1762), grandson of the first prince, was a distinguished soldier who rose to the rank of field marshal in 1758. He was succeeded by his brother Prince Miklós József (1714–1790), likewise a brilliant soldier and also a patron of the arts. This prince rebuilt the Esterházy castle in such a splendid Renaissance style that it came to be called the Hungarian Versailles; and he employed the composer Haydn for 30 years as the conductor of his private orchestra. He died in Vienna on Sept. 28, 1790. The emperor Joseph II had conferred the princely title on all Miklós József's descendants, male and female, in 1783.

Miklós József's grandson, Prince Miklós (1765–1833), was the founder at Vienna of a great collection of paintings and engravings (transferred in 1865 to the Hungarian academy in Budapest). He also erected a monument to Haydn at his summer place at Kis-Márton. When the French invaded Austria in 1797, he raised a regiment of 1,000 men at his own expense. In 1809, when Napoleon invited the Magyars to elect their own king to replace the Habsburgs, overtures were made to Prince Miklós, but he refused the honour and raised a regiment of volunteers in defense of Habsburg interests. He died at Como, in Italy, on Nov. 25, 1833.

His son, Prince Pál Antal (1786–1866), was a well-known diplomat. Born on March 10, 1786, he served during the Napoleonic Wars as secretary of the Austrian embassy in London in 1806 and in Paris under Metternich in 1807. Accredited to the Saxon court in 1810, he tried in vain to detach Saxony from Napoleon. He took part in the diplomatic negotiations of 1813–15, especially at the congress of Châtillon, and became ambassador to England at the conclusion of peace between the allies and France. He represented Austria at the coronation of Charles X of France in 1825 and at the London conferences of 1830–36. In 1848 he was minister of foreign affairs in the first responsible Hungarian cabinet, but resigned his post in September because he could see no way of reconciling the court with the nation. His last years were spent in obscurity. He died at Regensburg in Bavaria on May 21, 1866.

The Counts of the Fraknó Line.—The junior branch of the Esterházys of Fraknó, the members of which bear the title of count, was divided into three lines by the sons of the first prince's brother Ferencz Esterházy (1641–83).

The eldest of the latter's sons Count Antal (1676–1722), distinguished himself in the war against Ferencz Rákóczy II in 1703, joined Rákóczy next year and fought successfully against the Austrians in several battles until he was outnumbered and defeated in 1709. In 1710 he joined Rákóczy in Poland and accompanied him to France and Turkey. He died in exile at Rodosto on the shores of the Black sea in 1722. There is an edition of his campaign diary by K. Thaly (1901).

Count Bálint Miklós (1740–1806), whose father, Count Antal's son Bálint József, had entered the French service, became a member of Marie Antoinette's court and one of her favourites. He also stood in high favour with the comte d'Artois (later Charles X of France) and was raised to the rank of *maréchal de camp*. At the outbreak of the French Revolution he helped many royalists to emigrate. In 1791 he accompanied Artois to Vienna, from where he was sent to St. Petersburg to enlist the aid of Catherine II. He died at Grodek in Volhynia in 1806. There are editions of his *Mémoires* (1905) and of his *Lettres* (1906) in French, by E. Daudet. (T. K.)

ESTERS, in organic chemistry, compounds formed by the condensation of an alcohol and an acid, with elimination of water. Since this process was at one time considered analogous to the neutralization of an acid by a basic hydroxide in salt formation, these esters were formerly called ethereal salts. But this older term is misleading, since esters, unlike salts, are ordinarily not conductors of electricity in solution. In most polybasic acids, all the ionizable hydrogen atoms can be replaced in this way, and the compounds formed are known as neutral esters. If, however, some of the hydrogen of the acid remains undisplaced, then acid esters result. These acid esters retain some of the characteristic properties of the acids, forming, for example, salts, with bases. Esters may be prepared by heating the silver salt of an acid with an alkyl iodide; by heating the alcohols or alcoholates with an acid chloride or anhydride; by distilling the anhydrous sodium salt of an acid with a mixture of the alcohol and concentrated sulfuric acid; by heating for some hours on the water bath a mixture of an acid and an alcohol with a small quantity of hydrochloric or sulfuric acid; by treating an aqueous solution of the sodium salt of an acid with a dialkyl sulfate; or by treating an acid with an aliphatic diazo compound.

The esters of the aliphatic and aromatic acids are usually colourless neutral liquids, which are insoluble in water, but readily dissolve in alcohol and ether. Methyl oxalate, $C_2O_4(CH_3)_2$, is a solid (m.p. 54° C.), and numerous esters of high molecular weight are also solids. Many possess a fragrant odour and are prepared in large quantities for use as artificial fruit essences; for example, ethyl acetate, $CH_3.CO_2.C_2H_5$, butyl acetate, ethyl butyrate, $C_3H_7.CO_2C_2H_5$ (b.p. 121° C., odour of pineapple), isoamyl isovalerate, $C_4H_9.CO_2.C_5H_{11}$ (b.p. 196° C., odour of apples). The simpler esters hydrolyze readily when boiled with solutions of caustic alkalies or mineral acids, yielding the constituent acid and alcohol. When heated with ammonia, they yield acid amides (*q.v.*). With Grignard reagents (*q.v.*) they furnish tertiary alcohols, except in the case of alkyl formates, $H.CO_2.R$, where secondary alcohols are obtained.

N. Menschutkin (1879–82) examined the rate of esterification of many acids with alcohols and found that the normal primary alcohols were all esterified at about the same rate, the secondary alcohols more slowly than the primary, and the tertiary alcohols still more slowly. The nature of the acid used affected the result, for in a homologous series of acids it was found that as the molecule of the acid became more complex, the rate of esterification became less. The formation of an ester by the interaction of an acid with an alcohol is a reversible or balanced action; for as M. Berthelot and L. Péan de St. Gilles (1862) showed in the case of the formation of ethyl acetate (b.p. 75° C.) from ethyl alcohol and acetic acid, a point of equilibrium is reached beyond which the reacting system cannot pass, unless the system be disturbed in some way by the removal of one of the products of the reaction.

V. Meyer (1894) showed that in benzenoid compounds ortho-substituents hindered esterification of alcohols by acids in the presence of hydrochloric acid, this hindrance being particularly marked when two substituents are present in the ortho positions to the carboxyl group. In such a case the ester is best prepared by the action of an alkyl halide on the silver salt of the acid or of a dialkyl sulfate on the sodium salt, and when once prepared can be hydrolyzed only with great difficulty. The oils

and fats are largely the esters of the higher fatty acids and glycerol. The waxes, on the other hand, are largely esters of the higher fatty acids with higher monohydric alcohols. The esters of the higher fatty acids when distilled under atmospheric pressure are decomposed into an olefine and a fatty acid. (*See* OILS, FATS AND WAXES.)

Esters of the mineral acids may be prepared by the general methods described above. The neutral esters are as a rule insoluble in water and distill unchanged; on the other hand, the acid esters are generally soluble in water, are nonvolatile and form salts with bases.

Ethyl hydrogen sulfate (ethyl sulfuric acid), $C_2H_5.HSO_4$, obtained by the action of concentrated sulfuric acid on alcohol, is a colourless, oily liquid with acid reaction, slowly decomposing in aqueous solution. On heating, it forms diethyl sulfate and sulfuric acid.

Dimethyl sulfate, $(CH_3)_2SO_4$, a colourless liquid which boils at 187°–188° C., with partial decomposition, is used as a methylating agent. Great care should be taken in using dimethyl and diethyl sulfates, since the respiratory organs are seriously affected by the vapours.

Ethyl nitrate, $C_2H_5.ONO_2$, a colourless liquid which boils at 86.3° C., is prepared by the action of nitric acid on ethyl alcohol. Ethyl nitrite, $C_2H_5.ONO$, is a liquid which boils at 18° C.; the crude product obtained by distilling a mixture of alcohol, sulfuric and nitric acids and copper turnings is used in medicine under the name of sweet spirit of nitre. Amyl nitrite (*q.v.*), $C_5H_{11}.ONO$, boils at 96° C., and is employed in the preparation of anhydrous diazonium salts.

See also references under "Esters" in the Index volume.

ESTES PARK, a village at the eastern entrance to the Rocky Mountain National park (*q.v.*), in Larimer county, Colo., U.S., 63 mi. N.W. of Denver. It is located in beautiful Estes park, a level valley surrounded by mountains and lying at an elevation of 7,500 ft.

This area was visited by fur trappers in the early 1840s and was named for Joel Estes, first settler, who visited the region in 1859 and built his cabin there the next year.

The town of Estes Park is an important touring centre and summer resort. Trail Ridge road, highest continuous U.S. highway, goes through Rocky Mountain National park from Estes park to Grand lake.

The village is also headquarters for the Colorado-Big Thompson water diversion and power project. The population of the village is under 2,000. (L. R. HA.)

ESTHER, BOOK OF, a book of the Old Testament, one of the Megilloth, or rolls, group of the Hagiographa (the others being the Song of Solomon, Ruth, Lamentations and Ecclesiastes). Esther relates legendary events supposed to shed light on the origin and name of the festival Purim, the Feast of Lots (*see* JEWISH HOLIDAYS: *Semiholidays: Purim*). It is a romantic and patriotic tale, perhaps with some historical basis but with very little religious purport, and it may have been included in the Old Testament canon only for the sake of sanctioning the Purim celebrations. There is considerable evidence that the stories related in Esther actually originated among gentiles (Persian and Babylonian) rather than among the Jews. There is also reason to believe that the version given in the Greek Old Testament (the Septuagint) goes back to older sources than does the version given in the Hebrew Bible.

Contents.—Laying the scene at Susa, a residential city of the Persian kings, the book narrates that Haman, the vizier and favourite of King Ahasuerus (Xerxes I; reigned 486–465 B.C.), determined by lot the 13th of Adar as the day on which the Jews living in the Persian empire were to be slain. Esther, a beautiful Jewess whom the king had chosen as queen after repudiating Queen Vashti, and her cousin and foster father Mordecai, were able, however, to frustrate Haman's designs. His scheme to have Mordecai hanged resulted in Haman's being sent to the gallows erected for Mordecai, and the Jews throughout the empire were given permission to defend themselves on the day set for their extermination. The governors of the provinces learned in time that Mordecai, who had saved the king from being assassinated by two discontented courtiers, had succeeded to Haman's position as vizier; hence they supported the Jews in the fight against their enemies.

In the provinces, the Jews celebrated their victory on the following day, but at Susa where, at Esther's request, the king permitted them to continue the fight on the 14th of Adar, they rested and celebrated their success a day later. Thereupon Esther and Mordecai issued a decree obligating the Jews henceforth to commemorate these events on both the 14th and 15th of Adar.

Origins and Literary History.—Its theme and its language (which includes Aramaisms) characterize Esther as one of the latest books of the Old Testament. Nothing is known of its author. According to the postbiblical sources, its inclusion in the canon of sacred scriptures, as well as the observance of the feast of 14th and 15th Adar, as late as the 3rd century A.D. still met with strong opposition on the part of the Jewish authorities in Jerusalem; yet murals uncovered in the synagogue of Dura-Europos on the Euphrates prove that at that time the story of Esther was popular among the Jews of the Diaspora.

Purim.—In earlier times, the Purim festival does not seem to have been known under that name. II Maccabees calls its first day the Mardukian day, apparently defining it as a festival of the worshipers of the Babylonian god Marduk. Nor does the name Purim appear in the paraphrase of the Esther story found in book xi of Josephus' *Antiquities* (1st century A.D.). In distinction from this paraphrase, the Septuagint mentions the "casting of lots," but instead of the word *purim* ("lots") uses *phrurai,* which recurs as *phruraia* in Josephus. The term *phrurai* or *phruraia,* which is neither Aramaic nor Hebrew, seems to be an adaptation to Aramaic of the vocable *ferver,* known, *inter alia,* from the name Farvardigan, which denotes an Old Persian festival observed in ancient times in the same season as the Jewish festival of Purim. The fact that both II Maccabees and the Greek translations of Esther are unfamiliar with the term *purim* is hardly coincidental; indeed, the episode of the casting of lots is so insignificant that it is difficult to believe that Feast of Lots was the original designation of the feast.

Ishtar-Esther.—Susa, before becoming a residence of the Persian kings, was the capital of the kingdom of Elam. Inscriptions excavated at Susa prove, however, that around 600 B.C. the city was subject to the rule of the "Chaldean" kings of Babylonia, which implies that its population thenceforth included Babylonians and Aramaeans. The Babylonian goddess Ishtar figures frequently as "Ishtar, the Queen" and "Ishtar, the Bride" in the religious literature of the Babylonians. Istar or Esther being the Aramaic form of the name, the first of these designations means exactly the same as the Hebrew (and Greek) expressions for "Esther, the Queen." Similarly, another name of Esther, Hadassah (Aramaic), which occurs in the Hebrew text (ii, 7), goes back to *hadashatu,* a Babylonian word for "bride."

These data suggest that older legends reflected in the Book of Esther dealt not with a mortal queen of Persia but with one of the Babylonian deities worshiped at Susa. On the other hand, it appears that Vashti, the queen whom Esther superseded (ii, 1–18), reflects an Elamite goddess named Mashti. Hence it has been suggested that the introductory part of the Book of Esther goes ultimately back to a tale that described how Ishtar was elevated above other goddesses and became the queen of a divine king—a topic actually found in a Babylonian poem.

Marduk and Mithra.—Although the extant Book of Esther insists again and again on the Jewishness of Esther's cousin Mordecai, his very name characterizes him as a "worshiper of Marduk"; *i.e.,* a Babylonian. The Hebrew Bible provides the name of his antagonist, Haman, with the epithet "the Agagite" (Esth. iii, 1), a term intended to depict him as a hereditary enemy of the Jews. The Septuagint, on the other hand, calls him "a Bougacan" (Rest of Esther, xii, 6), the Greek form of an Aramaic term derived from a Persian designation of the god Mithra. Thus, it appears that the ancient tales used by the authors of the Greek narrative saw in Haman a worshiper of the Persian god Mithra.

The official recognition in the Persian empire of the Mithra

cult, which may well have resulted in clashes between its adherents and the worshipers of Marduk, took place during the reign of Artaxerxes II (404–359 B.C.). In characterizing Haman and Mordecai not as subjects of Xerxes (486–465) but of an Artaxerxes, the Septuagint seems therefore to be closer to the archetype of the Book of Esther than the Hebrew version is.

Rest of Esther.—The observation that the Esther legend reflects tales originally current among gentiles is corroborated by the conspicuous absence in the Hebrew version of any reference to Yahweh, the god of the Jews. The redactors of the version preserved in the Septuagint considered it necessary to interpolate many details which served the purpose of demonstrating Esther's and her cousin's religiousness. These are the so-called Additions to the Book of Esther, or Rest of Esther, which are not found in the Hebrew Bible and are included among the Apocrypha in the English Bible. They comprise: 17 verses preceding ch. i; 7 verses inserted in ch. iii; 11 verses and 2 chapters inserted between Esth. iv. 17 and v. 3 (the intervening verses of the Hebrew version being omitted in the Greek); 24 verses inserted in ch. viii; and 11 verses following Esth. x. 3, where the Hebrew version ends.

Bibliography.—L. B. Paton, *A Critical and Exegetical Commentary on the Book of Esther* (1916), "A Text-Critical Apparatus to the Book of Esther," *Old Testament and Semitic Studies in Memory of William Rayney Harper*, vol. ii (1908); Josephus, *Jewish Antiquities*, ix–xi, ed. by R. Marcus in the "Loeb Classical Library" (1937); J. Lewy, "The Feast of the 14th Day of Adar," *Hebrew Union College Annual*, 14:127–151 (1939); E. Bickerman, "The Colophon of the Greek Book of Esther," *Journal of Biblical Literature*, 63:339–362 (1944). (Ju. L.)

ESTIENNE (Étienne, latinized as Stephanus), the name of a distinguished family of French printers and humanists.

Henri Estienne (d. 1520), the founder of the family, is first recorded as working in Paris in 1502. In 1504 or 1505 he set up his own printing press in the rue Jean-de-Beauvais, near the university. On his death, in 1520, his three sons were still very young and the business was provisionally carried on for them by the foreman, Simon de Colines, who in 1521 married their widowed mother.

Robert Estienne (d. 1559), Henri's second son, had some hand in the printing of a Latin edition of the New Testament and the Psalms in 1522–23. His contract of marriage with Perrette, daughter of the respected Jodocus Badius, enabled him to become head of the family firm in 1526; and it was he who adopted the device of the olive tree for his title pages. In 1527–28 he published his first complete Bible in Latin; and in 1531 he completed his great *Dictionarium seu linguae latinae thesaurus*, which marks an epoch in the history of lexicography, not only for Latin but for all languages. Francis I of France made him king's printer for Hebrew and Latin works in 1539; and in 1540 he became *de facto* king's printer also for Greek, being commissioned in 1541 to supply the king's library with books printed in the Greek type of Claude Garamond. In the latter capacity he prepared the first printed editions of Eusebius (1544), Moschopulus (1545), Dionysius of Halicarnassus (1547), Alexander of Tralles (1548), Dio Cassius (1548), Justin Martyr (1551), Xiphilinus (1551) and Appian (1551; this appeared under his brother Charles' name) and the first edition of the Greek New Testament to contain a critical apparatus (1550). Of his Latin editions the early Virgil (1532) is noteworthy; and he also published grammars and educational texts.

Robert, however, became involved in the troubles of the Reformation. His relations with Jacques Lefèvre d'Étaples (Faber Stapulensis) and with Guillaume Farel indicate the liberal nature of his religious views and an earnestly critical attitude toward the Catholic dogmatism of the Sorbonne; and in Dec. 1547 the privy council of the new king Henry II, on representations from the Paris theologians, banned Estienne's series of Latin Bibles (1527, '32, '34, '45 and '46).

Estienne with some difficulty secured a guarantee of personal immunity from prosecution in 1548, but the continuance of the censorship was intolerable to him, and in 1549 he visited Geneva and conferred with Calvin. Next year, after careful preparations, he took up permanent residence in Geneva, where he embraced the doctrines of the Reformed Church.

Works printed by Estienne in Geneva include a Greek New Testament (1551), in which the present division of the text into verses occurs for the first time; a caustic *Ad censuras theologorum parisiensium responsio* (1552), in answer to the Sorbonne's condemnation of him; and several of Calvin's works. Received into full citizenship of Geneva in 1556, he died there on Sept. 7, 1559.

Charles Estienne (d. *c.* 1561), Robert's younger brother, not to be confused with Robert's son Charles (d. *c.* 1564), was educated for a medical career. For a time, however, he served as tutor to the future poet J. A. de Baïf and then, on Robert's flight to Geneva, succeeded to control of the family press in Paris and to the title of king's printer, as he remained a Catholic. Bankrupt in 1561, he is thought to have died in a debtors' prison.

Charles Estienne is more important as an author than as a printer. His writings cover a great variety of subjects: his "Épistre . . . en laquelle est déclairée la manière que les anciens ont observée en leur comédies" went through seven editions between 1542 (when it was attached to a translation of Terence's *Andria*) and 1578, and anticipates the renaissance of the drama in France; an anatomical work, *De dissectione partium corporis humani* (1545; French version 1546), is notable for its woodcut illustrations; *La Guide des chemins pour aller et venir par tous les pays et contrées du royaume de France* (1552; augmented ed. 1553; modern ed. by J. Bonnerot, 2 vol., 1936) is a pioneer work in the sphere of guidebooks.

The *Dictionarium historicum ac poëticum* (1553; augmented ed. 1561) went through many editions until it was superseded by the great French encyclopaedias of the following centuries; and *Praedium rusticum* (1554; augmented French version, *L'Agriculture et maison rustique*, 1564), a collection of tracts on husbandry, was also many times republished.

Henri Estienne (d. 1598), called the Great, was Robert's eldest son. Educated in classical literature, he traveled as a young man in Italy, England and Flanders, examining ancient manuscripts and visiting scholars, before joining his father in Geneva. There he began by publishing the results of his own researches in the first printed editions of several Greek texts, including a tract of Dionysius of Halicarnassus (1554), the works of the so-called Anacreon (1554), Athenagoras (1557), Maximus Tyrius (1557), the complete *Agamemnon* of Aeschylus (1557) and books xi–xx of Diodorus Siculus (1559; discovered by Estienne himself), as well as a Latin translation of the *Hypotyposes* of Sextus Empiricus (1559). In 1559 he succeeded to ownership of the press at Geneva under his father's will, which however bound him to keep it in that city and to remain in the Reformed religion. These conditions made him liable to much interference from the ecclesiastical authorities there.

The year 1566 saw three notable publications by Henri Estienne: the edition of *Poetae graeci principes heroici carminis*, which is admired both for its typography and for its scholarship; the *Traicté de la conformité du langage françois avec le grec* (this bears no date and may have appeared in 1565; modern ed. by L. Feugère, 1853) is the first of the author's works to assert the merits of his own language and may be compared with Joachim du Bellay's *Défense et illustration* (1549); and the Latin edition of Herodotus, with its "Apologia pro Herodoto." This "Apologia" was accompanied by a separate French version, *Apologie pour Hérodote* (full title *L'Introduction au traité de la conformité des merveilles anciennes avec les modernes, ou Traité préparatif à l'Apologie pour Hérodote*, 1566; modern ed. by P. Ristelhuber, 2 vol., 1879); and this, which is perhaps Estienne's most famous work, began his troubles in Geneva. Ostensibly designed to show how the strange stories in Herodotus are paralleled by equally strange ones in modern times, it is bitterly satirical of his own age; and some passages were most objectionable to the Genevan churchmen. Arrested and tried, Estienne was obliged to cancel the offending pages. Even so the book went through 12 editions in 16 years.

In the field of classical scholarship Estienne's output continued voluminously: his Greek and Latin text of Plutarch, 13 vol. (1572), may be cited as an example. He was, moreover, the first editor of classical texts to demonstrate the use of conjecture in the

restoration of a corrupt passage: but he resorted to conjecture only when all the manuscripts failed him. Estienne's greatest work, however, is his Greek dictionary, *Thesaurus graecae linguae*, 5 vol. (1572), a masterpiece and monument of lexicography of which new editions appeared as late as the 19th century.

With *Deux dialogues du nouveau langage françois, italianizé et autrement desguizé* (1578; modern ed. by P. Ristelhuber, 2 vol., 1885), a defense of pure French against italianizing innovations, Estienne displeased the Genevan authorities again. He therefore withdrew for a year to France, where he was very well received by the king, Henry III (now anxious to conciliate Protestant opinion); and his new book in praise of the French language, *Projet du livre intitulé de la Précellence du langage françois*, was printed in Paris in 1579. On his return to Geneva (1580) Estienne was sentenced to six days' imprisonment.

After 1583 Estienne spent all the time that he could away from his domicile, wandering from city to city in search of a congenial patron. The later publications of his press thus suffered to some extent from neglect. Of his own writings in this period one book deserves special mention: *Les Prémices ou le I livre des Proverbes épigrammatizez ou des Épigrammes proverbializez* (1594), provoked by Henry III's skepticism about the statements on proverbs in the work of 1579.

In 1597 Estienne left Geneva to visit his son-in-law, Isaac Casaubon, at Montpellier, France. From there he went to Lyons, where, at the end of Jan. 1598 he died.

Little need be said of the other members of the family. Henri's son PAUL (b. 1566) succeeded him in charge of the press in Geneva but had to leave the city because of his treasonable dealings with Savoy; he disappeared from history after his sale of the press in 1627. Paul's son ANTOINE (1592–1674), having abjured Calvinism, became king's printer in France in 1614 but was financially ruined in 1664 and died blind in a hospital, his son, HENRI (1631–61), also a printer, having predeceased him. The great Henry's brother ROBERT (d. 1571), who had remained a Catholic when his father, the first Robert, fled to Geneva, became king's printer in France in succession to his uncle Charles; his son, the third ROBERT (1560–1630) was appointed king's interpreter for Greek and Latin. FRANÇOIS, the third son of the first Robert, worked a press of his own in Geneva from 1562 to 1582; his family's activity, however, ended with his grandchildren.

BIBLIOGRAPHY.—A. A. Renouard, *Annales de l'imprimerie des Estienne*, 2nd ed. (1843); H. Stein, "Nouveaux documents sur les Estienne, imprimeurs parisiens, 1517–1665," *Mémoires de la Société de l'histoire de Paris*, vol. xxii (1895); Elizabeth Armstrong, *Robert Estienne, Royal Printer* (1954); H. W. Lawton, *Charles Estienne et le théâtre* (1928); Sir Herbert Fordham, *The Earlier French Itineraries* (1921); L. Clément, *Henri Estienne et son oeuvre française* (1899); and general histories of classical scholarship.

ESTIVATION: *see* HIBERNATION AND ESTIVATION.

ESTON, an urban district in the North Riding of Yorkshire, Eng., on the Tees, 4 mi. E.S.E. of Middlesbrough. Pop. (1961) 37,160. It is situated at an altitude of between 100 and 150 ft. at the foot of the northwestward-facing scarp of Eston Beacon (800 ft.). The chief industries are iron and steel production, shipbuilding and chemicals.

The Cleveland ironstones, of which the main seam was found in 1850, were nearly exhausted a century later. Numerous prehistoric earthworks fringe the hills to the south. Marton-in-Cleveland, 3 mi. S.W. of Eston, was the birthplace of Capt. James Cook (1728).

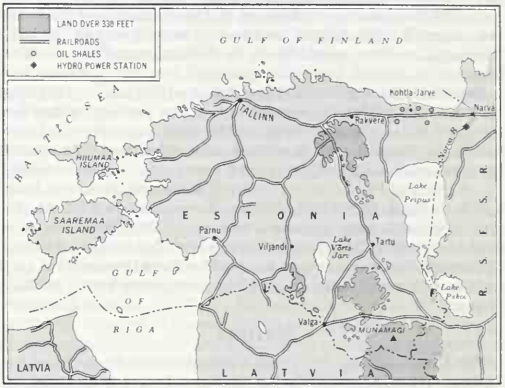
ESTONIA (Estonian EESTI; Russian ESTONIYA), since 1940 the Estonian Soviet Socialist Republic (Estonskaya Sovetskaya Sotsialisticheskaya Respublika), is the smallest of the three Baltic republics (Lithuania, Latvia, Estonia) of the U.S.S.R. It is 17,413 sq.mi. in area and has (1959) 1,196,791 inhabitants, with a density of 69 persons per square mile. Administratively it is divided into 39 *rayons;* the capital is Tallinn.

Estonia is bounded on the north by the Gulf of Finland, on the west by the Baltic sea. Its eastern boundary, with the Russian S.F.S.R., follows Lakes Pskov (Pihka Jarv) and Peipus (Peipsi or Chudskoye) and the Narva river, but the southern boundary, with Latvia, is not marked by any physical obstacle.

PHYSICAL GEOGRAPHY

Estonia consists of a low plateau of Cambrian and Silurian rocks, mainly limestones, flanked in the south by Devonian, mainly sandstones. The northern limit of the plateau is marked by a cliff which in places reaches the coast. In the west, the plateau continues in a number of Baltic islands, the largest of which are Hiiumaa (Khiuma) and Saaremaa. Almost the whole of Estonia is covered by glacial deposits of varying thickness, deepest where there are relics of terminal moraines. Even so most of the republic is below 300 ft. above sea level. This altitude is exceeded only in two main areas, while the highest point in Estonia reaches only 1,040 ft. (Munamagi).

Formed mainly from glacial materials, the soils are very varied,



ESTONIA, SHOWING RELIEF, RAILROADS, OIL SHALES AND HYDRO POWER STATION

ranging from coarse sands to clays, while the frequent occurrence of boulders is a hindrance to cultivation. In the north, where Silurian limestone underlying glacial deposits influences the make-up of the soil, lime is not deficient, but over most of the republic soils are podsolized and generally poor.

Climate.—For its latitude (approximately 57° 30′–59° 30′ N.) the climate of Estonia is mild. The mean temperature for the coldest month varies from 26° F. on Saaremaa to 19° F. in the southeast of the republic, while the July temperature is 60° along the coasts but reaches 63° inland. In spite of its small size there is thus a considerably greater range of temperature inland than along the west coast. This is reflected also in that ice cover lasts 130–145 days along the northeast coast but only 50–70 days in the west. Mean annual precipitation varies between 22 in. and 28 in. (Tallinn and Tartu each 26 in.), and falls mostly in summer (May–September).

Vegetation.—About 28% of the total area of Estonia is forested, but one-third of this is of poor quality and little of the original forest remains. The vegetation is mixed coniferous and broadleaf; the species of most frequent occurrence are the pine (*Pinus silvestris*), the spruce (*Picea excelsa*), the birch and the aspen. About 15% of the republic is marshy, and another 15% is sown land.

Animal Life.—Because of the long occupation of Estonia by man, animal life is restricted mainly to the smaller mammals, but elks, bears, roe deer, wolves and wild boars do occur, as well as hares and badgers. Birds include the wood grouse and hazel grouse, while seals, cod, plaice, salmon and eels are among the fauna of the coast. (J. P. Co.)

THE PEOPLE

The Estonians belong to the Finno-Ugrian family of peoples and constitute the core of the southern branch of the Baltic Finns, the other constituents of that group being the Livs and Votes. The Livs occupy the northern tip of Courland and number fewer than 1,000. The Votes are still less numerous and live in a few

villages in the vicinity of Narva. Formerly the Votes were said to have inhabited the whole of Ingermanland (Ingriya), whereas the expansion of the Livs along the Baltic shores to the south reached the Kurisches Haff (Kurskl Zaliv) toward the end of the 9th century. They were later absorbed by the Letts.

The Baltic Finns reached the Baltic about the beginning of the Christian era, coming from the great arc of the Volga, the area of the Fatianovo culture, which dates from the first half of the 2nd millennium B.C. On the shores of the Baltic they must have met a numerous Early German population (the Aesti of Tacitus) and it is possible that the Swedes of the island of Ruhnu (Runo) are the remnant of these ancient peoples. The Estonians and Livs intermarried with the Early Germans and also with the Lapps, but the extent of the admixture is not known. Culturally the Estonians were strongly influenced by the Germans and traces of the original Finnish culture have been preserved only in folklore. The majority of the people belong to the Lutheran Church. (For language *see* FINNO-UGRIC LANGUAGES.) (J. Cz.)

Population.—Of the total population of Estonia in 1959 (1,196,791), 56% (674,515) was urban, a higher percentage than for the U.S.S.R. as a whole (48%). The rural population is fairly evenly distributed, but the larger towns are either located on or near the north coast, such as Tallinn and Narva (*qq.v.*), or in the south, *e.g.*, Tartu and Parnu (*qq.v.*). Only Tallinn and Tartu (74,300) had more than 50,000 inhabitants. Approximately 73% of the total population were Estonians and nearly 22% Russians, who lived mainly in the larger towns. The remainder comprised Finns, Ukrainians, Belorussians and Jews. For the administration and social conditions of Estonia *see* UNION OF SOVIET SOCIALIST REPUBLICS. (J. P. Co.)

HISTORY

The Estonians (see *The People*, above) are first mentioned by Tacitus (1st century A.D.) in his *Germania* in which he writes of the keepers and tillers of the soil. Their political system was patriarchal, based on clans headed by elders. The first invaders of the country were Vikings (*q.v.*) who from the mid-9th century A.D. passed through Estonia and Latvia on their way to the Slavonic hinterland. These Vikings were soldiers and merchants, and there is much evidence to show that the Estonians learned from them: apart from ornaments, many of local make, archaeological finds in Estonia include Arabic, Byzantine, German and Anglo-Saxon coins as evidence of trade. In the 11th and 12th centuries, the Danes and the Swedes tried to christianize the Estonians, without success. Between 1030 and 1192 the Russians made 13 campaigns into Estonia but failed to establish supremacy there.

The German Conquest.—Meinhard, a monk from Holstein, landed in 1180 on what is now the Latvian coast and for 16 years preached Christianity to the Livs, a Finno-Ugrian tribe (*see* LIVONIA). His successor, Berthold of Hanover, appointed bishop of Livonia, decided that the sword must be used against the recalcitrant pagans. He was killed in 1198, in the first great battle. Albert of Buxhoevden, who succeeded him as bishop, proved himself a shrewd colonizer, pacifying the "treacherous Livs" and forcing them to build the fortress of Riga. To popularize recruitment for his army, Albert dedicated Livonia to the Virgin Mary. In 1202 he established the Order of the Knights of the Sword.

By 1208 the knights were firmly established on both banks of the Daugava (Zapadnaya), and Albert felt strong enough to proceed northward to the conquest of Estonia. In the following years the Estonians lost steadily in manpower while the knights replenished theirs with new crusaders from Germany. The Russian princes of Novgorod and Pskov also raided Estonia on many occasions, penetrating especially deeply in 1212 and 1216. All through the war the defense was aggressive, with raids as far as Riga (1215) and Pskov (pillaged in 1212 and 1217). The Estonian forces were still based on local levies, though these co-operated and helped one another to a large extent. Then, in a major battle in 1217, the knights defeated the Estonians and killed their commander Lembitu. Northern Estonia and the islands, however, remained free for another ten years. To complete the conquest, Albert concluded an alliance with King Valdemar II of Denmark, who in 1219 landed with a strong army on the northern coast, on the site of Tallinn. Moreover, in the summer of 1220 Swedes conquered part of western Estonia, but the garrison that they left there was wiped out shortly afterward. The Danes landed on the island of Saaremaa two years later but after their castle was bombarded with catapults, they surrendered on a safe conduct. Prompted by this victory fighting flared up all over Estonia and the Estonians raided as far as Sweden in 1226. By 1227 the conquest was complete.

In 1237 the Knights of the Sword joined the Teutonic Order (*q.v.*), which now assumed control of Livonia. Northern Estonia and the islands were under Danish rule; Livonia, that is, southern Estonia and Latvia, was shared between the Teutonic Order and the bishops. The terms under which the Estonian localities had submitted were not severe, but the conquerors violated them as their position became stronger, thus provoking a series of revolts. After major risings in 1343–45, the Danish crown sold its sovereignty over northern Estonia to the Teutonic Order in 1346. The Germans were now the masters in the "Land of the Virgin" and, with minor exceptions, formulated its history in colonization, commerce and the church for centuries to come. The Estonians, the Latvians and the Livs became the serfs of their conquerors, with little to sustain their national feeling save their folklore and their traditional crafts. In agriculture, however, the Germans made an improvement by introducing the three-field instead of the two-field system, and the towns enjoyed prosperity through the commercial activity of the Hanseatic League.

The Swedish Period.—By the end of the 15th century two major powers were emerging around Livonia: Poland-Lithuania already united in the south; and Muscovy, which had conquered Novgorod, in the east. A third factor, the Reformation, was to produce disunity and strife in this quarter as in many others. More by diplomacy than by victory in battle, Poland gained Livonia on the dissolution of the Teutonic Order in 1561. Three years before, northern Estonia had capitulated to the king of Sweden. The Muscovite tsar Ivan IV the Terrible had captured Narva in 1558 and penetrated deep into Estonia, bringing devastation with him, and it was not until 1581 that the Russians were expelled by the Swedes. In 1559 the bishop of Saaremaa had sold the Estonian islands to Denmark, but in 1645 they became part of the Swedish province. By the truce of Altmark (1629), which ended the first Polish-Swedish War, Poland surrendered to Sweden the major part of Livonia, so that all Estonian lands then came under Swedish rule.

Prolonged wars left the country devastated, and many farms were unoccupied. The vacancies were partly filled by foreign settlers who were soon assimilated. This also gave the German nobility the opportunity to enlarge their estates, to increase taxes and to exact more unpaid labour. Since a great part of the peasantry could not pay these taxes they became legally bound to work the land (*glebae adscripti*). The Swedish kings attempted to curb the power of the nobility and improve the lot of the peasants. Soon after Charles XI of Sweden came of age (1672) the nobles of Livonia were forced to show their title deeds, and those who failed to do so became tenants of the crown. A land survey was carried out, the yield and value of the peasants' holdings were registered for the first time, and the peasants were accorded the right to complain against their masters in the law courts. At the foundation of the university by Gustavus II Adolphus of Sweden in Tartu (1632) it was ruled that the "Academia Gustaviana, this temple of Protestant learning at the gates of the barbarous East," was to be open also to the peasants' sons of Estonia and Latvia. It was also in this period that the first books and grammars in Estonian were printed, and by the end of the Swedish era every parish had a few schools. The towns, however, did not seem to regain the prosperity that the Hanseatic merchants had brought.

The Russian Conquest.—The "good old Swedish days" for Estonia were more a legend than reality, but they ended with the Great Northern War. The Russian tsar Peter I was finally able to achieve the dream of his predecessors and conquer the Baltic provinces. After the defeat of Charles XII of Sweden at Poltava (1709), Russian armies seized Livonia. The barons did not re-

sist, angered as they were against the Swedish crown for its policy of reversion of estates. By the peace of Nystad in 1721 Sweden ceded to Russia all its Baltic provinces. The old Land of the Virgin was divided into the three *gubernyas* of Estonia, Livonia and Kurlandia (Courland). In 1740 a famous lawsuit by an Estonian peasant, Jaan the Miller from Vohnja, against his landlord ended in a decision by the College of Justice in St. Petersburg that the peasants had no right to sue their landlords. The peasants' lot became worse than ever. In 1804, however, under the emperor Alexander I, the peasants of Livonia were given the right of private property and inheritance; a bill abolishing serfdom was passed in Estonia in 1811 and in Livonia in 1819. Other agrarian laws followed, in particular that of 1863 establishing the peasants' right of free movement, that of 1866 abolishing the landowners' right of jurisdiction on their estates, including the right to flog and that of 1868 abolishing the *corvée*.

The Estonian National Awakening.—The Estonian peasants benefited by these reforms and at the end of the 19th century they possessed two-fifths of the privately owned land of the country. With the growth of urban prosperity as a result of industrialization the population increased. Improvement in education was such that by 1886 only about 2% of the Estonian recruits were unable to read. National consciousness increased, too. In 1838, at Tartu, F. R. Faehlmann organized the Estonian Learned society which decided to collect Estonian narrative folk songs into a national epic poem, as the Finns had done with their *Kalevala*. This was carried out by F. R. Kreutzwald, whose *Kalevipoeg* was printed in 1857–61. In 1857 J. W. Jannsen founded at Parnu the first Estonian newspaper, *Parnu Postimees*. The nation was stirred by the first national singing festival held in Tartu in 1869. Such festivals developed into a lasting tradition that played a most notable part in the national awakening. In 1878 C. R. Jakobson founded at Viljandi the periodical *Sakala*, in which he outlined a program of political reform, claiming for the Estonians absolute equality with the Germans and the inclusion of the Estonian-speaking districts of Livonia in the province of Estonia.

The accession of Alexander III marked the beginning of a period of more rigid russification. The Russian municipal constitution was introduced in 1882; Russian criminal and civil codes replaced the old Baltic ones; in 1887 Russian was made the language of instruction, instead of German and Estonian. In 1893 the University of Tartu (Dorpat), which was then an important centre of German learning, was russified. The first reaction of the Estonians was that poetic justice was being administered to their agelong oppressors, but they also feared the reactionary Pan-Slavism. The newspaper *Postimees*, transferred to Tartu in 1864, continued to be published there, and from 1896 its editor was Jaan Tonisson (*q.v.*), a Liberal intellectual. In 1901, in Tallinn, Konstantin Pats (*q.v.*) founded the newspaper *Teataja*, in which moderately radical ideas were expressed. In 1904, thanks to Pats, the Estonians won a clear victory on the Tallinn town council.

In Jan. 1905 a revolution started in Russia and spread immediately to Estonia. Tonisson founded a National Liberal party and organized its first congress in Tallinn on Nov. 27. The 800 delegates soon split into a Liberal and a Radical wing. Both voted resolutions demanding political autonomy for Estonia, with an administrative merger of all areas with an Estonian ethnic majority, abolition of all privileges of the nobility and distribution of crown lands to the peasants. But while Tonisson warned the nation against the use of violence, Jaan Teemant (*q.v.*), the Radical leader, called for the people to be armed and for expropriation without compensation of all baronial land. In December, Pats summoned a peasant congress in Tallinn, but martial law was proclaimed. Parties of workers scattered into the countryside and began to loot and burn the manor houses. Troops were drafted and repression started: 328 Estonians were shot or were hanged. Pats and Teemant fled abroad, both being sentenced to death *in contumaciam*. Tonisson's party was the only one to survive the revolution. At the elections to the first and the second Russian duma the Estonians returned five deputies.

The Birth of Independence.—The second Russian Revolution of March 1917 brought autonomy to Estonia. On April 12 the Russian provisional government passed a decree providing that all districts with an Estonian majority were to be united into one province; Jaan Poska, the National Liberal mayor of Tallinn, was appointed commissioner for Estonia and an Estonian national council was planned. Elections to the council, which came to be known as the diet (*maapaev*), took place on July 7–8. It met on July 14 and on Oct. 12 appointed a provisional government with Pats as premier.

The November *coup d'état* which brought the Communists into power in Petrograd made itself felt in Estonia. On Nov. 28, 1917, the Estonian diet decided to break away from the Russian state, but on Dec. 8 the Russian council of people's commissars appointed for Estonia a puppet Communist government headed by Jaan Anvelt, who seized power in Tallinn, but never obtained control of the whole country. (He was shot in Moscow in 1937.) In Feb. 1918 German forces advanced to Estonia. The Communists fled from Tallinn, and on Feb. 24 the Estonian provisional government formally declared the independence of Estonia. The following day German troops entered Tallinn (Reval). Estonian leaders, except Pats, who was arrested, and Juri Vilms, who was shot by the Germans, went abroad or underground. On March 3 the Brest-Litovsk treaty was signed and by a protocol signed in Berlin on Aug. 27 the sovereignty over the Baltic countries was transferred from Russia to Germany.

Germany, however, lost World War I and capitulated on Nov. 11, 1918. The Estonian provisional government met in Tallinn on the same day and once more proclaimed the independence of Estonia. Two days later the Soviet government declared the treaty of Brest-Litovsk null and void. Pats, released from internment, reassumed his post. On Nov. 28 the Red army took Narva and started the invasion of Estonia, which was denuded of all arms by the retreating Germans. The government ordered general mobilization. Finland sent guns and rifles, while on Dec. 12 a British naval squadron under Rear Adm. E. A. Sinclair arrived in Tallinn and put ashore 1,000 light machine guns. On Dec. 23 Col. Johan Laidoner (*q.v.*) was appointed commander in chief of the Estonian forces. A few days later Sinclair handed over to him two Soviet destroyers captured near Tallinn. In Jan. 1919 Laidoner started a counteroffensive, helped by 2,000 Finnish volunteers. On Feb. 24, 1919, the first anniversary of the republic, Laidoner was able to report that the invaders had been driven out of the national territory. The war, however, continued, because of new Communist offensives and also because of the need to help Latvia in its two-front fight against the Red army and a German force commanded by Rüdiger von der Goltz. In their march against Estonia the Germans were beaten (June 21–23) and pursued to the gate of Riga, where an armistice was signed.

On Feb. 2, 1920, the peace treaty with the Soviet Union was signed at Tartu. Russia "voluntarily and forever" renounced its sovereign rights over the territory and people of Estonia. The war of liberation had cost the Estonian army 2,236 dead. At its peak in Dec. 1919 the army numbered 86,600 men.

On June 15, 1920, the constituent assembly (elected in April 1919), with August Rei as president, voted the new constitution with a single-chamber parliament (*riigikogu*) of 100 members elected for three years, with a system of proportional representation, and a chief of state (*riigivanem*) who was also the premier. As no party had an absolute majority, government by coalition became the rule, and from May 1919, when the first constitutional cabinet was formed, to May 1933 Estonia had 20 coalitions headed by ten statesmen, of whom Pats, Tonisson and Teemant each held the position of *riigivanem* four times.

On Dec. 1, 1924, 300 conspirators, mostly Russians working on the transit base at Tallinn or smuggled in, tried to seize communications and to call in Soviet troops, but failed ignominiously. The Communist party was outlawed and the movement became virtually extinct. The world depression of the early 1930s caused the falling off of agricultural prices and unemployment. The strong government action necessary to cope with the situation was precluded under the 1920 constitution. Two amendments to the constitution were rejected by referendums of 1932 and June 1933. A third campaign for the revision was begun in 1931 by the

Vaps (Vabadussojalaste Liit, or Freedom Fighters' league) under the leadership of Artur Sirk. When the Vaps proposal was presented to the country in a referendum in Oct. 1933, it was adopted by 416,879 votes (56.3% of the electorate). The new constitution gave sweeping powers to the president. Pats became acting president and was expected to prepare the ground for the first presidential election. Rather than risk the success of a Vaps nominee, he proclaimed on March 12, 1934, a state of emergency; the Vaps was dissolved, its leaders were arrested and the political activities of all parties forbidden. Laidoner assumed responsibility for public order. For three years Pats ruled as a benevolent dictator. In Dec. 1936 a new constituent assembly was elected. It prepared a third constitution with a chamber of 80 deputies elected by the majority system and a national council of 40 members. The election was held in Feb. 1938; in the lower chamber there were 63 members of the Patriotic League founded by Pats in Feb. 1935, and an opposition of 17 led by Tonisson and Teemant. In April Pats was elected president for a term of six years.

Independence Lost.—The fate of Estonia was decided by the so-called nonaggression treaty of Aug. 1939 between Nazi Germany and the U.S.S.R. A secret protocol to this treaty assigned Finland, Estonia, Latvia and eastern Poland to the Soviet orbit. After the defeat of Poland this arrangement was revised on Sept. 28 and a secret supplementary clause extended the Soviet sphere of influence to Lithuania. On the same day the Soviet government imposed on Estonia a treaty of mutual assistance, which reaffirmed the Tartu peace treaty and also the Soviet-Estonian nonaggression treaty of May 4, 1932, but conceded to the U.S.S.R. several military bases on Estonian territory which were manned forthwith. A broadly based nonpolitical government under Juri Uluots was appointed. On June 16, 1940, a Soviet ultimatum demanded a new Estonian government "able and willing to secure the honest application of the Soviet-Estonia mutual assistance treaty." The following day Soviet armed forces occupied the whole country, and on June 19 Andrei Zhdanov, a member of the Kremlin Politburo, arrived in Tallinn to complete the designs of the Soviet government. On June 21 a puppet government headed by Johannes Vares took over. On July 14–15 the "election" of a new chamber of deputies took place under a new procedure ignoring the provisions of the constitution and electoral law. Only the single list of Soviet-sponsored candidates was allowed to stand. On July 21 the new chamber was presented with a resolution to join the U.S.S.R.; it was unanimously adopted the following day in spite of being contrary to constitutional procedure. On Aug. 6 the Moscow supreme soviet incorporated Estonia in the U.S.S.R. Meanwhile Pats, Laidoner and many political leaders were arrested and deported to Russia. In the first 12 months of Soviet occupation more than 60,000 people of all classes and ages were killed or deported: more than 10,000 were removed in a mass deportation in the night of June 13–14, 1941. All the Soviet moves in the Baltic countries from 1939 were synchronized, sometimes to the hour.

On June 22, 1941, Germany attacked the U.S.S.R. (*see* WORLD WAR II). Large areas of Estonia were freed from Soviet forces by improvised Estonian units before the German front reached Estonia. For three years Estonia was under German occupation. It became part of the Ostland province. The Germans were in no hurry to restore self-government or expropriated property, let alone independence. By Feb. 1944, however, the Russians were back on the Narva front. A few thousand Estonian youths who had escaped to Finland to fight the Reds in the Finnish army rather than with the Germans returned, but of course could not save the situation. About 30,000 Estonians escaped by sea to Sweden and 33,000 to Germany; many thousands perished on the sea. On Sept. 22, 1944, Soviet troops took Tallinn. It is estimated that about 20,000 Estonians were deported in 1945–56. The third large deportation took place on March 24–27, 1949, and comprised about 40,000 persons, mostly farmers who resisted collectivization.

After restoration of the Soviet regime, Vares became the first chairman of the presidium of the Estonian supreme soviet, but he committed suicide in 1946 and was succeeded in turn by E. N. Pall, A. M. Jakobson (1950), J. H. Eichfeld (1958) and A. A. Muurisepp (1961). A. T. Veimer, chairman of the council of ministers, was dismissed in 1951 and succeeded by Muurisepp, replaced in turn, in 1961, by V. I. Klausson. J. G. Kabin was from 1950 first secretary of the Communist party.

Nominally the 15th republic of the U.S.S.R. (in the early 1960s still unrecognized as such by many governments, including the United States and Great Britain, but accorded *de facto* recognition by some governments), Estonia is subjected to the laws and practices of the union.

The Communist party (more than 40,000 members and candidates in the early 1960s), as well as the government, administration, down to collective farms, have been permeated by ethnic Russians or by Russians of Estonian extraction.

(K. SM.; E. AR.)

THE ECONOMY

Estonia was predominantly agricultural until 1945, but since then industrial production has been increasing fast, a trend reflected in the rapid increase of urban population (350,000 in 1939 to 674,000 in 1959) and the accompanying decrease in rural population (697,000 to 522,000). Industry, which includes forestry and fishing as well as mining and manufacturing, accounts for more than 80% of the value of production in Estonia, but this is no guide to the numbers employed in different activities, and agriculture is probably still the largest employer of labour. After World War II there was progressive collectivization of agriculture on the Soviet model.

About 1,830,000 ac. (15% of the total area) was sown land. About one-third (690,000 ac.) of this was under cereals, with rye the principal crop and wheat very restricted in acreage. Nearly half (850,000 ac.) was under fodder crops, while 230,000 ac. were devoted to potatoes. Limited quantities of flax and vegetables were also grown. As suggested by the large acreage under fodder crops, the emphasis in agriculture is on livestock farming. The livestock population of Estonia in the early 1960s exceeded 1,000,000 (40% cattle, 39% pigs, 21% sheep). Cattle are of outstanding importance and there is a greater emphasis on dairying than in the U.S.S.R. as a whole (63% of all cattle, against 47% in the U.S.S.R.). Estonia has a surplus of dairy products and meat.

The oil shales of the northeast are the principal source of energy in Estonia and in the early 1960s production averaged about 9,000,000 tons annually. These were roughly equivalent in heating value, when refined into liquid fuel and gas, to 3,000,000 tons of coal. The shales were at first heavily exploited to serve Leningrad to which a gas pipeline was built from Kohtla-Jarve in 1948, but later were consumed locally as well. The shales are supplemented by limited quantities of peat, wood and hydroelectricity (Narva power station).

Engineering is the chief branch of manufacturing. Shipbuild-

CAMERA PRESS LTD.—PIX

RESIDENTIAL SECTION OF TALLINN WITH DOMES OF THE ORTHODOX CATHEDRAL IN THE BACKGROUND

ing and electrical engineering are of all-Union importance, and precision and agricultural engineering are also represented. The chief centres are Tallinn, Tartu and Parnu. Cotton textiles are produced in Tallinn and Narva, linen in towns in the south. Industry is geared to the Soviet economic planning and trade is almost entirely internal within the U.S.S.R.

By Soviet standards Estonia has a good network of railways (860 mi. but mostly single track) and roads (6,400 mi. with a hard surface). Estonian shipping was nationalized by the U.S.S.R. and the chief ports are Tallinn and Parnu. There are air communications with the Soviet Union. *See* also references under "Estonia" in the Index volume. (J. P. Co.)

BIBLIOGRAPHY.—E. Uustalu, *The History of Estonian People* (1952); J. Hampden Jackson, *Estonia*, 2nd ed. (1948); V. Raud, *Estonia* (1953); A. Kaelas, *Das Sowjetisch besetzte Estland* (1958); O. Loorits, *Grundzüge des Estnischen Volksglaubens*, 3 vol. (1949–60); S. Kuri, *Estonia: a Selected Bibliography* (1958); G. N. Cherdantsev *et al.*, *Ekonomicheskaya geografiya S.S.S.R.*, pp. 144–154 (1957); M. I. Rostovtsev and V. Yu. Tarmisto, *Estonskaya S.S.R.* (1957); Tsentralnoye Statisticheskoye Upravleniye pri Sovete Ministrov S.S.S.R., *Narodnoe khozyaystvo S.S.S.R. v 1958 godu. Statisticheskiy yezhegodnik* (1959). (J. P. Co.; K. Sm.; E. Ar.)

ESTORIL, a fashionable seaside resort in Estremadura province, central Portugal, faces south onto the Tagus estuary, 25 km. (15½ mi.) W. of Lisbon by road. Pop. (1961 est.) 40,000. The chief feature of the town is the magnificent avenue of palm trees leading from the sea front to the casino. There are many luxury hotels and the usual tourist attractions. Estoril is noted for its mild climate and is both a winter and a summer resort. It is situated on the railway from Lisbon to Cascais. The nearby Monte Estoril is situated on the slopes of the hill of the same name, 358 ft. high. There are fine views over the bay. The hill is noted for its trees, especially pines, eucalyptus and palms. São João de Estoril is another nearby resort.

ESTOURNELLES DE CONSTANT, PAUL BALLUAT, BARON D' (1852–1924), French diplomat and parliamentarian who devoted most of his life to the cause of international co-operation and peace and in 1909 was awarded the Nobel peace prize for his work, was born on Nov. 22, 1852, at La Flèche, Sarthe. In the French diplomatic service he reached the rank of minister plenipotentiary. Later he entered politics and was elected deputy for the Sarthe *département* in 1895 and re-elected in 1898 and 1902. From 1904 he sat in the senate and was re-elected in 1909 and 1920. He took an active part in the interparliamentary international conferences which began in 1889 and was a member of the French delegation to The Hague peace conference of 1899, which led to the formation of the Court of Arbitration at The Hague. When that court seemed likely to be ignored by governments he persuaded Pres. Theodore Roosevelt to submit to it a minor dispute between the United States and Mexico in Sept. 1902. This example was then followed by other governments. In 1905 he set up in Paris the Association for International Conciliation, with branches in many other countries. He was active in securing the summoning of the second Hague conference in 1907. His contribution to internationalism lay chiefly in his tireless work as propagandist and organizer. He died at Bordeaux on May 15, 1924. His publications include *La Politique française en Tunisie* (1891); *La Conciliation internationale* (1906); *Le Rapprochement franco-allemand* (1909); *Pour l'aviation*, 2nd ed. (1909); *Les États-Unis d'Amérique* (1913; Eng. trans., *America and Her Problems*, 1915); and *Pour la Société des Nations* (1921). (D. Tn.)

ESTOUTEVILLE, GUILLAUME D' (1403–1483), French ecclesiastic, archbishop of Rouen, was sent to France as legate by Pope Nicholas V to make peace between Charles VII and England (1451), and undertook, ex officio, the review of the trial of Joan of Arc; he afterward reformed the statutes of the University of Paris. He then went to preside over the assembly of clergy which met at Bourges to discuss the observation of the Pragmatic Sanction, finally returning to Rome, where he passed almost all the rest of his life. He was a great builder, Rouen, Mont St. Michel, Pontoise and Gaillon owing many noble buildings to his initiative.

ESTRADA PALMA, TOMÁS (1835–1908), first president of Cuba, was born near Bayamo, Oriente province, July 9, 1835. He became a general in the revolutionary army during the Ten Years' War (1868–78) against Spain and in 1875 president of the provisional government. In 1877, however, he was captured by the Spaniards. After his release he went to Honduras and later moved to Orange county, N.Y., where he opened a boys' school. In 1895, when the Cuban revolution again broke out, he went to New York city and became active in the work of the Cuban junta. After the death of José Martí, Estrada Palma became the real head of the revolution. In 1902, when the United States turned the island over to the Cubans, Estrada Palma became the first president of the republic by an almost unanimous vote. His rule was honest and intelligent and brought Cuba to a high degree of prosperity. He joined no political party at first but was forced more and more to ally himself with the Conservatives to get his measures through congress. He became the candidate of the Conservatives for re-election in 1905, an election in which the Conservative leaders probably used fraudulent methods at the polls without Estrada Palma's knowledge. The Liberals charged a corrupt election and fomented the revolution of 1906, which resulted in the resignation of Estrada Palma on Sept. 6, 1906. The United States was forced to take control of the government temporarily. Estrada Palma died in Oriente province on Nov. 14, 1908.

See P. D. Camacho, *Estrada Palma, el gobernante honrado* (1938). (R. H. F.)

ESTRADES, GODEFROI, COMTE D' (1607–1686), marshal of France and one of Louis XIV's ablest diplomats, was born at Agen. He served with distinction in the Low Countries during the Thirty Years' War, conducted a famous defense of Dunkerque (1651–52) and took part in later campaigns in Catalonia (1655), Italy (1657) and Holland (1672). He was *maire perpétuel* of Bordeaux (to which the title of *comte* was attached) from 1653 to 1674 and was made *chevalier de l'Ordre du Saint Espirt* in 1661 and marshal of France in 1675. Having been ambassador extraordinary to Holland (1646) and a participant in the conferences at Münster, he went as ambassador to London in 1661; the attempt of the Spanish ambassador there to have precedence over him led to a great triumph for France; and in 1662 he negotiated the purchase from England of Dunkerque, of which he became governor. He was again ambassador to Holland from 1663 to 1668 and also French *premier plénipotentiaire* in the negotiations at Breda in 1667 and at Nijmegen from 1675 to 1678. He died in Paris on Feb. 26, 1686. The first collections of the letters and diplomatic papers of the comte d'Estrades (1709, 1710, 1718 and 1743) contain material of dubious authenticity. A better edition was begun by A. de Saint-Léger and L. Lemaire, *Correspondance authentique de Godefroi, comte d'Estrades*, vol. i (1924).

BIBLIOGRAPHY.—P. Lauzun, *Le Maréchal d'Estrades* (1896); A. de Saint-Léger, *Les Diverses Éditions des Lettres, Mémoires et Négociations de M. le comte d'Estrades et la propagande anti-française . . .* (1923); J. Jusserand, *Le Maréchal d'Estrades et ses critiques* (1928). (F. C. S.)

ESTRÉES, GABRIELLE D' (1573–1599), mistress of Henry IV of France and by him the ancestress of the Vendôme branch of the house of Bourbon, was the daughter of Antoine d'Estrées, marquis de Coeuvres, and Françoise Babou de La Bourdaisière. At the court of Henry III she met the seductive Roger de Saint-Lary, later duc de Bellegarde, whose mistress she seems to have become and who finally introduced her to Henry IV, then occupied in conquering his kingdom. Henry, who stayed at the château of Coeuvres in Nov. 1590, fell violently in love. It was probably during the siege of Chartres (1591) that she became his mistress. Henry arranged a purely formal marriage (June 1592; annulled in 1594) between her and Nicolas d'Amerval, seigneur de Liancourt; this did not prevent his acknowledging her publicly as his mistress from Dec. 1592. Henry, indeed, was often accused of compromising his victories in order to visit her. She had his entire confidence and influenced him in his decision to become a Catholic (1593), in the hope that the pope would then annul his marriage to Queen Margaret. Surrounded by a rapacious entourage, Gabrielle showed herself neither disinterested nor faithful. She gave the king three children who were legitimized, namely César, duc de Vendôme (1594–1665), Catherine Henriette (1596–1663) and Alexandre (1598–1629), later grand prior of France.

In 1595, when her son César was legitimized, Gabrielle was made marquise de Monceaux. Subsequently she was made duchesse de Beaufort (1597) and duchesse d'Étampes (1598). Master of France at last (1598), Henry seriously considered marrying her despite the opposition of the Holy See and of the duc de Sully (Maximilien de Béthune, whose protector she had been). He gave her the ring which he had received at his coronation, and she already saw herself as queen of France when she died suddenly, during an attack of eclampsia, in Paris on April 10, 1599. Her funeral took place with royal honours at St. Denis.

BIBLIOGRAPHY.—J. B. Capefigue, *Gabrielle d'Estrées et la politique de Henri IV* (1859); A. Desclozeaux, *Gabrielle d'Estrées*, Eng. trans. (1907); E. de Lanouvelle, *Gabrielle d'Estrées et les Bourbon-Vendôme* (1936); R. Ritter, *Charmante Gabrielle* (1947). (P. Er.)

ESTRÊLA, SERRA DA, the highest mountains in Portugal, are situated in the centre of the country between the basins of the Tagus and the Mondego to the north. They run about 40 mi. from northeast to southwest and are about 10–15 mi. wide; their summit reaches 6,532 ft. A block mountain range composed of Paleozoic schists and slates, they are the western continuation of the Cordillera Central. They consist of a series of high plateaus trenched on the borders by deep valleys—notably those of the Zêzere and upper Mondego—and fault scarps. With an annual rainfall of more than 90 in. above 4,000 ft., the range is one of Portugal's chief pastoral districts, transhumance of sheep and goats still being practised. Woolen mills utilizing local power are sited around the edge of the massif. Covilhã is the chief centre and is a winter sports resort. The permanent snow line lies at about 5,250 ft., with snow in the winter months to 3,600 ft. Winter sports are popular, with ski lifts organized by the Ski Club of Portugal into the mountains from Covilhã. (J. M. Ho.)

ESTREMADURA, a coastal province of central Portugal, comprises southern Leiria, most of Lisboa and northern Setúbal districts. Area 2,057 sq.mi. Before 1933 it consisted of the whole of Leiria and Lisboa and, for a time, Santarém and all Setúbal. (For the historical explanation of the name, *see* EXTREMADURA.)

The limestone peninsulas of Lisboa and Setúbal are separated by the estuary of the Tagus and present striking contrasts. The Lisboa peninsula, with its basaltic Serra de Sintra (1,706 ft.) facing the Atlantic, has more than 30 in. of rainfall annually. Its parks of Pena and Monserrate have a lush vegetation, descendants of plants imported from Madeira and elsewhere. Along the northern shore of the Tagus estuary there are many tourist resorts. Lisbon and its surrounding countryside, which has a relatively dense rural population, is the centre of economic activity in the peninsula. There are rich wheatlands on the basalt plateau to the west of the capital and on the Miocene clays to the north, but more important are the vineyards of Colares and Mafra.

The drowned estuary of the Tagus, the Mar da-Palha, forms Lisbon's excellent natural harbour, covering nearly 100 sq.mi. The low, flat lands to the south on the Setúbal peninsula, sheltered by the limestone anticlinal Serra da Arrábida, were colonized from about 1850, with vineyards and cork estates. The valley of the Sado has long been malarial, sparsely settled (apart from Setúbal) and largely waste, but in the 1960s there was some cultivation of irrigated rice fields.

The backbone of northern Estremadura consists of Jurassic and Cretaceous hills deeply dissected in clays and sands. The area is bordered on the east by the Serra de Montejunto (2,178 ft.) and the Serra de Caldeiros (2,062 ft.). North-south valleys carry important roads and form agricultural districts, the most important centres being Tôrres Vedras, where there are more vineyards, and Alcobaça. Abundant springs make it possible to grow many different crops, including fruit. The coast constitutes an abrasion platform of Miocene or Pliocene Age, varying from a broad lowland in the Nazaré district to a narrow shelf in the south. Near Sintra, the cliffs of Cape Roca rise more than 300 ft. from the sea at the most westerly point of Portugal (9° 30′ W.).

Apart from the provincial and national capital of Lisbon (pop. [1960] 818,382), the chief towns of Estremadura (with municipal population) are: Setúbal (59,024), the chief sardine port and a busy canning centre; Barreiro (35,758) and Almada (71,075), suburbs of Lisbon across the Tagus estuary; Caldas da Rainha (38,749), a spa; and the fishing towns of Peniche (22,898) and Nazaré (13,564). (J. M. Ho.)

ESTRUP, JACOB BRØNNUM SCAVENIUS (1825–1913), Danish statesman, Conservative prime minister from 1875 to 1894, was born April 16, 1825. In 1846, he took over his father's estate. In 1864 Estrup, as a member of the so-called National Landowners' party, entered the *landsting*. As minister of the interior from 1865 he carried through major improvements in the railways and in Esbjerg harbour. He set his mark on the new constitution of July 1866; it secured the landowners a commanding position in the *landsting*, where he became the leader of a powerful group. The first Conservative governments were not strong, and when the Liberals obtained a majority in the *folketing* in 1872, a long dispute over the constitution ensued (*see* DENMARK: History). In 1875 Estrup formed a government, backed by the *landsting* majority and by the king, Christian IX. The *folketing* rejected his motion on fortifications, but his government nevertheless promulgated provisional finance measures, which the Liberals considered unconstitutional. Estrup counteracted public unrest with police action and a stricter penal code. The Liberal leaders, faced with a strong Conservative party, disagreed among themselves, and the dispute was resolved by a compromise. Some social reforms were carried through by Estrup in collaboration with the moderate Liberals, whose majority supported the finance act of April 1894; this contained grants for temporary military installations but repealed Estrup's earlier police and press measures. In Aug. 1894 Estrup resigned. As a member of the *landsting* until his death on Dec. 24, 1913, he was a prominent critic of the Liberal governments after 1901.

See K. G. Brøndsted, *J. B. S. Estrup* (1925). (F. Sk.)

ESTUARY, an inlet or embayment, the lower end of a river valley invaded by the sea, resulting either from a rise of sea level or sinking of the land. Because such valleys or bays narrow upstream and have smaller tributary valleys, tidal movement is commonly constricted and heightened and therefore tidal effects are conspicuous; hence the name estuary (Lat. *aestuarium*, "a place reached" by *aestus*, "the tide").

Estuaries usually are not bays of simple form. The sea has invaded a river system with tributaries, and small bays extend up many of the drowned tributary valleys and in some cases even into the tributaries of the tributaries, thus forming an extremely intricate set of bays and inlets.

An estuary normally is bordered by extensive marshes and mud flats. The unsubmerged divides between the stream valleys furnish sediment that gradually accumulates as flats along the shores. The upper end of nearly every inundated valley extends above sea level, and the upper reaches of the streams in these valleys continue to bring sediment down and deposit it as small deltas (*see* DELTA) or mud flats at the head of each inlet or bay. Numerous fresh-water streams ordinarily flow into an estuary, and the net movement of water is toward the ocean; consequently the water in estuaries is less salty than that in the adjacent ocean. Many long estuaries receiving large streams are nearly fresh in their upper reaches.

Estuaries are widely distributed geographically. The northeast coast of the United States exhibits many. Chesapeake bay and Delaware bay are excellent examples, as are many of the numerous bays of the Maine shore. Most of the bays and inlets of the English and French coasts are estuaries also.

BIBLIOGRAPHY.—D. W. Johnson, *The New England-Acadian Shoreline* (1926); P. H. Kuenan, *Marine Geology* (1950); W. H. Twenhofel, *Principles of Sedimentation* (1950); A. Guilcher, *Coastal and Submarine Morphology* (1958). (W. C. C.)

ESZTERGOM (Slovak OSTRIHOM; Ger. GRAN; Lat. STRIGONIA), one of the oldest towns in Hungary, lies 40 km. (25 mi.) N.W. of Budapest, in the *megye* (county) of Komarom, on the right bank of the Danube, which there forms the frontier between Hungary and Czechoslovakia. The site is at the western end of the gorge cut by the Danube between the Pilis and Börzsöny hills, dividing the Little Alföld from the Great Alföld or Hungarian Plain. Pop. (1960) 23,065 (mun.). Opposite Esztergom, across

the Danube, the Hron tributary enters the main stream from the north. In the angle formed by the confluence lies the Slovak town of Sturovo (formerly Parkan or Parkány). The two towns were for long connected across the Danube by a bridge which was destroyed in World War II. Esztergom was the capital of the early Arpád kings. Stephen I was born in the city and crowned there in 1001. The various forms of its name are all references to Esztergom's importance as a grain market. It was sacked by the Mongols in 1242 and occupied by the Turks in the 16th and 17th centuries. The archbishopric, also one of the oldest in Hungary, was moved to Trnava in Slovakia during the Turkish occupation and Esztergom received it again only in 1820. The cathedral is a small copy of that of St. Peter's, Rome. There was some manufacturing development in heavy industry and textiles after World War II. (H. G. S.)

ETA, a Japanese outcaste group of butchers and leatherworkers. Traditional Japanese society, in the period before the advent of westernizing influence, was rigidly structured. Historically, there were nobility and a military aristocracy, below these the commoner farmer and artisan, and, at the bottom of the social hierarchy, several pariah or outcaste groups. Two of the most numerous of these untouchables, the so-called *senmin,* were the Hinin and the Eta. Both groups engaged in conceptually degrading and menial occupations, the Hinin being executioners, itinerant actors and mountebanks, while the Eta were those engaging in butchering and leatherworking. Of the two, the Eta were considered the lower. They were limited to certain areas and districts, were obliged to marry only into their own group and could not serve other Japanese as domestics. All upper classes in Japan, in fact, avoided contact with the Eta, whose very presence, it was felt, was defiling. The Hinin, although themselves avoided by other Japanese, were also contemptuous of the Eta. A societal situation not unlike that of the caste distinctions traditional in India was created.

Eta origins may depend, in fact, on contact with India, especially Buddhism and its social stigma against those who are in any way concerned with the taking of life, whether human or animal. This *ahimsa* concept of Hinduism-Buddhism may have influenced the attitudes of the early Japanese against those whose hereditary occupations were butchering and leatherworking. But this is speculation, since Eta origins are much in dispute. Another theory, widely held in Japan, derives the group from Korean war prisoners and slaves.

Modernization, following the Meiji restoration of 1868, saw laws passed to abolish the medieval castes, including the Eta, Hinin and similarly depressed groups, as well as the samurai or knightly aristocracy. Such legislation removed the restrictions on residence, marriage, landholding and occupation to which the lowest castes had been traditionally subject. In 1922, the Eta organized to form an equality association of 1,300,000 members, a fact which indicated both the numbers of the caste and the continuing sense of discrimination. In the 1960s discriminatory attitudes were still reported against the descendants of Eta, expressing themselves in the avoidance of marriage to such individuals or in a general unwillingness to take food allegedly prepared by an Eta. *See* also SANKA.

See S. Ninomiya, "An Inquiry Concerning the Origin, Development, and Present Situation of the Eta . . . ," *Transactions, Asiatic Society of Japan,* 2nd ser., x: 47–154 (1933). (R. F. SR.)

ETAH, a municipal town, *tehsil* and district in the Agra division of Uttar Pradesh, India. The town, 46 mi. N.E. of Agra, mainly important as district headquarters, is also the seat of a college affiliated to Agra university. Pop. (1961) 24,663. It is on the Grand Trunk road and a spur of the Northern railway main line. ETAH TEHSIL (482 sq.mi.) had a population in 1961 of 360,719.

ETAH DISTRICT (1,715 sq.mi.; pop. [1961] 1,300,199), mainly an alluvial plateau irrigated by the Ganges canal network, dips along its northeastern border into the Ganges valley. Between the modern river bed and the swamps and hollows to the southwest marking its ancient channel lies a belt of fertile siltland. Kasganj (31,554), 15 mi. N. of Etah, is a centre of cotton and sugar processing and trade, a major grain market and a busy junction on the Northeastern railway. Soron (13,119) is a Hindu pilgrimage place with important fairs. (B. SI.)

ÉTAMPES, ANNE DE PISSELEU, DUCHESSE D' (1508–1580), mistress of Francis I of France, was the daughter of Guillaume de Pisseleu, lord of Heilly in Picardy. One of Louise of Savoy's maids of honour, she supplanted Madame de Châteaubriant (Françoise de Foix-Lautrec) in the king's affections after his release from Madrid (1526). Having married her to Jean de Brosse (probably in 1530), Francis conferred the countship of Étampes on the couple in 1534, erecting it into a duchy in 1537. Everywhere the antagonist of the dauphin Henry's mistress, Diane de Poitiers, the duchesse supported Philippe Chabot (*q.v.*) against the constable duc de Montmorency and was a protectress of the Reformed religion and of such writers as Clément Marot. After Francis' death (1547), her backing of Jarnac against the favourite La Châtaigneraie (*see* DUEL), together with Diane's enmity, made her dismissal from court inevitable. She died in retirement in 1580, having been deprived of her duchy for nine years (1553–62) and estranged from her husband (d. 1565).

ETAWAH, a municipality, district and *tehsil* in the Allahabad revenue division of Uttar Pradesh, India. The city is on the east bank of the Jumna, 67 mi. E.S.E. of Agra. Pop. (1961) 69,553. It is intersected by deep fissures over which bridges and embankments carry the broad streets from the old city on the south to the new on the north. The town's chief monument, the 16th-century Jami Masjid or principal mosque, built on high ground from the fabric of Hindu buildings, has a façade 13 ft. long. There are also fine Hindu temples about the mound on which stands a ruined 15th-century fort. Etawah is on the main Northern railway between Agra and Kanpur and is connected northward by a metaled road to the Grand Trunk route. It is an important handloom centre and produce market from which ghee (clarified butter) is distributed to many parts of India.

ETAWAH DISTRICT (1,669 sq.mi.; pop. [1951] 970,695; [1961] 1,182,988) extends southwestward across the Ganges-Jumna doab and beyond the Jumna valley to the gorges of the Chambal river and the rocky outliers of the Vindhya mountains. Much of the southwestern tract is cut by great ravines, and near the rivers there is recurrent soil erosion. Reclamation and afforestation schemes have been undertaken. *See* DOAB. (B. SI.)

ETCHING is the process of biting lines or areas by means of an acid. By etching generally is meant the process of biting these lines in a metal plate with a view to its being printed from; by an etching is meant the print taken from a plate so etched. Lines may indeed be etched on a metal plate which is itself intended to serve some decorative purpose, with no idea of prints being taken from it, but etching of this sort should more properly be treated in connection with decorative metalwork and has no significance here. Etching, as above defined, has this important point in common with line engraving, that the lines which are to appear black on the print are incised, as opposed to those in wood engraving, which are in relief. The process of printing from an etched plate is, therefore, identical with that of printing from an engraved plate. Etching differs from line engraving only in the process by which the lines are incised; in the latter case, by an instrument of triangular section which scoops a shaving out of the metal; in the former, by chemical action. A print from an etched plate may generally be distinguished from a print from an engraved plate by the fact that in an etching the lines do not diminish or increase gradually in thickness but do so in more or less abrupt stages, and that the endings of the lines are square whereas in an engraving they taper gradually to an end. These differences are inherent in the different processes. It is obviously impossible to end a line abruptly by means of a triangular gouging instrument, such as a burin; if the burin were stopped suddenly the shaving it forms in its course would be left. In etching, on the other hand, the thickness or thinness of the lines being obtained by successive bitings, and each successive biting being comparatively uniform over the plate, a line, which is intended gradually to diminish in size, has (microscopically perhaps) a form like that of an extended telescope. Engraving has, of course, been con-

stantly used in combination with etching, and it is often a matter of difficulty to distinguish between the parts which are purely etched, those which have been first etched and later strengthened by the engraving tool, and those which have been directly engraved. A further word should be added on the subject of dry point. This is merely the process of scratching with the etching needle directly on the plate. In its passage the needle leaves an irregular ridge on either side of the line which it makes (burr) to which the ink adheres, so that a dry-point line when printed has at first a slightly blurred but rich effect; this burr wears away quickly, and the scratched line by itself, when printed, is then faint and meagre. Dry point, though it is independent of the use of mordants, is generally employed in conjunction with etching for the richness of its effect.

The number of satisfactory prints which may be taken from an etched plate varies according to the hardness of the metal and the depth to which the lines have been etched. A fine line on a copper plate may be almost obliterated after 200 or 300 impressions; a dry-point line, as has been indicated, will lose its characteristic effect very much sooner.

Some of the prints taken from Rembrandt's plates in the latter part of the 19th century were almost, though not quite, as good as those taken during his lifetime. *See* also Dry Point; Engraving, Line; Woodcut and Wood Engraving.

HISTORY

Etching, like engraving, was probably invented north of the Alps, and it is in Germany, France, England and, above all, the Netherlands that the greatest triumphs of the process have been achieved. It will be most convenient in a brief summary of its history to deal with its progress in the Teutonic countries from its beginning to the end of the 17th century, and then to return to Italy, France and Spain. As a means of decorating metal, particularly armour, etching was practised at least as early as the middle of the 15th century. It seems probable that, as the earliest engraving was the work of goldsmiths, so the earliest etching is to be credited to the armourers. The idea of printing from such a plate they no doubt borrowed from the goldsmiths.

16th and 17th Centuries

Germany.—The first etching to which an approximate date can be given is a portrait of Kunz van der Rosen by Daniel Hopfer (working 1493–1536), which a rather complicated line of reasoning assigns to the year 1504 or earlier. Daniel Hopfer was one of a family of armourers working in Augsburg, but was an artist of much originality. Hans Burgkmair the Elder (1473–1531), the dominating personality of the Augsburg school, upon whom the Hopfers largely depended, executed a single etching, no doubt learning the process from the lesser artists. The earliest etching which actually bears a date is one by the Swiss goldsmith, soldier and draftsman Urs Graf (d. 1527), of the year 1513, most probably done at Basel. Albrecht Dürer (1471–1528), with the eagerness for experiment which characterized him, tried etching, but apparently found it an unsympathetic medium and after a few experiments gave it up in favour of engraving. His half dozen etchings dated between 1515 and 1519, impressed though they are with that great artist's personality, are among the least satisfactory of his reproductive work. Iron, the metal used in all these first attempts, did not allow of much delicacy, and the result, compared with line engraving on copper, for which it was regarded as merely a less laborious substitute, was unsatisfactory. However, it was a new process which everyone must try out for himself. Nicolas Hogenberg of Munich (working 1523–37), whose artistic career was passed at Malines in the service of Margaret of Austria, was one of the first northern artists to use etching with rapidity and freedom, and his frieze of the entry of Charles V into Bologna in 1530, in a number of plates, is important from its size as well as from its subject. Frans Crabbe (the Master of the Crayfish; d. 1553), who was associated with Hogenberg, also followed his example in etching, but his work was unequal in quality.

Dürer's followers in Nuremberg, Hans Sebald Beham (1500–50) and Georg Pencz (1500–50), etched only a few plates, for the most part in the years immediately following Dürer's experiments. Albrecht Altdorfer (about 1480–1538) of Ratisbon, no doubt also impelled by Dürer's example, tried his hand at etching as early as 1519, and later used it for the first pure landscapes produced. Augustin Hirschvogel (1503–53?) and Hans Sebald Lautensack (working 1524–63) in Vienna followed Altdorfer's example in landscape but without quite the freshness and charm of its originator. The only succeeding etcher of importance in Germany is Jost Amman (1539–91), who worked at Nuremberg.

The Netherlands.—Lucas van Leyden (1494–1533), who probably learned the process from Dürer during the German's visit to the Netherlands in 1520–21, etched a few plates. These are technically superior to Dürer's and of importance as probably the first examples of the use of copper for etching, which made it possible for line engraving to be used in combination with it. The fine portrait of Maximilian which Lucas van Leyden made in 1521 is an example of this combined use of etching and engraving, the whole of the emperor's face being finished with the burin. Dirick Vellert of Antwerp (working 1517–44), an artist of delicate accomplishment, used a technique much resembling Lucas', though the majority of his works were engraved.

In the Netherlands line engraving usurped the field in the latter half of the 16th century, and comparatively little etching was done. Jan Corneliszoon Vermeyen (1500–59), the court painter of the emperor Charles V, had etched a number of plates. He is probably the earliest example in the Netherlands of the painter, with no training as an engraver, turning his hand to etching. His plates, some of them of large size, were executed in rather a formal style but are extraordinarily fresh and effective. Marcus Gheeraerts the Elder (*c.* 1521–1604) of Bruges was the author of a charming series of small plates illustrating Aesop (1567), which were etched with a delicacy and sensitiveness hardly equaled by Adrian van Ostade or Hollar. Gheeraerts spent the latter part of his life in England, where he issued further etchings, the first to be published in that country. Hieronymus Cock (1510?–1570), the publisher, did some admirable landscapes of the conventional type, and the great Pieter Bruegel (1525?–1569) and Frans Floris (*c.* 1517–70), the fashionable painter of the time in Antwerp, each etched one plate, while Hans Bol, the landscape painter and draftsman of Malines (1534–93), within rather narrow limits, was an etcher of considerable charm. Paul Brill of Antwerp (1554–1626), who passed practically the whole of his life in Italy, is chiefly interesting as a link between the conventional Italian landscapes evolved in the Venetian school and the native style of the Netherlands. Hercules Seghers (*c.* 1590–1645), an artist of real originality, was dependent on Brill's theatrical conventions, while still in contact with Rembrandt. Seghers' curious experiments in printing and colouring by hand are interesting and unique though they cannot actually be classed as colour prints, as he never attempted to print with more than one colour at a time. Esaias van de Velde (*c.* 1590–1630) and his brother Jan (*c.* 1596–1641) were the earliest of the Dutch etchers who looked to Holland for the subjects of their landscapes. Simple, almost meagre in technique as their etchings are, they rendered with a naïve charm and directness the views in their native land. Willem Buytewech (*c.* 1585–1625), from whose drawings the Van de Veldes worked, is well known as a painter of individuality, which appears in his few etchings.

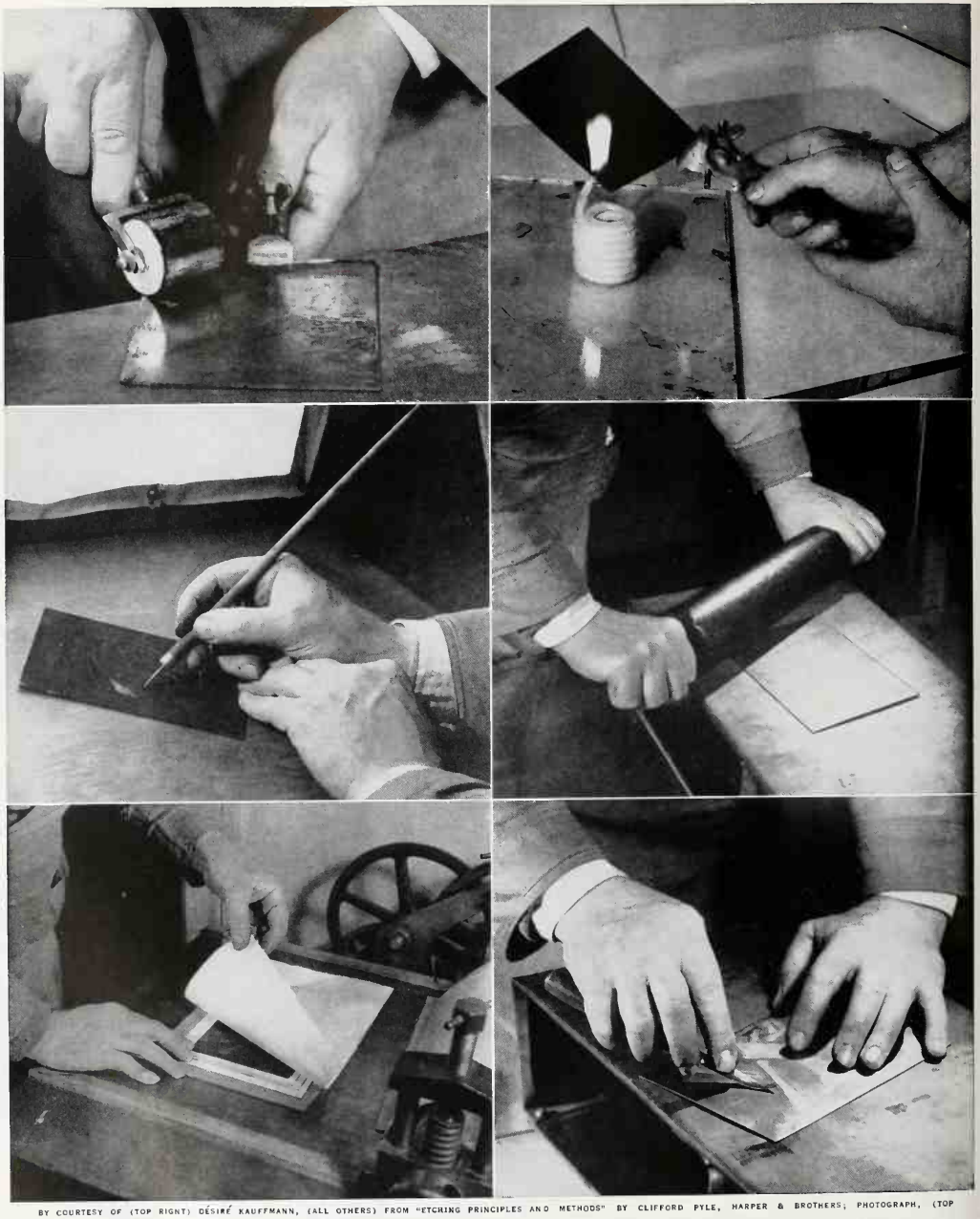
Van Dyck.—The enumeration of the names of these few etchers who worked in the Netherlands at the end of the 16th and beginning of the 17th century leads up to but does not explain the importance which, with surprising suddenness, the medium attained in the hands of Rembrandt (1606–69) and Anthony Van Dyck (1599–1641). The latter's achievement, remarkable as it is, is limited in scope and quantity, but as it precedes Rembrandt's in point of time it will be dealt with first. The famous "Iconography" was planned by Van Dyck as a series of engraved portraits of contemporaries eminent in the arts. Etching was to play only a subsidiary and preliminary part in it and, in fact, it was only for 18 of the 100 plates issued that Van Dyck himself executed preliminary etchings. The rest were engraved in line by the professional engravers trained in the school of Rubens. These 18

BY COURTESY OF (TOP LEFT) ANTHONY GROSS, (TOP RIGHT) CRESSET PRESS LTD., (CENTRE LEFT) BRITISH MUSEUM, (BOTTOM) MISS R. BIRNIE PHILIP

17th-20th CENTURY ETCHINGS

Top left: "La Maison de Poète" by Anthony Gross
Top right: "Lady Reading a Book" by Charles Keene (1823–91)
Centre left: Etching of a shell by Rembrandt (1606–69)
Bottom: "Black Lion Wharf, London" by James McNeill Whistler (1834–1903); second state

ETCHING



BY COURTESY OF (TOP RIGHT) DÉSIRÉ KAUFFMANN, (ALL OTHERS) FROM "ETCHING PRINCIPLES AND METHODS" BY CLIFFORD PYLE, HARPER & BROTHERS; PHOTOGRAPH, (TOP RIGHT) COPYRIGHT, D. VAN NOSTRAND COMPANY, INC., 1948

PROCESSES AND TOOLS USED IN MAKING AN ETCHING

Top left: Applying the ground to plate
Top right: Smoking the grounded plate
Centre left: Using the etching needle on the smoked plate
Centre right: Inking the plate
Bottom left: Lifting a print
Bottom right: Using the burnisher in preparation for final proof

etchings, too, were subsequently finished by these same engravers, and it is only in the very rare impressions taken from these plates, before the additions were made, that Van Dyck can be appreciated as an etcher. His methods were quite independent of previous practice and of the conventions of line engraving. No doubt the fact that he regarded the etching as merely preparatory was partly responsible for their freedom from preconception. His natural instinct for form, aided to a certain extent by the example of Baroccio and Italian etchers, led him directly to that solution of the problem of etching which has since become recognized as the right one. His methods seem so obvious that it is difficult as well as superfluous to try to describe them. Economy of line and contrast between this and the whiteness of the paper are the main secrets. The heads stand out, realized with quite extraordinary force and precision. There is no use of dry point or of tone on the surface of the plate; the effect is gained by clean draftsmanship.

Van Dyck's example as an etcher had comparatively little immediate effect; it was in modern times that his influence was most apparent. The great influence which he exercised during his lifetime and for the rest of the 17th century was on portrait engraving, through the "Iconography" in its completed, engraved state.

Rembrandt.—Rembrandt's importance as an etcher is greater than Van Dyck's in proportion to the superiority and greater versatility of his genius. The number of etchings attributed to him, varying with the trend of criticism from 140 to 300, is at all events considerable. These engravings comprise every variety of subject, scenes from the Old and New Testaments, genre, portrait, still life, landscape. In his treatment of each subject in turn, his method was as fresh and untrammeled as was Van Dyck's in relation to portrait. But unlike Van Dyck's, Rembrandt's approach was gradual, and final and complete as each successive achievement appears to us, to him nothing was satisfying or final. His earliest dated etching, the small portrait head of his mother, of 1628, is a masterly achievement, reflecting with great exactness the style of portraiture of his early Leyden period. Between this date and about 1632 his style underwent no material change, though his etchings vary enormously in quality and elaboration. About 1636 the influence of Rubens, apparent in his painting, was correspondingly seen in his etching in a different and lighter chiaroscuro, with a tendency to sharper contrasts in light and shadow. The so-called "100-guilder print" of the 1640s ("Christ Healing the Sick") marked the apogee of his next period, in which dry point played an important part, and where the whole plate was covered with a network of fine lines, giving almost the appearance of mezzotint. To 1653 belong two of his largest and perhaps most wonderful plates, the "Ecce Homo" and the "Three Crosses," in the latter of which his feeling for the dramatic moment and his tremendous power of expressing emotion in terms of chiaroscuro were triumphantly exhibited. The technique became slightly less elaborate; heavy individual lines were more apparent, there was less use of crosshatching, a tendency to model in vertical parallels and a masterly use of the effects of tone left on the surface of the plate and of dry point. In the etchings which followed up to the last dated one of 1661, there was a further tendency toward the still broader and more drastic treatment visible in his latest paintings and a more complete reliance on heavily etched line without dry point. The landscape etchings, most of which belong to the years about 1640 and 1650, stand somewhat apart in the comparative simplicity of their aim as studies from nature, with the exception of the most famous, the "Three Trees" of 1643, which is as dramatic as any of his subject etchings. The portraits are, for the most part, in the more careful and elaborate technique of his earlier period, perhaps insisted on by his sitters. Among the greatest are the numerous self-portraits he etched from the mirror in every variety of costume and character.

Rembrandt's contemporaries in Holland, except for his immediate associates and followers, Jan Lievens (1607–74), Ferdinand Bol (1618–80), Jacob (*c.* 1616–1708) and Philips Koninck (1619–88) and J. G. van Vliet, remained comparatively uninfluenced by his example. Anthonis Waterloo (1609?–1677?), who enjoyed in the 18th century an exaggerated vogue, has indeed some claims to importance as a landscape etcher, and Allart van Everdingen (1621–75), in his small plates, mostly of Norwegian scenes, showed a delicate and observant talent. Jacob van Ruysdael (*c.* 1628–82) may claim a place next to, though indeed far below, Rembrandt's as a landscape etcher. His delicate and exquisitely finished etchings are in the nature of detail studies of particular woodland scenes, diversified by still water reflecting the trees. Of the Dutch painters who looked for their themes to Italy, Nicholas Berchem (1620–83) and Karel du Jardin (1622–78) are the best known. Their landscapes included as important parts of the composition human figures and animals, and they serve as a link connecting the landscape with the animal and genre painters. Of the animal painters Paul Potter (1625–54) is the most famous, of the genre painters Adrian van Ostade (1610–85). Ostade used etching with a painter's eye, and knew equally well how to render with extraordinary subtlety the subdued light of a squalid interior and the glow of sunlight on a scene of rustic festivity.

Great Britain.—The only etcher of note to work in England during the 17th century was Wenceslaus Hollar (1607–77), who was born at Prague but spent the greater part of his life in the British Isles. Of the very large number of etchings which he executed, the majority were intended to serve merely utilitarian and topographical purposes, but in spite of this nearly everything that he did shows at least a touch of real artistic feeling.

Italy.—Etching, as has been seen, was probably invented north of the Alps, but it was in Italy that its potentialities as an independent graphic medium were first realized. Though Marcantonio Raimondi (*c.* 1460–*c.* 1530) certainly used etching in a few of his prints and was probably the first to do so in Italy, generally in conjunction with line engraving, Parmigiano (1504–40) was the earliest to employ the etching needle not as a substitute for the burin but with the freedom of the pen. His etchings are obvious reproductions of his masterly but rather facile drawings, which enjoyed a great contemporary popularity, and he most probably used etching as a method of satisfying the demand for his drawings. Following his example, most of the great Italian painters thought it incumbent on them to make at least a few etchings. Of Parmigiano's immediate following, the Venetian painter Schiavone (d. 1582) was the only etcher of importance, and his work in this direction was almost entirely derivative from that of his prototype. Battista Franco (1498?–1561), the Venetian follower of Michelangelo, did a certain amount of etching, but in a lighter and more formal style, and the work of the other Venetians or Veronese such as Battista and Marc' Agnolo del Moro, G. B. and Giulio Fontana and Paolo Farinati shows rather a reaction against the freedom of Parmigiano's style, to which, however, Jacopo Palma the Younger (Palma Giovane) (1544–1628) returned with some measure of success. Federigo Barocci of Urbino (1528–1612), in his two or three etchings, worked out a new and striking method of his own, detailed and careful but at the same time quite different from the conventional engraver's style, and produced a masterly rendering of the peculiar effects of chiaroscuro at which he aimed in his paintings. Annibale Carracci (1560–1609), the most brilliant exponent of the so-called eclectic school of Bologna, in his few etchings was less original, but they still have the stamp of a real artistic personality.

In etching, as in painting, the artists of Italy were roughly divided into two schools during the 17th century, the one following the orthodox teaching of the Bolognese Carracci, the other deriving from Michelangelo da Caravaggio. The division was no longer a territorial one, and the eclectics worked side by side with the *tenebristi*, or followers of Caravaggio. This close contact led by degrees to a mutual influence by one school on the other, and the original line of demarcation gradually became obscured. Guido Reni (1575–1642), an actual pupil in the Carracci school at Bologna, was the most eminent of the eclectic etchers, and his well-balanced, delicately sentimental style became the classic type for most of the etchers in France and Italy during the century. Simone Cantarini, Giovanni Antonio and Elisabetta Sirani were agreeable echoes of Guido. G. F. Barbieri (Guercino) (1591–1666) in his few etchings was more original and showed some Caravaggesque influence as did Giuseppe Caletti (*c.* 1600–60). G. F. Grimaldi (1606–80?), the chief landscape etcher of Bologna,

was conventional and dull. Carlo Maratta (1625–1713) carried on the Carracci style into the 18th century. The greatest etcher working in Italy in the 17th century was undoubtedly the Spaniard José de Ribera (1588–1652). His style, insofar as it is not to be classed as Spanish, was definitely Caravaggesque, but his technique as an etcher was partly derived from the Bolognese. The sensitiveness of his outline, which played an important part in his work, and the sureness with which he knew how to render the changes from brilliant light to darkest shadow were among his great merits. Ribera's pupil, the Neapolitan Salvator Rosa (1615–73), the painter of romantic, bandit-infested landscapes, in his etchings, the less pretentious ones particularly, showed some of the charm which is associated with his fantastic world. In Genoa Benedetto Castiglione (1616–70), whose talent had something in common with Salvator's, is attractive from the grace which characterizes his etchings and interesting from the fact that he was influenced by Rembrandt's work. In Venice Giulio Carpioni (1611–74) etched with vividness and grace romantic scenes in a technique derived from the Carracci. Pietro Testa (1611–50), who worked in Rome, was an unequal but not uninteresting artist, with a curiously Venetian feeling for light which anticipated Tiepolo.

France.—Little etching was done in France in the 16th century. The frescoes decorating the palace at Fontainebleau, executed by Rosso Fiorentino, Primaticcio and their followers, were reproduced in etching by Antonio Fantuzzi, Leonard Thiry and other artists engaged in the work, but these were crude and hasty works. Jean Cousin (d. 1590) is credited with two or three etchings of more merit, and the architect and designer Jacques Androuet du Cerceau (*c.* 1510–80) used etching for his delightful architectural compositions with a truly French grace. But with the 17th century and the names of Jacques Callot (1592–1635) and Claude Lorrain (1600–82) France attained a position of real importance in the history of etching. Callot, unlike most of his contemporaries and successors, who depended almost entirely for their inspiration on Italian classicism, was essentially and originally French. Technically his innovations consisted in his handling of the individual line, by the thinning and thickening of which his effects were exclusively obtained, but he was also the inventor of a new world of minute fantastic figures, a sort of French stage fairyland. Besides their extraordinary elegance, such series as the *grandes* and the *petites* "Misères de la Guerre" ("Miseries of War") have a point and an emphasis in their narrative which is surprising on so minute a scale. The technique of subsequent etching could not and did not remain uninfluenced by Callot's practice, but his only close follower was an Italian, Stefano della Bella (1610–64).

The French academicians of the *grand siècle*, Eustache le Sueur, Charles le Brun and the rest, with their ready-made classical formulas had no chance of success in a medium as personal as etching should be, and it remained for Claude Lorrain, with his direct reactions to the landscape which he saw around him, to produce the only great etchings of the period besides those of Callot. Though his conception of landscape painting was bound by certain theatrical conventions, in his drawings and etchings he approached his subject directly and at the same time with the true landscape painter's instinct for selection. He was not a good etcher technically; there is something fumbling and amateurish in his work, but in spite of these disadvantages, he somehow succeeded in producing landscape etchings of an extraordinarily moving quality.

18th Century

The 18th century did not find in etching its most characteristic means of expression. In France, whose artistic preponderance in Europe during the period was undisputed, line engraving was more extensively and more successfully practised. Etching was to a large extent employed, it is true, but mostly as a preliminary to or in conjunction with engraving, and such mixed productions may more legitimately be classed as engravings. Antoine Watteau (1684–1721), François Boucher (1703–70), Honoré Fragonard (1732–1806) and the brothers Augustin (1736–1807) and Gabriel de St. Aubin (1724–80) all practised etching to a limited extent. Jean Duplessi-Bertaux (1747–1813) etched historical scenes, largely of republican and Napoleonic times, with a spirit and delicacy quite in the tradition of Callot.

It was in Italy and Spain, however, that the great etchers of the century flourished. Giovanni Battista Tiepolo of Venice (1696–1770), whose brilliant decorative painting was the culmination of rococo art, showed as an etcher equal brilliance. His technique, recalling that of José Ribera and G. B. Castiglione, was yet entirely original in its almost complete avoidance of heavy defining shadow and his method of rendering the broad, gray shadows by systems of herringbone and irregularly arranged short lines. His scenes, laid in a brilliant and all-enveloping sunlight, comprised some of the stock 18th-century sylvan and Arcadian genre, as well as more imaginative subjects, but treated alike with a peculiar point and irony. Canaletto (1697–1768), Tiepolo's contemporary in Venice, distinguished as the most brilliant portrayer of his native city, was an etcher of almost equal distinction. His 31 etchings of Venice and the neighbourhood reproduce faithfully the quality of his painting. It is, indeed, remarkable in him and in Tiepolo how their etching was directly dependent on their painting. They seem each independently to have discovered an extraordinarily complicated and original method of etching which would most exactly correspond to the essential quality of their painting. In each case the result was not, as might have been anticipated, a lifeless reproductive technique, but a brilliant addition to the repertory of etching. Giovanni Battista Piranesi (1720–78), the third great Italian etcher of the century, though also a Venetian by birth, worked nearly all his life in Rome. The bulk of his extensive etched work is archaeological, but his feeling for the dramatic in architectural composition, as exemplified in the "Carceri" ("Prisons"), and for the quality of the etched line certainly entitle this work to be regarded from the artistic point of view.

19th and 20th Centuries

Goya.—Francisco Goya y Lucientes (1746–1828) inevitably finds a place at the beginning of any account of modern art, and indeed his work looks forward into the 19th rather than backward into the 18th century, so extraordinarily fresh and original it is and so free from any of the typically rococo elements which marked 18th-century art. Yet technically Goya derived from Tiepolo (he visited Italy during his youth), and his earliest work in etching, though it reproduced paintings by Velázquez, was very much in Tiepolo's style. But his original compositions, the series of the "Caprichos," the "Proverbios," the "Desastres de la Guerra" ("Disasters of War") and last of all the "Tauromaquia" ("Bull-fighting"), showed the evolution of an original technique with the use of aquatint for the backgrounds. The mysterious satirical works, the "Caprichos" and "Proverbios," although the exact targets at which their shafts were aimed is uncertain, are overwhelming in the bitterness and intensity of their satire and the horror of the imagination which they show.

With the notable exception of Goya's work there was comparatively little of interest to record in the field of etching during the first half of the 19th century. It was one of those periods of stagnation which occur before a revival.

France.—The painters of the Barbizon school of landscape in France, Théodore Rousseau (1812–67), Charles Jacque (1813–94), C. F. Daubigny (1817–78), J. F. Millet (1814–75) and Camille Corot (1796–1875), all etched. The most prolific and important for his influence on the etchers of the next generation was Jacque, whose style shows the influence of Ostade and the Dutch etchers of the 17th century, while Daubigny shows great power and originality as an etcher. Millet in his plates, executed in a style which resembles that of Ostade's etchings magnified to about four times their original size, portrayed those subjects of peasants at work, familiar in his paintings, with the same instinctive understanding. Corot used the etching needle only occasionally but in a manner which, for all its seeming scratchy incompetence, vividly reproduced the atmospheric effects of his paintings. The greatest figure in French etching of the century, that of Charles Meryon (1821–68), stood rather curiously aloof from his contemporaries. Precluded by colour blindness from the practice of painting, he was that rather rare phenomenon, an etcher who is not a painter.

Influenced by a study of the work of the admirable but comparatively little-known Dutch 17th-century topographical and marine etcher Reynier Zeeman, and obviously, to a certain extent, by Piranesi, rather than by contemporaries, Meryon evolved for himself a system of line which in its clarity and incisiveness is unrivaled for the rendering of architectural subjects. His reputation, resting as it does on a small number of plates of the streets and churches of Paris, might, on the face of it, seem exaggerated, but the extraordinary perfection and inevitability of design which these show create out of a restricted subject matter a formula of universal application.

In France during the 20th century etchings have been mainly painters' by-products, frequently designed as illustrations to books in limited editions. The etchings of Henri Matisse (1869–1954), Pablo Picasso (1881–) and Marc Chagall (1887–) have distinctively personal characteristics, and those of André Dunoyer de Segonzac (1884–) are informed with a rare and sensitive understanding of the medium's qualities.

Great Britain and the United States.—In England, work of distinction, based on that of Rembrandt's contemporaries such as Ruysdael, was done by John Crome (1768–1821) of Norwich, as well as by John Sell Cotman (1782–1842), most of the latter in the process, invented at the end of the 18th century, called soft-ground etching. The names of Thomas Girtin (1775–1802) and J. M. W. Turner (1775–1851) can hardly be omitted in view of their pre-eminence as landscape painters, but their work in etching, masterly as it is, was in neither case intended to be final. Girtin's etchings were preparatory to aquatints and Turner's to mezzotints. Andrew Geddes of Edinburgh (1783–1844) was the author of a number of plates of real distinction, marked by a considerable and intelligent use of dry point, and Samuel Palmer (1805–81) was a landscape etcher of individuality.

The revival of etching in England in the 19th century was attributable in large measure to the influence of three men of whom undoubtedly the greatest was James McNeill Whistler (1834–1903). American by birth but thoroughly cosmopolitan, Whistler passed a large part of his life in England, and it was there that his influence was perhaps most pronounced. His conception of etching, as of art in general, even more than his practice, had the profoundest effect. Rembrandt was undoubtedly the chief influence on his etched work. His idea of an etching as an exact and inevitable composition, from which no single one of the innumerable lines which go to make it up could be removed or displaced by a hairbreadth, is almost justified by the perfection of some of his exquisite plates. Sir Francis Seymour Haden (1818–1910), Whistler's brother-in-law, though an amateur, exercised a direct influence on English etching even greater than that of Whistler, who had little immediate following. Haden was by no means an echo of Whistler; though inspired by him in the first instance, his was too independent a character to allow of anything in the way of imitation. He also went back to Rembrandt for inspiration and etched with an incisive strength and clarity which are admirable, though his work can hardly compare with the greatest. The third important influence on English etching was that of Alphonse Legros (1837–1911), a Frenchman and first Slade professor of art in London, where most of his life was spent. His was not a talent of extreme originality, but he had a great feeling for and understanding of etching and an extraordinary power of assimilating the methods and thoughts of the great masters.

Whistler had in Théodore Roussel (1847–1926), another Frenchman settled in England, a follower as fastidious as himself who later evolved an original method of colour etching, and, as a student, Walter Richard Sickert (1860–1942), an artist of real originality who developed along his own lines and founded an important school. In the tradition of Haden may be counted Sir David Young Cameron (1865–1945), whose landscape etchings and dry points are marked by faultless taste and distinction; and Sir Muirhead Bone (1876–1953), who specially developed the architectural theme. The most distinguished pupil of Alphonse Legros was William Strang (1859–1921), whose portrait etching is admirable but who perhaps failed by a lack of concentration in his powers to convince as a great etcher. The etched work of Augustus John (1878–1961) is the characteristic occasional work of an important painter, in the tradition of the great masters, while that of Sir Frank Brangwyn (1867–1956) is distinguished by its impressive size and feeling for composition in landscape and architecture. Sir Frank Short (1857–1945), for many years director of the school of engraving at South Kensington, exercised, after Haden and Legros, the greatest influence on etching, and was himself an etcher of great technical ability, though perhaps better known as the reviver of the process of mezzotint.

The succeeding generation was extraordinarily prolific in etchers of real distinction. As an etcher of landscape as well as of subject, largely eastern, James McBey (1883–1959) certainly attained an important position by the originality and strength of his work. Henry Rushbury (1889–), in a style somewhat akin to Sir Muirhead Bone's, did fine work in landscape, while Sir George Clausen (1852–1944), E. S. Lumsden (1883–1948) and many others had a merited reputation in the same genre. F. L. Griggs (1876–1938) revived the imaginary architectural composition originated by Piranesi and made of it a delicate and original means of expression. In portraiture, Francis Dodd (1874–1949), Gerald Brockhurst (1890–) and Malcolm Osborne (1880–1963) did admirable work.

After Whistler, the interest in etching in the United States was very intense, and a great deal of work was done, largely, it is true, by artists working abroad, of whom Joseph Pennell (1860–1926), D. S. Maclaughlan (1876–1938) (a Canadian by birth), Frank W. Benson (1862–1951), Herman A. Webster (1878–) and Arthur W. Heintzelman (1891–) were notable.

In England, however, in the second quarter of the 20th century, interest in etching showed a marked decline. The craft, though still practised by the older etchers, found little favour with the newer generation of artists. Etching became more of an occasional and tentative method of expression and less of a specialist's occupation. As an etcher who used the technique with lively inventiveness and a delightfully fresh approach, Anthony Gross (1905–) acquired a wide and deserved reputation. The brilliant experimental work of Stanley William Hayter (1901–) exerted a considerable influence both in England and France.

Other European Countries.—In the Netherlands, as might well be expected, much work of importance was done in the second half of the 19th century. J. B. Jongkind (1819–91), Jozef Israëls (1824–1911) and Matthijs Maris (1839–1917), though the work of the latter is the occasional essay of a painter, are artists of the first rank, while C. Storm van's Gravesande (1841–1924) and Marius Bauer (1867–1932) contributed something considerable to etching.

Scandinavia produced an etcher of astonishing virtuosity in the person of Anders Zorn (1860–1920). In Germany, much etching of real distinction was done. (A. E. P.; F. W. W.-S.; X.)

TECHNIQUE

The basic technique of etching is simple, but the process is so conditioned by unpredictable factors that the production of each etching has the character of an experiment in chemistry. Consequently there is often a need for alterations and corrections, and it is here that technical difficulties may occur. Broadly speaking, the process depends on the action of acid on metal. A thin sheet of metal, usually copper, sometimes zinc, is coated with an acid-resisting ground. The design is drawn on this coating with a needle which, by removing the ground, lays bare the metal without making any incision in its surface. The plate is then put in an acid bath. The acid attacks the lines of exposed metal and bites—that is, etches—them into furrows of a required depth. Thereafter the ground is removed and prints are taken from the intagliated surface.

Plates and Their Preparation

Copper offers the most suitable surface for etching. Zinc is quite widely used, but it has certain disadvantages: the action of acid on zinc is rapid and the resulting etched line coarse; zinc plates, because they are soft, yield a limited number of good prints, particularly when they are lightly bitten. Iron is satisfactory

when used with nitric acid, but is highly susceptible to atmospheric conditions. The best thickness for etching plates is 16 to 18 gauge. Thinner plates are apt to buckle in printing. Thicker plates have nothing to commend them; they are merely more expensive. The edges of all plates must be beveled; otherwise they will cut the paper under pressure of the rollers. Edges are taken down with a flat file to an angle of 45°, the corners are rounded and file marks are removed with a scraper and polished with a burnisher. Then the whole surface is polished free of scratches and cleaned of all trace of grease. A good cleaning agent is made of finely ground whitening compounded with a half-and-half solution of household ammonia and water. A piece of cotton wool or a soft clean rag is steeped in this paste and rubbed vigorously over the face of the plate. The paste is washed off under a tap; if the plate is perfectly clean, the water will spread in an even, glasslike film over its surface. The plate is thenceforward picked up carefully and held only by its edges to save the surface from contact with the fingers.

Grounding.—The plate is then ready to be coated with an acid resist. There are many formulas for making up etching grounds from such ingredients as wax, pitch and resin, but reliable mixtures can be bought ready-made from reputable firms. They are sold in the form of hard, black balls or round, flat cakes. To enable the ground to be applied in a molten state, the plate has to be heated. The most convenient stove is a small iron box in which a gas ring, Bunsen burner or spirit stove can be housed. When the plate is warmed, a sufficient quantity of the ground is melted out on its surface and spread, first with a piece of printing muslin, and then patted into a thin, even coat with a dabber. A dabber is a closely packed wad of cotton wool tightly encased in a covering of fine silk or kid. Dabbers can be bought or homemade. Another method of laying a ground is to apply it in a flat film with a leather-covered roller. The ground is rolled in all directions over the plate; the roller is cleaned immediately after use. When the coating is properly made, it is like a brown skin of unvarying thickness, and no gleam of bare metal is perceptible. Liquid grounds are also obtainable and satisfactory. They are poured on, and the plate is then maneuvered until the surface is wholly covered. It is highly important, especially when using liquid grounds, that the process be carried out under conditions that are as dust free as possible.

Smoking the Ground.—Many etchers prefer to work on a blackened surface because it enables them to trace their designs (through red or yellow carbon paper) and to watch the progress of their drawing as the needle begins to open up the lines. To smoke a ground, the plate is held upside down in a hand vise and its surface protected from the jaws of the vise by a folded piece of card or stiff paper; then it is passed slowly backward and forward over the flame of two or three lighted wax tapers twisted together. The brown surface of the ground is thus transformed into an even, sooty black; but the greatest care is needed to avoid scorching the ground.

Needling.—An etching needle proper or almost any steel point can be used. Sooner or later, each etcher decides on the particular tools that suit his hand. The best points are blunt and rounded. The needle is held nearly upright because in this position it moves most easily through the ground without scratching the metal. A fairly heavy instrument is desirable; its own weight will remove the ground without a consciously—and sometimes erratically—exerted pressure from the hand. Different weights of line are produced by needles of different thicknesses and, of course, by more or less prolonged immersion in the acid.

Acids.—The three acids in most general use are nitric acid, Dutch mordant and perchloride of iron. For etching copper, three parts of nitric acid are diluted by five parts of water to make a mixture of average working strength. A more concentrated solution bites more quickly and coarsely. For zinc, the proportion is one part of acid to three of water. Nitric attacks the metal fiercely and produces bubbles of gas which are a useful indication of the acid's action. Nevertheless, they should be dispersed with a feather, otherwise they will hinder the process of biting. Nitric is not suitable for fine, detailed work unless it has been diluted considerably. Dutch mordant is made by dissolving 4 parts of chlorate of potassium in a little hot water; when it has cooled, a solution of 20 parts of hydrochloric acid in 76 parts of water is added. This is the strength for copper. For zinc, the proportions are 2 parts of chlorate of potassium and 10 of hydrochloric acid to 88 of water. Dutch mordant bites more temperately than nitric acid; it etches deeper, but not wider, lines. Iron perchloride is a good, consistent mordant, but its action produces a sediment which settles on the lines and impedes the action. It is necessary, therefore, to lay the plate face downward on small wooden rests so that the sediment will fall to the bottom of the bath. Before the plate is immersed in acid, the sides and back are protected by a quick-drying coat of some such acid resist as Brunswick black. This is known as a stopping-out varnish.

During the needling of the plate, grease from the hand is sometimes deposited on the exposed lines. As a precaution, the plate is usually washed in acetic acid before etching begins.

Biting.—The plate is placed in the acid until the lightest lines are bitten to a sufficient depth. The duration of this process is infinitely variable. Much depends on the strength of the acid, the nature of the metal and many other factors. For example, the acid will bite more energetically at warmer temperatures; again, it will be more active in a network of close lines than in lines drawn at wider intervals. The depth of the biting can be gauged approximately by examining the lines under a magnifying glass or linen tester, or by feeling a sample line with the etching needle and letting the point travel along the groove. There is, however, no certainty in such tests nor, indeed, any reliable guide other than long experience.

After the first biting, the plate is removed, washed thoroughly in water and dried with blotting paper. Those lines which are considered deep enough are then coated carefully with stopping-out varnish, and the plate is returned to the acid to allow the other lines to be strengthened. The process may be repeated again and again. The success of an etching is not measured, however, by the number of bitings; many admirable etchings have been made with only one immersion. Some etchers adopt a reverse method. The darkest lines are opened first and then bitten; the plate is removed from the acid, the less dark lines are added, and so on. In this way, the most delicate lines are immersed for the shortest time. Alternatively, the grounded plate may be placed in the acid and needled without removing it from the bath until the whole design is completed. In the latter process, though the etching needle itself is attacked by the acid, it does not lose its point. Both techniques call for a good deal of certainty and control, but they are expeditious and they dispense with the need for stopping out. A further method, known to have been used by Whistler, is to let a little acid fall on the plate and guide it over the surface with either a feather or a piece of cotton wool wrapped round a stick.

Regrounding.—If it should be necessary to take an interim proof during the process of biting, the plate is washed, dried, placed on the stove and warmed. First, the varnish is cleaned from the back and then the ground from the front, with turpentine and a rag. If the proof shows that the plate is underbitten, it may be regrounded with a leather roller. The new ground is rolled out on a warmed, unused plate. When it is as smooth and as thin as possible, it is applied to the used plate. If the proof shows that the plate is overbitten, the surface can be lowered with a scraper or a burnisher. The scraper is a fluted, triple-edged tool; the burnisher is oval in section with a highly polished surface. If a correction, involving complete redrawing, has to be made, the faulty area is taken out with the scraper or, if the biting is shallow, the burnisher. The area on the back of the plate which corresponds exactly with the depression on the front (caused by scraping or burnishing) is located with callipers. This area is then beaten up with a hammer, or if a small area with a hammer and punch, until the front surface is level again. The marks of the correction are removed by rubbing them with snakestone and water, and the surface is restored with charcoal and finally polished with an oil rubber—a tightly bound roll of woolen cloth which, dipped in a little olive or sperm oil, makes an effective polisher.

Defects and Faults.—Defective areas and blemishes on plates

are, to a large extent, avoidable. Nevertheless, the possibility of error is considerable. The most commonly found defects are: (1) foul biting, due either to a badly laid ground or to a ground scorched during the process of smoking or to grease on an inadequately cleaned plate; (2) pitting, caused by impurities in the ground; (3) broken lines, made by irregular needling; (4) scratches, made inadvertently and all too easily.

Printing

To print from an etched plate, the ink has to be forced into the furrows, the surface area cleaned and damped paper pressed onto the plate with such weight as to pick the ink out of the bitten lines.

Ink.—It is possible to buy good printing ink in fresh condition. Alternatively, it can be made from ink powders. One-third of french black is mixed with two-thirds of frankfort black and pounded with a muller until, with the gradual addition of copperplate oil, all grit and lumps have been crushed into a smooth consistency. The ink should be of such a thickness that, if some is picked up with a palette knife, it will hang from the edge without quite dropping off. Ink is applied with a dabber or a felt roller. If the plate is warmed, the ink fills the lines more readily; on the other hand, if the plate is inked cold, the resulting impression has a richer quality. After inking, the plate is inspected thoroughly. Any lines which show a gleam of uncovered metal can be filled by rubbing ink into them with the finger.

Wiping the Plate.—Both hard and soft printing muslins are used to wipe the plate. It is convenient to fold the muslin into flat pads of a size which will lie comfortably in the hand. Hard muslin takes off most of the ink; the pad is swept with a firm, circular—but not scooping—movement over the surface. The soft muslin is used with scarcely any pressure and, when the metal looks clean and shining, it is wiped yet again, being brushed in the lightest possible way with the palm of the hand. Finally, in an operation known as *retroussage,* a piece of soft muslin may be trailed lightly over the plate to persuade a little ink from the furrows and produce an enhanced quality of line.

Paper.—Paper for proving can be of any quality; even blotting paper will serve. For final prints, only the best handmade papers are good enough. Sheets should be thoroughly damped, placed between blotting paper, laid between sheets of glass and left for about 24 hours before printing.

The Press.—The copperplate press usually consists of a flat metal bed which passes between heavy rollers. The rollers are turned by handles which operate either directly or through a gear. Pressure is increased or slackened by adjusting screws which bear down on packing at either end of the top roller. The plate is laid on blotting paper in the centre of the bed and overlaid, first with two thin fronting blankets, and then with two thicker felt blankets. The blankets are straightened and tensioned; their edges are stepped to make an easy approach for the roller; the handles of the press are turned over evenly and slowly, the blankets folded back and the print lifted from the plate by one corner. Prints are allowed to dry completely, then damped again and laid under a light pressure to flatten them. Defective prints are destroyed. Perfect prints are numbered 1/50, 2/50, etc. (the second number indicating the total of the edition). Proofs taken to reveal alterations and corrections are given the description of states and referred to as first state, second state and so on.

Soft-Ground Etching and Combined Techniques

The soft ground is made with ordinary etching ground melted into and thoroughly mixed with an equal quantity of tallow. It is applied evenly to the surface of a warmed plate with a leather roller. The orthodox method of drawing on the plate is to lay a thin sheet of slightly textured paper on the ground and make the design with a pencil. The soft ground is extremely impressionable; therefore, while drawing, the hand is not allowed to rest on the paper surface. When the sheet is peeled off, the ground will cling to the lines which have been pressed by the pencil point. The plate is then etched. The Dutch bath or iron perchloride is the best mordant for soft grounds. The method of printing is the same as that of an ordinary etching. The linear quality of soft-ground etching is precisely similar in character to grit-textured lines drawn with a soft pencil or chalk. The drawing is not necessarily made with pencil, nor is the overlaid paper indispensable. Any kind of point can be used to draw directly on the ground, and, indeed, any kind of texture can be pressed into the ground and etched in the image of its own impression.

Etchings and soft-ground etchings are freely used in conjunction with other intaglio techniques. Dry-point and line engraving are often combined with etching, and the laying of an etching ground and subsequent immersion in acid are the basis of the stipple, crayon and dot techniques of engraving. Aquatint (*q.v.*) is primarily an etching process with a rich tonal character which is frequently associated with the typically linear qualities of either soft- or hard-ground etching.

See also references under "Etching" in the Index volume.

(F. W. W.-S.)

Bibliography.—*History:* A. M. Hind, *A History of Engraving and Etching* (1927); British Museum (A. M. Hind), *Guide to the Processes and Schools of Engraving* (1914); H. W. Singer, *Die moderne Graphik,* 2nd ed. (1920); "Modern Masters of Etching Series," ed. by M. C. Salaman (1923–26); J. Laver, *History of British and American Etching* (1929); W. Shaw Sparrow, *A Book of British Etching From Francis Barlow to Francis Seymour Haden* (1926); M. H. Grant, *Dictionary of British Etchers* (1952); E. T. Chase, *The Etchings of the French Impressionists and Their Contemporaries* (1946); F. W. H. Hollstein, *Dutch and Flemish Etchings, Engravings and Woodcuts, ca. 1450–1700,* 10 vol. (1954), *German Engravings, Etchings and Woodcuts, ca. 1400–1700* (1954); Annual Catalogues, National Exhibition of Prints, Library of Congress, Washington, D.C.

Technique: F. Seymour Haden, *About Etching* (1878); W. P. Robins, *Etching Craft* (1922); E. Hesketh Hubbard, *On Making and Collecting Etchings* (1923); W. Silsby, *Etching Methods and Materials* (1943); H. Sternberg, *Modern Methods and Materials of Etching* (1949); Levon West, *Making an Etching* (1947); J. Buckland-Wright, *Etching and Engraving; Techniques and the Modern Trend* (1953).

ETERNITY, timelessness, or the state of that which is held to have neither beginning nor end. *See* Infinity and the Infinite; Theism; *see* also Eschatology; Immortality.

ETESIAN WIND, a remarkably steady southbound drift of the lower atmosphere over the eastern Mediterranean and adjacent lands in summer. From about mid-May to mid-September, with occasional breaks, the etesians dominate the Adriatic, Ionian and Aegean seas, and such adjacent countries as Greece, Egypt and Israel.

The name is derived from the Greek *etos* ("year") and is suggestive of the winds' regularity. They are of such significance to human activities that the ancient Greeks announced in the market places their expected beginning.

The winds, which reach maximum intensity in the early afternoon and may cease during the night, are part of the general inflow toward an intense area of low pressure usually centred over northwestern India in summer. They are not of the class generally termed monsoon winds. The etesians depart from the monsoon type in that they are practically rainless, are not accompanied by high relative humidities and are not replaced in winter by drifts from the opposite direction.

An extreme example of the constancy of the etesians is Cairo, Egy., where July winds blow from the northwest, north or northeast 98% of the time. Somewhat smaller percentages are typical of most of the region.

Rather analogous wind regimes and climates, which are called etesian climates and are characterized by dry summers and winter rainy seasons, are found in California, Chile, South Africa and southwestern Australia. *See* Wind: *Local Winds.*

(W. A. Bm.)

ETEX, ANTOINE (Tony) (1808–1888), French sculptor, painter and architect who executed the two groups of "Peace" and "War" placed at each side of the Arc de Triomphe, Paris, was born in Paris on March 20, 1808. He studied under the neoclassical sculptor F. J. Bosio at the École des Beaux-Arts. He designed the Ingres memorial at Montauban and painted a large allegory, "The Glory of the United States," for the city hall, New York city. Among his best-known architectural works are the tomb of Napoleon I in the Invalides and a monument of the 1848 revolution.

He died at Chaville, Seine-et-Oise, on June 14, 1888.

ETHANE is a colourless, gaseous compound of hydrogen and carbon, the second member, after methane, of the paraffin series of hydrocarbons. Its formula is C_2H_6. The second most important constituent of natural gas, it also occurs dissolved in petroleum oils and as a by-product of oil refinery operations and of the carbonization of coal. Structurally ethane is the simplest hydrocarbon containing a single carbon-carbon bond. This is shown by the structural formula $CH_3—CH_3$ and the alternative name dimethyl.

The industrial importance of ethane is based upon the ease with which it may be converted to ethylene, C_2H_4, and hydrogen. This transformation is accomplished by a process of pyrolysis or cracking, involving passage through hot alloy-steel tubes. Like propane, and to a lesser extent butane, ethane is a major raw material for the huge ethylene petrochemical industry, which produces many important products including polyethylene plastic, ethylene glycol and ethyl alcohol. By the 1960s, total world reserves of ethane were estimated at over 10,000,000,000 tons. Annual usage, almost entirely in the U.S., was about 0.2% of this figure. With over 90% of the ethane produced burned as fuel without separation from natural gas, the potential for future chemical production remained large.

Although the boiling point of liquid ethane is −127° F., ethane gas can be liquefied under pressure or at reduced temperatures and thus separated from natural gas. Unlike propane (*q.v.*), liquid ethane is not in common use as an industrial or domestic fuel.

See CHEMISTRY: *Analytical Chemistry; Methane and Ethane; see* also references under "Ethane" in the Index volume.

(H. G. DA.)

ETHANOLAMINES are organic compounds that can be considered as derivatives of ammonia in which the hydrogen atoms are replaced by the organic radical (CH_2CH_2OH). Mono-, di- and triethanolamine were first prepared by Charles Adolphe Wurtz (1860) by heating ethylene chlorhydrin with aqueous ammonia in a closed tube. It was not until Ludwig Knorr (1897) employed fractional-distillation methods that these products were separated. In 1927 a U.S. chemical corporation was granted a patent for the production of triethanolamine from ethylene oxide and ammonia. This process resulted in the formation of the coproducts monoethanolamine and diethanolamine in a definite ratio. The resulting mixture of the three ethanolamines is separated by distillation.

Uses.—The ethanolamines are widely used in industry, having some application in every manufacturing or processing industry. Several hundred use patents on triethanolamine and the closely related products monoethanolamine and diethanolamine were issued in the United States alone. A large volume of published literature appeared in both German and English.

The established uses for monoethanolamine are numerous and varied. In combination with fatty acids, it forms soaps which have the ability to emulsify mineral oil, waxes and resins in water. Such emulsions are extensively applied in the processing of textiles, to aid metal cutting and processing and in household specialties such as floor waxes. As a raw material, monoethanolamine is important in the manufacture of other industrial chemicals and pharmaceuticals. Other important uses involve monoethanolamine as an acid gas absorbent, in the manufacture of textile finishing agents and in the production of synthetic detergents and insecticidal sprays.

Diethanolamine is commercially important because of its strong affinity, in aqueous solutions, for hydrogen sulfide and carbon dioxide. Thus, it is used in the removal, recovery and concentration of carbon dioxide from flue and other waste gases. Textile specialties, such as finishing agents and dyeing and scouring assistants, are produced from diethanolamine and fatty acids. A radiopaque medium which permits exceptionally clear X-ray photographs for urology is produced from diethanolamine. A synthetic analgesic, Demerol, can be made from diethanolamine. This compound is comparable with morphine in its analgesic powers and has antispasmodic properties similar to those of atropine.

The use of triethanolamine as a corrosion inhibitor in automobile antifreeze solutions and high-temperature airplane-engine coolants is an outstanding commercial application of this compound. Large quantities are used in floor wax polishes where the emulsifying properties of triethanolamine soaps are advantageous. In the textile industry triethanolamine soaps are employed for many purposes, particularly in the preparation of mineral-oil emulsions for lubricating fibres and yarns. The ability of triethanolamine to combine with acids either to counteract the undesirable effects of acidic conditions or to produce new compounds results in applications which serve useful purposes in the manufacture of cosmetics, insecticides, metal-cutting oils, petroleum chemicals and cement-grinding aids.

Properties.—In having an ammoniacal odour and yielding typically alkaline solutions in water, the ethanolamines exhibit properties similar to other amines. The ethanolamines also possess some of the characteristics of alcohols but boil at higher temperatures than industrial alcohol. However, it is the amine portion of the molecule, through its ability to neutralize acids, that predominates and makes possible a great many commercial applications of the ethanolamines.

The chemical reactions of ethanolamines are similar in many respects to those of other organic nitrogen compounds. Thus, with acids they form salts or soaps; with esters, anhydrides and acyl halides they form substituted amides; and important derivatives are obtained by reaction with alkyl halides, aldehydes, ketones and carbon disulfide.

BIBLIOGRAPHY.—Emil J. Fischer, *Trianthanolamin und andere Athanolamine;* C. B. Kremer, "Ethanolamines," *Journal of Chemical Education,* vol. 19, pp. 80–81 (1942); Carbide and Carbon Chemicals Company, *Synthetic Organic Chemicals,* 13th ed. (1952).

ETHER (IN CHEMISTRY), any member of a certain class of substances (of which the well-known anesthetic, diethyl ether, commonly called "ether" or "aether," is one) composed of carbon, hydrogen and oxygen and having the general formula R.O.R′, where R and R′ are alkyl or aryl groups (*see* CHEMISTRY: *Organic*). The term ether formerly included the esters (*q.v.*) of organic acids, such as acetic ether, now termed ethyl acetate.

(X.)

Medical Uses.—(*See* ANESTHESIA AND ANESTHETICS.) Ether was still the most widely used general anesthetic in the early 1960s. This is because in ether anesthesia the various stages are well demarcated and the ratio between toxic and therapeutic levels is relatively large. Ether depresses neither the respiration nor the circulation until high concentrations are present in the blood. It is frequently employed to augment other anesthetics, such as nitrous oxide and cyclopropane. It is occasionally used rectally, after being mixed with oil, in prolonged and severe asthmatic attacks. Ether was used in the past for other medical purposes. Because its rapid evaporation causes intense cold, it was once used as a local anesthetic, but it has largely been replaced by other, more effective agents.

When ether is taken internally, the effect is quite similar to that of ingested alcohol.

(R. G. PE.)

Diethyl ether, $(C_2H_5)_2O$, the ether of pharmacy, is a colourless, volatile, highly inflammable liquid, of specific gravity 0.736 at 0° C., boiling point 35° C. and freezing point −117.4° C., with a powerful characteristic odour and a hot, sweetish taste; it is soluble in ten parts of water and in all proportions in alcohol; it dissolves bromine, iodine and, in small quantities, sulfur and phosphorus, also the volatile oils, most fatty and resinous substances, pure rubber and certain vegetable alkaloids. Mixed with ethyl alcohol it is an important solvent for nitrocellulose. A mixture of the vapour with oxygen or air is violently explosive. The making of ether by the action of sulfuric acid on alcohol was known about the 13th century, and later Basil Valentine and Valerius Cordus described its preparation and properties. The name ether appears to have been applied to the drug only since the time of F. G. Frobenius, who in 1730 termed it *spiritus aethereus* or *vini vitriolatus.* It was considered to be a sulfur compound, hence its name sulfuric ether; this idea was proved to be erroneous by Valentin Rose about 1800. Ether is manufactured by the distillation of five parts of 90% alcohol with nine parts of concentrated

sulfuric acid at a temperature of 127°–140° C., a constant stream of alcohol being caused to flow into the mixture during the operation. The distillate is purified by treatment with lime and calcium chloride and by subsequent distillation.

The presence of even small amounts of water or alcohol in ether can be shown by the continued evolution of hydrogen gas on treatment with metallic sodium. Chromic acid oxidizes ether to acetaldehyde, acetic acid and ethyl acetate, the proportions depending upon the experimental conditions. Ozone oxidizes it to a mixture in which acetaldehyde, hydrogen peroxide and organic peroxides have been identified. Peroxides are also formed in the presence of air under many conditions of storage. This action is inhibited by certain metals, notably copper. In contact with hydrogen iodide at 0° C. it forms ethyl iodide, and with water and a little sulfuric acid at 180° C. it yields alcohol. It forms crystalline compounds with bromine and with many metallic salts. Ether may be transported in iron drums, glass bottles or tin cans. Its principal use is in the manufacture of smokeless powder, in organic synthesis, as a solvent, in analytical chemistry and for medicinal purposes. No flames or sparking electrical equipment may be used in connection with the industrial application of ether.

Absolute ether is ether from which water, alcohol and acidic impurities have been removed to the greatest practicable extent. It is prepared from the ether of commerce by washing with a saturated aqueous solution of calcium chloride, then treating with sodium until the evolution of hydrogen ceases, and finally distilling from the excess of sodium.

Ether is used as a solvent in various organic syntheses. *See* GRIGNARD REAGENTS.

Other Ethers.—*Dimethyl ether*, $(CH_3)_2O$, first obtained by J. B. Dumas and E. Peligot, 1835, is best prepared by heating methyl alcohol and sulfuric acid to 140° C. and leading the evolved gas into sulfuric acid. The sulfuric acid solution is then allowed to drop slowly into an equal volume of water, so that the methyl ether is liberated (E. Erlenmeyer and A. Kriechbaumer, 1874). It is a pleasant-smelling inflammable gas, condensing to a liquid which boils at −23.6° C. It is somewhat soluble in water and readily soluble in alcohol and concentrated sulfuric acid. It combines with hydrogen chloride to form a compound, $(CH_3)_2O.HCl$. Methyl ethyl ether, $CH_3.O.C_2H_5$, prepared from methyl iodide and sodium methoxide, is a liquid boiling at 10.8° C. The homologous ethers are also liquids, with boiling points rising with increase of carbon content.

Diisopropyl ether, $(CH_3)_2CH.O.CH(CH_3)_2$, is a liquid boiling at 68° C. It is made from propylene, a by-product of the cracking process (*see* PETROLEUM). To some extent it has displaced diethyl ether as a solvent, and it finds some use in high-octane motor fuel.

Cyclic Ethers.—Ethylene oxide, CH_2CH_2O, boiling at 14° C., is used as a fumigant and also as a reagent in organic synthesis. Dioxane, $C_4H_8O_2$, boils at 101° C. and is widely used as a solvent.

Preparation.—The true ethers are formed by elimination of one molecule of water from two molecules of the alcohols; the two hydrocarbon radicals are the same in simple ethers, and different in mixed ethers. They may be prepared by the action of concentrated sulfuric acid on the alcohols; alkyl sulfuric acids are first formed and yield ethers on being heated with alcohols. The process is rendered continuous by running an alcohol slowly into the heated reaction mixture of alcohol and sulfuric acid. Benzene sulfonic acid has been used in place of sulfuric acid (F. Krafft, 1893). A. W. Williamson explained the mechanism of this action in 1850; in 1851 and 1852 he prepared diethyl ether (*see* above) by the action of sodium ethoxide on ethyl iodide, and showed that all ethers possess the structural formula given above. They may also be prepared by heating the alkyl halides with silver oxide. Hydrogen halides convert them into alkyl halides. With chlorine they yield substitution products.

See also references under "Ether (In Chemistry)" in the Index volume. (X.)

ETHER (IN PHYSICS), a hypothetical substance filling all space, inclusive of those volumes occupied by ordinary matter, and serving to transmit those forces (gravitational, electric, magnetic) which one material object exerts on another located at a distance. During the 19th century the ether hypothesis was accepted by all competent authorities, although there was diversity of opinion as to the ether's properties. After 1900 the opposite opinion gained ground; namely, that the hypothesis is unnecessary for the explanation of any observed phenomena. A few physicists still hold to the older view that an ether must exist and that otherwise philosophical difficulties would arise in connection with the concept of action-at-a-distance.

This radical change of attitude in regard to the ether was brought about by Albert Einstein, who showed that many of the properties formerly ascribed to the ether can equally well be ascribed to space and time.

Traditionally, space has been conceived as an unchangeable, passive constituent of the universe, neither affecting nor being affected by the dynamic changes occurring in the material parts of the universe. On this view, space is analogous to a moving-picture screen which remains unaltered even when the most violent scenes are projected upon it, and neither aids nor hinders the action, though its presence is necessary for the perception of the pictures. If space is unchangeable, in this sense, then it is necessary to assume the existence of a more active medium, occupying all space, and taking a dynamic part in the motions and other phenomena of the universe.

Twentieth-century physicists agree with their predecessors on this point. In fact, there has been no discontinuity in the chain of reasoning, and the Einstein theory of space and time rests on foundations that were laid in the distant past. The present conception of a dynamic space-time continuum has evolved gradually out of the original conception of a pair of continua: passive space and dynamic ether.

The Development of the Ether Theory from Descartes through Huygens.—Many ancient philosophers stressed the necessity of postulating an invisible intangible substance which takes a causal part in the motions of the planets, etc. The name "aether" or "ether" is derived from their writings. Some writers postulated many ethers, each occupying its own region in space. The idea of a single, all-pervasive ether may have originated with René Descartes (1638). In any case, Descartes exercised a dominant influence on all later physical theories of the ether.

The characteristic feature of Descartes' cosmology was its rejection of action-at-a-distance. Force was communicated only by contact, from one particle of matter to its immediate neighbours. As the sun warms and illuminates the earth, it was therefore necessary to assume that the space between the earth and sun is filled by some form of matter—the imperceptible ether. Light and heat were considered to be pressure, transmitted instantaneously from the sun to the earth. A visible object is seen, according to this theory, because the ether transmits a pressure from it to the beholder's eyes. The ether is thus conceived as a sort of blind man's stick.

Descartes' theory of light was shortly challenged from two directions. Pierre de Fermat (in 1657) doubted that light is transmitted instantaneously, and this was confirmed (in 1675) by Ole Roemer, who showed that the eclipses of Jupiter's moons gave definite evidence of the time needed by light to travel from Jupiter to the earth (*see* LIGHT, VELOCITY OF). Robert Hooke (1667) pointed out that Descartes had given no explanation of colour, and proposed the theory that light was an oscillatory motion of the ether, such that the particles of ether move back and forth in the same direction as that in which the light is being transmitted (longitudinal waves). Light of different colours is characterized by different rates of vibration (*see* OPTICS; SPECTROSCOPY).

Another investigator into the cause of colour was Sir Isaac Newton (1672). He became involved in a painful controversy with Hooke; perhaps because of this, or possibly because of uncertainty in his own mind, his writings on the ether problem are confused. Apparently Newton did not entirely reject the idea of an ether capable of longitudinal vibrations. These might even have something to do with light, but fundamentally he thought light consisted of streams of particles or corpuscles. These particles, emitted by the source of light and moving away from it

at a high speed, might interact with the ether and set up waves, much like a ship moving over the sea. Moreover, their emission was in some way connected with the vibrations of the luminous flame or other source of light. It cannot be said, however, that Newton really developed a coherent theory of light; however important his contributions to optics are, the pages on which he outlines basic concepts show indecision and a reluctance to commit himself to any one hypothesis.

This is also apparent in his treatment of gravitation and the ether, but his famous inverse square law is susceptible to precise formulation without using the picturesque terms of the ether theory. Consequently, all consideration of the ether is brusquely postponed indefinitely. Descartes' principle that forces result only from the contact of one material particle on another is not definitely rejected, and many of Newton's followers considered that the ether transmitted the force of gravity; no detailed theory of this process was ever developed, however.

In 1690, Christiaan Huygens published an explanation of the double refraction of light by Iceland spar. He developed Hooke's theory of light as a form of longitudinal wave motion to the stage where it could give quantitative explanation of additional phenomena. Newton was justly critical of this theory, in that it did not account for all of the known optical phenomena associated with Iceland spar.

The 18th Century.—The following century was one of great scientific activity. Many discoveries were made in the field of electricity and magnetism. The ether—even a multitude of ethers—was invoked by many writers to explain the phenomena, but no general agreement was reached.

Newton's corpuscular theory of light dominated the writings on optics. Thomas Melvill and Gaspard de Courtivron (1752) advanced the hypothesis that the colour of light was determined by the velocity of the corpuscules. This was disproved by observing that the satellites of Jupiter did not change colour at the moment of eclipse; thus different colours must travel with the same velocity.

James Bradley discovered the aberration of light (*q.v.*) in 1725; this is the apparent displacement of the stars because of the motion of the earth in its orbit. It can be explained by the corpuscular theory and is analogous to the slanting path of a raindrop down the window of a moving car. It can also be explained by the wave theory, although not so picturesquely. However, since all stars show the same displacement it follows that the velocity of light must be independent of that of its source. This is easily understood if light is a series of waves in a stationary ether, but it is difficult to see how corpuscules emitted by a moving body can fail to share in that motion.

The mediocre success of the corpuscular theory of light left the wave theory with some adherents, notably Benjamin Franklin, Jean Bernoulli (the younger) and the mathematician Leonhard Euler. The latter epitomized the current state of the theory in the words "light is in the ether the same thing as sound in the air."

The Triumphal Period of the Ether Theory.—The phenomena of the colours of thin films (such as soap bubbles or oil on water) and of the colours seen when light is passed through fine-meshed screen or cloth were known throughout the 18th century, but received little attention. The first triumph of the wave theory of light was Thomas Young's explanation, in 1800, of the effect of films on light. He reasoned that light may be reflected from both surfaces of the film. If the incident light consists of a single train of waves, the reflected light must consist of two trains moving through the same part of the ether. The one train will lag behind the other by a distance equal to twice the thickness of the film. Since the waves consist of alternate rarefactions and condensations, it may happen that the rarefactions of the reflected train coincide with the condensations of the other so that the two neutralize each other. The two trains are said to interfere destructively with each other. Whether or not this interference will occur depends on the spacing of the waves (their wave length) and on the thickness of the plate. If the colour of light is determined by its wave length, a thin film will not reflect light of some colours. Those colours which it does reflect are the only ones perceived by the eye. Young was able to develop this principle of interference into a complete quantitative account of the observed phenomena.

Young also endeavoured to apply his principle to the explanation of the coloured fringes which appear to surround fine fibres or wires—which is usually called diffraction. He was not successful, and this triumph was reserved for Augustin Fresnel (1815) who wrote several brilliant papers showing how the principle of interference when combined with Huygens' earlier theory explained this very complex set of phenomena.

The adherents of the corpuscular theory of light could not furnish any equally satisfactory explanation of diffraction, but the triumph of the ether theory was not yet complete. Newton's critique of the theory had been based on the phenomena of the double images seen through a crystal of Iceland spar. The phenomena had been investigated by others, and Etienne L. Malus (1808) had shown that when light is reflected from a window-pane, it will not pass through a crystal if the latter is held in certain positions. (This basic discovery has been applied in the manufacture of sunglasses.) This polarization of light, which prevents it from passing through a crystal in one position, but enables it to pass when the crystal is turned through 90°, has no analogue in acoustic phenomena. This, in essence, is Newton's criticism of the wave theory in the form summarized by Euler. A beam of polarized light is characterized by two mutually perpendicular directions: its direction of propagation and its direction of polarization. The relation of the latter to a crystal determines whether or not the light will be transmitted.

The solution of this problem was suggested to Young by the discovery (made by François J. Arago in 1816) that two beams of light will not interfere if they are polarized at right angles to each other. Instead of assuming that the ether particles vibrate back and forth in the *same* direction as the waves are traveling, as is the case in sound, Young assumed that the direction of vibration is at *right angles* to the direction of the wave propagation. This theory of transverse waves was first published in the *Encyclopædia Britannica* article on CHROMATICS, written by Young in 1817. With its aid, Young and Fresnel were ultimately able to give a complete explanation of the phenomena of polarization. This completed the triumph of the wave theory over the corpuscular theory of light, but new problems immediately arose.

The Solid Ether.—Previous to Young's investigations, it had been tacitly assumed that the ether was a fluid, more tenuous than air. How else could the planets move through the ether without encountering a resistance that would have brought them to rest long ago? Now, a fluid cannot transmit transverse waves; such vibrations require a degree of rigidity not possessed even by liquids. Thus Young's hypothesis amounted to supposing that the ether is solid. It eliminated one objection to the wave theory of light, but only to replace it by another. George G. Stokes (1845) gave a partial explanation of the manner in which the planets move through the solid ether. He pointed out that sealing wax and similar compounds are rigid so far as rapidly changing forces are concerned, but flow like liquids under long continued forces. Compared with the very rapid vibrations of light waves, even the motions of the planets are slow.

However, Fresnel and his immediate followers found much to be done, and postponed this question. The theory of a vibrating solid had not been worked out, and presented many technical problems. These can only be mentioned although they occupied a full generation of mathematical physicists.

One guiding principle governed all their researches, and raised interesting problems. The ether was conceived to be qualitatively identical with elastic solids such as glass or steel. Thus the theoretical developments could be guided by experiments with tangible solids, as well as by optical experiments with the otherwise intangible ether. The ether was assumed to differ only quantitatively from ordinary solids. Thus, it was shown that in order to account for the very high velocity of light (as compared with the velocity of sound in, say, steel) its rigidity and density must both be enormously much greater than that of steel. Estimates

of the density of the ether varied, but ran as high as 1,000 tons per cubic millimetre.' This accentuated the paradox of an intangible solid through which the planets move without resistance.

Another problem is presented by the fact that an ordinary solid can and does transmit both longitudinal and transverse waves. An experimental search for two kinds of light waves was made, but only the transverse waves were found. This ultimately led to the abandonment of the principle of qualitative similarity, and the recognition that the ether must be qualitatively different from ordinary solid matter. Thus a first step toward the ultimate identification of the ether with space itself was taken in the early 19th century.

The Electromagnetic Theory.—In accordance with Descartes' views, not only light, but gravitational, electric and magnetic forces must be transmitted through an ether. In order to leave open the question of the number of ethers required, that ether which transmits light was often called the luminiferous ether. During the 18th century, the ether hypothesis was most often invoked to explain the electric and magnetic forces. Thus, Henry Cavendish, in 1771, wrote a paper entitled "An Attempt to Explain . . . Electricity by Means of an Elastic Fluid." Since writers on optics, electricity and magnetism usually concentrated on the problems of a single field, it is often not easy to determine whether they assumed that there were many ethers or only one. Occasionally a writer, such as Joseph Priestley, hinted at the next great development in the ether theory, which was the recognition that a single ether sufficed for the three groups of phenomena. The question of a gravitational ether remained in the limbo to which Newton had consigned it.

During the earlier phases of investigation, electric and magnetic quantities were measured in similar but quite independent units. The possibility of a single system of units was brought about by Hans Christian Oersted's discovery (1820) that an electric current produces a magnetic field, and the discovery by Michael Faraday and Joseph Henry (1832) that a changing magnetic field produces an electromotive force. Two such complete systems of measurement were established, the one based on the original magnetic, the other on the electric system. Any quantity, such as electric current, could then be measured in either of two units. The ratio of the units of electric current was called c and was found to be equal to 3.1×10^{10} cm./sec. This constant became the centre of a remarkable development.

The experimental investigations initiated by the discoveries of Oersted, Faraday and Henry were supplemented by many theoretical investigations, and by 1854, William Thomson (later Lord Kelvin) was able to calculate the velocity with which telegraph signals were propagated along submarine cables. In doing so, he had to introduce constants whose values depended on the materials of which the wire and its insulating cover were made, the way in which it was supported, etc. Gustav R. Kirchhoff, in 1857, set himself a simpler problem, supposing the wire to be bare and suspended in free space. He was able to show that the velocity with which an electric signal would be propagated along such a wire is simply equal to $c = 3.1 \times 10^{10}$ cm./sec.

In itself, this is not surprising. The remarkable thing is that the velocity of light is also 3.1×10^{10} cm./sec., to within the limits of the accuracy of the measurements previous to that time. Kirchhoff recognized that this could not be a coincidence, and thus the way was opened for a unification of the optical and electromagnetic theories. (*See* LIGHT, VELOCITY OF.)

However, light is not propagated along wires, and the question remained: are electromagnetic waves propagated through free space? If so, the identity of optical and electromagnetic phenomena can be considered as proven. This technical problem of proving that electromagnetic waves are transverse and can be propagated with the velocity of light through free space was solved by James Clerk Maxwell. He was able to show that longitudinal electromagnetic waves cannot travel through transparent substances or free space, thus establishing the superiority of the electromagnetic theory of light over those theories based on an ether qualitatively similar to ordinary elastic solids. Finally, he was able to show that all electromagnetic and optical phenomena could be explained by a single system of stresses in the ether, which obey quite different laws than do the elastic stresses in such substances as steel.

While the principle of qualitative similarity was thus no longer tenable in its simplest form, attempts were made to reconstruct it in a more complicated form. Thus Lord Kelvin (1887) suggested that the ether might be a fluid in rapid motion—filled with minute vortices. A similar theory had been suggested by Jean Bernoulli more than a century before. Such a vortex ether had some of the properties needed to support the Maxwell stresses.

The experimental verification of Maxwell's electromagnetic waves by Heinrich R. Hertz, and their application to wireless telegraphy by Guglielmo Marconi, are worthy of mention, although not strictly relevant to the ether theory.

After Maxwell, there was no further mention of a plurality of ethers, and physicists were convinced that a single ether sufficed for the transmission of all known forces, including the gravitational. But a detailed ether theory of gravitation was not forthcoming.

The Ether Drag Experiments.—The manner in which solid objects move through the ether was the subject of much experimentation and more speculation. It has already been noted that this problem was especially acute during the period in which the ether was conceived to be a solid, but it continued to be studied for more than a century.

Young had pointed out that the observed aberration of light could most easily be fitted into the wave theory by supposing that the earth moves through the ether without disturbing it, or imparting any motion to it. However inexplicable such a state of affairs might be, it was at least simple.

It was also necessary to suppose that the ether permeated all matter, otherwise how could light be transmitted through transparent substances? Moreover, the density of the ether inside glass or water would have to be greater than in free space—otherwise the phenomenon of refraction could not be explained. Fresnel considered how a material object could move without setting the surrounding ether in motion, and yet allow the ether inside it to be denser than its surroundings. This is possible only if the ether inside the moving object also is in motion, but has a velocity which is less than that of the object itself. Ether flows into the front of the object and out at the rear in such a way that it is compressed while inside. The ratio of the ether's velocity to that of the object became known as the Fresnel drag coefficient, and can be calculated from the index of refraction of the object. This motion of the entrained ether affects the velocity with which light is transmitted. This was investigated experimentally by Armand H. Fizeau (1851) who measured the velocity of light in a stream of water. His results confirmed Fresnel's theory.

This relatively satisfactory solution of the problem was upset in 1887 by the famous Michelson-Morley experiment (*q.v.*). If the ether surrounding the earth does not share the earth's motion, then the velocity of light relative to the earth should depend on the direction of the ray. If the light is transmitted in the direction of the earth's motion, the earth will partially overtake it; hence in this direction light will seem to have a lower speed than in any other. In 1881 Albert A. Michelson had designed an experiment to test this conclusion, and with E. W. Morley's aid it was carried out. The result was negative: light appeared to travel in all directions with the same speed. It seemed necessary to suppose that the earth carried the ether with it, having imparted its full velocity at least to the ether in its immediate vicinity.

This was in direct contradiction with Young's explanation of aberration. Even worse, Sir Oliver Lodge soon after (1892) performed an experiment whose simplest interpretation was also in conflict with the idea of an ether set into motion by nearby matter. Lodge reasoned that, if the earth imparts its velocity to the ether, then the rotation of a heavy steel disk would set the ether into motion. This ether motion must in turn affect the propagation of light in a detectable manner. But on performing the experiment, Lodge was forced to conclude that the disk im-

parted less than 1/200 of its velocity to the surrounding ether. If the same fraction applied also to the motion of the earth, the Michelson-Morley experiment would have yielded a positive result.

The first attempt at a theoretical resolution of these conflicts ignored Lodge's experiment. Hendrik A. Lorentz recalled that Stokes had proposed an alternative to Young's explanation. Stokes (1845) showed that if the earth imparted its full motion to the ether in immediate contact with its surface, and if the surrounding ether moved in a certain way, then the phenomenon of aberration could still be explained. Lorentz pointed out that Stokes' theory reconciled the two experimental results, but that the kind of motion required of the ether was impossible unless the ether was very compressible. Max Planck (1899) examined this suggestion in detail, assuming that the earth carried along an atmosphere of ether in the same way it carries along an atmosphere of air. The difficulty of conceiving an ether which is at once very compressible and very rigid could be avoided as already suggested by Stokes (*see* above).

However, even this explanation did not survive experimental test, for it predicted that the velocity of the ether at a height above the earth would be different than immediately at the earth's surface. Michelson made a direct comparison of the velocity of light at a height of about 50 ft. with that a few feet beneath the earth's surface, and found no difference. The Michelson-Morley experiment was repeated from a balloon and the same negative result was found.

Many variants of these experiments were performed without clarifying the paradoxical situation. The Michelson-Morley experiment was also repeated with increased precision. In 1925–26, D. C. Miller repeated it both near sea level and on Mt. Wilson (6,000 ft. altitude). He believed that his experiments showed a small velocity of the ether relative to the earth. (In 1955, a group of physicists headed by R. S. Shankland restudied Miller's original notebooks and concluded that he had overestimated the accuracy of his results, and that the most probable value of this velocity is zero.)

Working at the same place as Miller, and also in 1926, R. J. Kennedy repeated the experiment, without detecting any relative motion of the earth and the ether. Miller's results attracted considerable public attention, and he endeavoured to draw various conclusions from them. His theories failed to account either for aberration or for the result of Lodge's experiment.

It will be recalled that this and the experiments on aberration are most readily explained by assuming the ether to be stationary and not set into motion by nearby matter. George Francis FitzGerald (1892) suggested that the Michelson-Morley experiment might also be consistent with this hypothesis, although at first glance it contradicts it flatly. The simple theory of the experiment assumes that the apparatus is constructed of perfectly rigid material and does not change its shape when rotated from one position into another. He suggested that motion through the ether might cause solid objects to contract in the direction of motion, and that this contraction might be of just the proper amount to bring about the negative result of the experiment.

This contraction hypothesis has an unpleasant *ad hoc* character; it was relieved of this onus by the investigations of Lorentz (1905), who showed that it was almost a logical consequence of Maxwell's electrodynamic equations. When these are supplemented by a very simple postulate, they predict that an object moving through the stationary ether will contract by just the necessary amount to explain the negative result of the Michelson-Morley experiment. In addition, they predict that a moving clock will run slow.

In view of these theoretical and experimental results, it became customary to speak of "a conspiracy of natural laws" to prevent any observation of motion relative to the ether.

The Principle of the Relativity of Motion.—It remained only to recognize that this interplay of natural laws was not accidental, but itself the consequence of a more general law. In 1905, Einstein recognized this more general law as the principle of the relativity of motion. In its most general form, this reads as follows:

> The motion of a single object cannot be observed experimentally; only changes in the spacial relations of two or more material objects can be observed.

In a lecture delivered in 1904, H. Poincaré had already formulated this principle; his lecture did not receive widespread notice, and was forgotten for some years.

A restricted form of this principle had been accepted as valid since the time of Newton, and is even common knowledge among the general public. Most people have been in a railway car drawn up beside another in a station, and have experienced momentary confusion when one of the two began to move. Only by glancing at a building or some other third object is it possible to be certain which train has started.

One consequence of the old principle of qualitative similarity was that the ether could serve as such a third material object, despite its intangibility. The negative results of the many modifications of the Michelson-Morley experiment indicated that the ether was not a material object in the sense of the principle of relativity.

However, the many lines of evidence for an all-pervasive something which takes part in the dynamic processes of the universe remained valid. While it was not immediately recognized that space itself has some of the properties formerly ascribed to the ether, Einstein and Poincaré recognized that an acceptance of the principle of relativity in its most general form had implications for all natural laws, including those of geometry.

To understand this, it is necessary to make a brief review of the development of geometric theory. This began in earliest history, and there is every reason to believe that at first geometry bore the same relation to the practical affairs of the architect and engineer as, say, the theory of electricity does today. Even in the highly academic writings of Euclid (300 B.C.) there is internal evidence of this mundane origin. Later writers, puzzled by the remarkable logical rigour of Euclid's theorems, advanced the idea that the concepts of space and time are a congenital part of human reason. The denial of any of the "self-evident" axioms of geometry would, according to this view, result in demonstrable contradictions.

This a priori position of geometry was rendered untenable by the investigations of Nikolai Lobachevski (1826) who showed that it was possible to construct logical systems analogous to Euclidean geometry, based on sets of axioms which are incompatible with the "true" axioms of geometry.

At first, these non-Euclidean geometries were merely mathematical curiosities, but they gradually forced a return to the older conception of geometry. The decision as to the "true" axioms was not a matter of pure logic, nor were all of the axioms of geometry "self-evident" (*see* GEOMETRY). Perhaps it was Einstein who first clearly formulated the view that the axioms of geometry are generalizations based on experiments with solid objects, and as such, are subject to reformulation in the light of more refined experiment.

In the same way, geometrical concepts are abstractions from reality, just as are the concepts of electrical theory, etc. The primary concept of Euclidean geometry is that of a rigid body, which can be moved without changing its shape. It was from this point of view that Poincaré and Einstein approached the Lorentz-Fitz-Gerald contraction hypothesis and its implications for geometry. A less obvious concept of Euclidean geometry is that of a signal which can be transmitted instantaneously from one place to another. A study of the ether drag experiments led Einstein to the hypothesis that no signal could be transmitted with a velocity greater than that of light. This necessitated other revisions of geometry and also of the theory of time. Ultimately, it was possible to construct a single theory which embraced all of geometry, kinematics, mechanics and electromagnetic theory.

The achievement was similar to that of Maxwell, who had welded the previously separate theories of electricity and magnetism into a single logical unit. Perhaps not surprisingly, Maxwell's electromagnetic theory was the only one of the four that could be included in Einstein's system without critical revision.

Einstein's Theory of Gravitation.—It has repeatedly been

noted above that the theory of gravitation had resisted all efforts to make it an integral part of the theories whose development has been traced. Neither had any major contribution to our knowledge of gravitation been made since Newton's time. The necessity for a revision of the Newtonian theory was now acute: for it implied that the gravitational forces were transmitted instantaneously. This was a flat contradiction to Einstein's 1905 hypothesis of a maximum signal velocity. While this was at once apparent, it was not until 1916 that Einstein constructed a detailed theory which embraced gravitational as well as electromagnetic forces.

It is characteristic of the force of gravity that it produces the same acceleration in all objects, large or small, light or heavy. This has been embodied in the apocryphal tale of Galileo Galilei and the leaning tower of Pisa. It was the subject of Evangelista Torricelli's famous experiment showing that a feather and a coin fall side by side in a vacuum. Einstein generalized it into the proposition that it is impossible to distinguish between an acceleration and a gravitational field. With the advance of aviation, this principle has become part of ordinary language, and it is common to hear such statements as, "The pilot weighs half a ton as he pulls his plane out of a power dive." Einstein's principle of equivalence asserts that this is neither metaphor nor exaggeration. The theory of gravitation is part of geometry and kinematics.

The actual incorporation of this principle into the theory of space-time required abstract logical reasoning of considerable complexity. The basic mathematical investigations were already complete, however. They had been begun by Carl F. Gauss (1821–1848) and carried through by Georg F. Riemann (1854). It is no disparagement of Einstein's achievement to emphasize the continuity of his work with that of his predecessors. They had been tentatively developing the idea that the properties of space are somehow determined by its material content of planets, houses, men. It was a major advance to assert definitely that the local properties of both space and time are direct consequences of the existence of nearby matter, and that, in turn, these properties have a causal influence on the motions of that matter. The development of this assertion into a detailed scientific theory was an even greater step.

This theory of gravitation has been confirmed experimentally. For a long time, a certain small irregularity in the motion of the planet Mercury (the precession of its perihelion) had defied explanation. The modified theory accounted perfectly for it.

Furthermore, the new theory predicted that light rays would be curved when passing close to a large object such as the sun. This has been confirmed by observations of the stars seen near the sun during a total eclipse. (*See* SPACE-TIME; RELATIVITY.)

The Photon Theory.—It might be satisfactory to some if this partial account of man's attempt to understand the universe could be closed with an implication of success and completion. The successes of the project are obvious, but it is not yet completed.

The history of the ether theory is essentially that of the rise of the wave theory of light and the decline of the rival corpuscular theory. Even while the former was being completed, new experimental discoveries gave the corpuscular theory a new vitality. The discovery of the photoelectric effect by Philipp Lenard (1904) and of the change of wave length of X-rays on scattering (Arthur H. Compton, 1923) unexpectedly disclosed new properties of light. Had these been known at the time of Fresnel, it is doubtful whether the wave theory of light could ever have won adherents.

It was Einstein who pointed out that these newly discovered properties could easily be explained by the corpuscular theory, but would be very difficult to reconcile with the wave theory.

Detailed theories were developed by Erwin Schrödinger, Werner Heisenberg and Paul A. M. Dirac (1925) but have no direct bearing on the ether theory and therefore do not come into the present considerations. (*See* QUANTUM MECHANICS.) It is certain that they, also, are only the first stage of a further evolution of physical theory. *See* also ELECTRICITY; ELECTROMAGNETIC WAVES; RADIO; LIGHT; RELATIVITY; QUANTUM MECHANICS.

See E. T. Whittaker, *History of the Theories of the Aether and Electricity*, 2 vol., 2nd ed. (1951). (C. ET.)

ETHEREAL OIL: *see* ESSENTIAL OIL.

ETHEREGE, SIR GEORGE (*c.* 1635–1692), English dramatist, was the creator of the Restoration comedy of manners. His grandfather, George Etherege, a vintner, was one of the original Bermuda adventurers, and his father, Capt. George, lived for some time in the Bermudas. The dramatist was the eldest of seven children. He may have been born at the home of his grandfather in Maidenhead, Berkshire, *c.* 1635, and probably accompanied his father to France during the Great Rebellion, as he is said to have traveled in that country in his youth. His father died there in 1650. About 1653 his grandfather apprenticed him to an attorney in Beaconsfield, Buckinghamshire.

Etherege's first comedy, *The Comical Revenge; or, Love in a Tub,* was brought out at Lincoln's Inn Fields theatre in March 1664. An immediate success, it is said to have brought £1,000 to the company in a month. Its novelty lay in its exploitation of contemporary manners, especially in the intrigue of the stylish Sir Frederick Frollick, but it still followed earlier tradition in containing a romantic plot, in heroic couplets and blank verse, and a farcical sub-plot.

Its success gave Etherege the entrée into the world of fashion where he became the boon companion of the literary rakes Sir Charles Sedley, the earl of Rochester and the earl of Dorset. He was thereafter involved in several escapades, the most notorious being a riot at Epsom, Surrey, in 1676 when one of the participants was killed. *She wou'd if she cou'd,* Etherege's second comedy, was staged at Lincoln's Inn Fields theatre on Feb. 6, 1668, but failed because of poor acting. It was the first comedy of manners to attain unity of tone by shedding the incongruous romantic verse element.

From 1668 to 1671 Etherege was in Turkey as secretary to the English ambassador, Sir Daniel Harvey. His account of conditions and personalities there is remarkably well-informed and makes entertaining reading. After his return he wrote the prologue for the opening of the new Dorset Garden theatre on Nov. 9, 1671. There his last and wittiest comedy, *The Man of Mode, or, Sir Fopling Flutter,* was produced with acclaim in March 1676.

About 1679–80 Etherege was married to Mary Arnold (née Sheppard), a rich widow. His knighthood is said by contemporary gossip to have been purchased for his marriage or obtained as a result of it. The marriage was unhappy, and the antiquary William Oldys says that Etherege had a daughter, who died young, by Elizabeth Barry, the actress.

Etherege was appointed envoy to the diet in Ratisbon in 1685. His two Letterbooks from there (in the British museum, London, and the Harvard Library Theatre collection) include personal as well as official correspondence, and introduce him as a man. The ceremonious life and stolid Germans bored and exasperated him, and he made little attempt to adapt himself to them; instead, he sighed for London diversions and scandalized his colleagues by flaunting a liaison with a strolling actress. He made an enemy of the powerful Austrian envoy, the Count de Windisgratz, and of his own secretary, Hugh Hughes, who wrote home malicious reports of his character and conduct. But although tactless and irresponsible, Etherege showed qualities of astuteness and loyalty, and he stayed at his post until 1689, when he followed James II to Paris. There he died on or about May 10, 1692.

Known to his friends as "easy" and "gentle," Etherege prided himself on his laziness. Nevertheless, he had a relish for life and a shrewd knowledge of men. His invention of a new type of high comedy was due to his ability to project on the stage, with a fine eye for telling detail, an ideal comic subject—the fashionable world of his time. Though the comedy of manners had been foreshadowed by Caroline dramatists, Etherege brought it to fruition in the setting of the amoral, cynical, gay and frivolous society of the Restoration. His characters, shorn of the last shreds of Jonsonian-type "humours," are individuals: the witty libertines, ineffable fops, high-spirited ladies and vain dupes whom he saw

around him. His finest creations, the heartless gallant Dorimant and the exquisite dandy Sir Fopling, set his contemporaries identifying their originals. Etherege's light-textured but subtly cadenced prose was an admirable vehicle for his dialogue, in which combats of wit and duels between the sexes predominated.

The brilliant style of comedy which he established was successfully cultivated by his successors, William Wycherley, William Congreve and Sir John Vanbrugh, and has persisted to modern times. His own plays, however, failed to hold the stage after the mid-18th century. Etherege also wrote a few light lyrics, letters in verse and lampoons. The love lyrics have a feeling for the value of the fleeting moment, and are among the most charming and gracefully turned of their day.

BIBLIOGRAPHY.—*Works*, ed. by A. W. Verity (1888), *Dramatic Works*, ed. by H. F. B. Brett-Smith (1927); *Letterbook of Sir George Etherege*, ed. by S. Rosenfeld (1928); E. Gosse, *Seventeenth Century Studies* (1883); B. Dobrée, *Restoration Comedy* (1924); D. Underwood, *Etherege and the Seventeeth-Century Comedy of Manners* (1957); D. Foster, "Sir George Etherege: Collections," *Notes and Queries*, 153 (1927); S. Rosenfeld, "Sir George Etherege in Ratisbon," *Rev. of Eng. Studies*, X (1934), "The Second Letterbook of Sir George Etherege," *Review of English Studies*, "New Series," II (1952). (S. M. R.)

ETHICAL MOVEMENT. The aim of the Ethical Movement is to assert the supreme importance of the moral factor in all the relations of life, the goal being a society embodying the moral values of justice, love, loyalty, brotherhood and peace. Devotion to the ethical ideal is central and quite apart from any theological, metaphysical or ceremonial considerations.

The Ethical Movement was initiated in the United States in 1876 by Felix Adler (*q.v.*). While completing his rabbinical studies in Berlin Adler was impressed by the works of Immanuel Kant, with whose philosophy he at the time identified himself, and by *Die Arbeiterfrage* of Friedrich Albert Lange. His first action on returning to America consisted, as he says,

> in founding among men of my own or nearly my own age a little society which we ambitiously called a Union for the Higher Life, based upon three tacit assumptions:—sex purity, the principle of devoting the surplus of one's income beyond that required for one's genuine needs to the elevation of the working classes, and, finally, continued intellectual development.

Adler later broke with the Jewish religion although his friends urged him to maintain outward conformity in order to adapt the truth of the old faith to modern conditions. But he felt that what was of real value in Jewish or any other religion needed to be "restated and fitted into a larger synthesis." He compared religious growth to the growth of a tree and said

> to expect that development should continue along Hebrew or Christian lines was like expecting that a tree would develop along one of its branches. There is a limit beyond which the extension of a branch cannot go. Then growth must show itself in putting forth a new branch.

The first actual Ethical society was inaugurated in New York city on May 15, 1876, and the bulk of it, in the words of Adler, "consisted of what would be called average people, especially of fathers and mothers who felt the need for themselves and their children of something to take the place of the consecrating influence of the old religions." This desire for a consecrating influence expressed itself in the initiation of very simple Sunday services, which were without prayer or ritual and were preceded by beautiful music to induce a suitable atmosphere, the most important feature being the address.

Adler's ideals as to the qualifications necessary for an ethical speaker were high: "that he shall give his whole life to the problems of ethical living, having no professional or business interests in competition with his dedication to these problems."

Societies in the U.S.—Members are not committed to any creed. They build their world view on the basis of modern knowledge, their program of action on the method of science. There is thus a substantial unity of outlook. The organization pioneered in many areas of social work. Each society has a weekly platform meeting, usually on Sunday morning, consisting of music, reading and an address. The societies conduct forums, classes, institutes, men's clubs, women's clubs, young people's groups and religious education programs for children. In 1945 the New York society set up an annual summer "encampment for citizenship," for young persons of all colours, classes, creeds and racial stocks to learn the practice of democracy by living and studying together.

The U.S. societies, numbering nearly 30, are united in the American Ethical union, with a total membership of more than 6,500. The union publishes the bimonthly *Ethical Outlook*.

British Ethical Societies.—The first English Ethical society, called the South Place Ethical society, was founded in 1888 by Stanton Coit, who had been attached to Adler's society in New York.

The Ethical societies of England were incorporated as the Ethical union in 1928, with the following main object:

> To promote by all lawful means the study of ethical principles; to advocate a religion of human fellowship and service, based upon the principle that the supreme aim of religion is the love of goodness, and that moral ideas and the moral life are independent of beliefs as to the ultimate nature of things and a life after death; and, by purely human and natural means, to help men to love, know, and do the right in all relations of life.

The official publications of the union are *Plain View* and *News Notes*.

The Secular Education league, founded in 1907, regards religion as a personal and private matter which all should be free to pursue in a private capacity.

International.—In 1952 the Ethical union of England, the American Ethical union and the Ethische Gemeinde of Vienna joined with Humanist organizations of many countries to form the International Humanist and Ethical union. The first meeting was held in Amsterdam in that year. The effort was to meet the demand for "an alternative to the religions which claim to be based on revelation on the one hand and to totalitarian systems on the other." The alternative was considered to be ethical humanism, a religious philosophy of life based on respect for man as a spiritual and moral being.

BIBLIOGRAPHY.—*Fiftieth Anniversary of the Ethical Movement, 1876–1926* (1926); Felix Adler, *Reconstruction of the Spiritual Ideal* (1924), *An Ethical Philosophy of Life*, rev. ed. (1946); H. J. Blackham, *The Ethical Movement During Seventy Years* (1946); David Saville Muzzey, *Ethics as a Religion* (1951); Henry Neumann, "Ethical Culture," ch. 22 in H. W. Schneider, *Religion in the Twentieth Century* (1948); Gustav Spiller, "The Ethical Movement" in J. Hastings (ed.), *Encyclopaedia of Religion and Ethics* (1908–25). (A. E. HN.; X.)

ETHICS (from Greek *ethos*, "character") is the systematic study of the nature of value concepts, "good," "bad," "ought," "right," "wrong," etc., and of the general principles which justify us in applying them to anything; also called "moral philosophy" (from Latin *mores*, "customs"). The present article is not concerned with the history of ethics (*see* ETHICS, HISTORY OF), but treats its general problems apart from their historical setting.

While the philosophical theory of ethics must be distinguished from the everyday task of making moral decisions, it has usually been held by moral philosophers that the chief test which can be applied to an ethical system is to ask if it can be harmonized with what is often called common-sense ethics; *i.e.*, with those ethical judgments which at our best we feel constrained to make, apart from philosophical argument, in our ordinary ethical thinking. It can hardly be maintained that common-sense ethics is infallible, but its relation to philosophical ethics may be compared to that between our ordinary perceptions and physical science. As the scientist must start with our perceptions of physical objects, so the moral philosopher must start with our common-sense ethical judgments, for he has no other data. Both will try to bring their data into a system, and in the process of doing so will have to amend some in order to make them coherent with others.

THE GOOD

By "the good" here is meant what is intrinsically good (or good-in-itself), not what is good only as a means to something else. More often than not, when we use the word good, we apply it, as with physical objects, to something which is good only as a means, but if nothing were really good in itself, nothing would be of value as a means, and to ask what things are good-in-themselves

seems obviously very relevant to the question what we ought to do. This is often expressed by asking what the supreme end of life is, but it is better not to put the question in this form, since it assumes that there is only one ultimate end, which may be doubted. Nor should we push the distinction so far as to fail to recognize that what is good as a means may also be good in itself or to treat the earlier stages of a complex process as mere means to the later stages (as if a whole book were written simply for the sake of its last paragraph). It is generally held that an inanimate object or a mere state or quality of such cannot be good in itself; what is good in itself must be a kind of experience or state of mind, but this is compatible with many differences of opinion as to the exact nature of the good.

Happiness as the Only Good.—The philosophers who have maintained that happiness is the only good and therefore the sole proper aim of action have generally made no distinction between happiness and long-continued pleasure. They have commonly been called hedonists (Greek *hedone*, "pleasure") or utilitarians. (*See* HEDONISM; UTILITARIANISM.) Some have held the doctrine in the egoistic form, according to which the sole ultimate valid reason for doing anything is that it is conducive to one's own greatest pleasure; others have put forward a theory of universalistic hedonism, according to which our duty is to produce the greatest total pleasure, our own share in this to be counted as neither more nor less important than any other man's. We must remember that the term "pleasure" is normally used by philosophers to cover all kinds of joy and satisfaction, not only the "lower" ones to which the term is most commonly applied in popular discourse, and that the major part of our duty in promoting the pleasure of others will consist in the noninfliction and removal of pain. The hedonist does not maintain that we normally ought to settle what we are to do by a direct calculation of the amount of pleasure and of pain likely to be produced but would admit that we can usually rely on the recognized general ethical rules such as those commanding honesty and justice, on the ground that the felicific value of these is shown by the past experience of the race.

Hedonism has been supported by a psychological doctrine that the only possible object of desire is pleasure, but this is generally rejected by modern psychologists. The element of truth in it is that there is a close connection between desire and pleasure, but the connection is not adequately expressed by saying that we desire only pleasure. It is true, indeed a tautology, that a man can only desire what pleases him in some way, but this is different from saying that he desires only pleasure itself, not that which gives pleasure. On the contrary, we could get little pleasure without desiring or at least valuing and liking for its own sake something other than pleasure.

Egoistic hedonism is not necessarily a selfish theory in the sense that it advocates external behaviour of a kind which would ordinarily be described as selfish, for the philosophical holders of the view have almost always insisted that a man must do what will benefit others if he is to obtain the greatest happiness for himself. It is indeed easy to show that most wrong actions are detrimental to the agent's own happiness as well as to that of others. But it can hardly be maintained that this is true of all, especially where the agent's life or health would be endangered by his doing his duty, unless we postulate reward or punishment in a future life. With the help of this postulate egoistic hedonism has been defended even by theologians, but it is generally felt that, insofar as a man is motivated solely by the desire for his own happiness, he is not really virtuous, however impeccable his outward conduct may be. And if the theory that we can desire only our own pleasure is abandoned, it is difficult to find any argument for egoistic hedonism. It cannot be refuted logically, but it certainly seems in conflict with our moral consciousness.

Universalistic hedonism is more plausible. It accords with the strong empiricist tendency in modern philosophy, pleasure being a definite empirically given feeling. It has also been commended as a simple theory, but it is certainly not at all simple to apply in view of the difficulty of measuring very different kinds of pleasure against each other. It is difficult to tell how far it agrees with common-sense ethical beliefs as to what we ought to do because of the great difficulty of estimating the relative pleasure value of the consequences of alternative actions and so of deciding in doubtful cases which action the hedonistic theory really supports; but it certainly disagrees with common-sense ethics in three respects. First, if consistent, it can acknowledge no distinction of quality between different pleasures, so that for it the pleasures of a pig are as good as those of an artist or philosopher, provided only that they are equally intense, durable and free from pain. If consistent, it must admit even that the state of mind of a sadist in torturing his victims, being pleasant, is as good in itself as an equally pleasant state of mind involving the most unselfish devotion and that it only has effects which are evil. Secondly, hedonism cannot attribute any intrinsic value to actions or motives, at least except insofar as these are pleasant, yet we cannot help valuing very highly deliberate acts of self-sacrifice although they are painful. We cannot attribute this merely to their results because there are many cases where we should value A's act in sacrificing something to B, yet despite its similar results, condemn B's act if he enforced the sacrifice against A's will. Thirdly, hedonism can only value the amount of pleasure produced and must value the same amount equally, however it is distributed. If the total sum of happiness could be ever so slightly increased by taking away all happiness from one man or even by plunging him into the extremest misery, it ought to be done on the hedonist view; if ever so little additional happiness could be secured by punishing an innocent man, he ought to be punished. It must be admitted that no strict disproof of hedonism is possible, but if these divergences from utilitarianism in common-sense ethics represent not only what people actually think but what they ought to think, they constitute a fatal argument against hedonism. It may well be the case that everything which is intrinsically good is pleasant, but it is much less plausible to maintain that its goodness consists in, or is in strict proportion to, its pleasantness.

Some thinkers would insist on making a sharp distinction between happiness and pleasure. They normally mean by happiness some state of feeling involving a qualitatively especially valuable kind of pleasure more durable than ordinary transient pleasures. But even when so understood, the term, if its distinctive meaning is not to be altogether destroyed, still has a reference exclusively to the feeling side of our nature; consequently the doctrine that happiness is the only good denies by implication any intrinsic value to the cognitive and conative side.

Perfection as the Good.—This view finds the good of a being in the fullest development of his faculties. Happiness would then be valued only as a sign that our faculties were functioning well, or at least as the feeling inseparable from their good functioning. The theory of perfectionism has sometimes (as apparently in Plato and Aristotle) taken an egoistic form, which is less hopelessly irreconcilable with common-sense ethics than is egoistic hedonism but is still open to the objection that, if one's own perfection is of value, other people's must be so also and it is impossible to see why the mere fact that a perfection belongs to somebody else rather than to me should exempt me from any obligation to produce it. It is not as if the sense of obligation were specially connected with our own good; on the contrary, it seems to require impartiality between ourselves and other men. But if our good consists in the development of our capacities, what are capacities? A capacity is not anything actual by itself but only something hypothetical. Apart from its actualization in behaviour and states of mind, it is nothing that exists in the proper sense of the word. So the view that our good is the perfection of our faculties reduces itself to the view that the good lies in certain states of mind and actions.

Pluralistic Views of the Good.—But now it seems impossible, hedonism having been discarded, to find a single formula which will cover all cases of intrinsic goodness. It has sometimes been said that moral virtue is the only real good, but it is hard indeed to hold that it does not matter at all whether the virtuous man is happy or miserable. It is also hard to avoid ascribing an intrinsic value above the pleasure involved to knowledge, to aesthetic experience, to love or to certain religious experiences. Nor is it

clear that there are not also goods of a more complex order dependent on the manner of the distribution of good or at least of the means to it. This is supported by the fact that a good man prefers a just distribution to an unjust even apart from its consequences. A further complication was pointed out by G. E. Moore in *Principia Ethica* (1903) when he laid down the principle of "organic wholes," according to which the value of a whole need "bear no regular proportion to the sum of the values of its parts."

Thus consciousness of a beautiful object may have great value, even though the object by itself and mere consciousness by itself have severally no or practically no value. A great drama or a great picture may contain elements which by themselves would be valueless or even worse than valueless and yet greatly enhance the value of the whole to which they belong; many experiences involving sympathy or courage owe much of their value to an element of pain and all pain is in itself bad. Moore's statement of the principle is not above criticism, but surely there are wholes the value of which depends not only on the elements included in them but on the interrelation of these elements.

THEORIES OF OBLIGATION

Under this heading is considered what makes an action obligatory. We may distinguish four types of answer to the question.

1. It may be answered by reference to the good produced. This position is known as utilitarianism, in a wider sense. Hedonistic utilitarianism is a variety but not the only variety of it, since hedonistic utilitarianism asserts the further proposition that the only good is pleasure and the only evil pain.

2. Reference may be made to universal laws known a priori independently of their consequences. The most famous example of this is provided by Kant's ethics. It is not indeed fair to say that Kant ignored consequences. What he meant was that we should take account of consequences in order to apply a moral law but not in order to establish the validity of the law or to make an exception of it. Kant tries to show that there is a kind of contradiction in denying certain general ethical laws. He insists that a man should ask, not what the particular consequences of his action will be, but what would happen if everybody acted in the way in which he proposes to act; and then argues that the general violation of certain ethical laws would be either impossible or self-defeating in the sense of thwarting the very purpose for which the law was violated. Thus he argues that, if everybody lied whenever it suited him, this would be self-defeating because the lies would never be believed. In this way Kant thought that he had shown it to be wrong to lie under any circumstances whatever, even in order to save life; but few philosophers have been prepared to go as far as that. Kant's second principle is more teleological and has had enormous influence: "Act so as to treat humanity always as an end and never only as a means." Kant's account is, however, of more value as an analysis of the nature of moral action, distinguishing it sharply from prudential action, than as a discussion of the criteria whether something is our duty or not. Yet there do seem to be a good many cases where Kant's criterion—"Act as if the maxim of your action were to become by your will law universal"—works better and more in accord with our moral convictions than does utilitarianism. There are many cases where we think it wrong to perform an act not because the consequences of the particular act are bad but because we consider what would happen if everybody acted as we propose to act. Scholastic writers on ethics also defend the absolute universality of some ethical laws based on the ground that our capacities ought not to be used in an unnatural way; *i.e.*, so as to frustrate the purpose for which they were intended.

3. It may be held that we see what we ought to do by intuition in each particular case. It would hardly be possible to hold that we could do so without taking consequences into account at all, since we can hardly be expected to see whether an action is right without knowing its factual nature and we can hardly separate the nature of an action from at least its relatively immediate consequences, but it has been held that there is no discernible logical connection between "good" and "ought" and therefore that we can never infer from the goodness of its consequences that an action ought to be done.

4. Another view recognizes certain a priori laws as holding independently of the good done by observing them yet regards them not as absolute laws but as instances of "prima-facie duties." By a "prima-facie duty" is meant one which holds only if it is not overruled by a superior moral obligation. Among prima-facie duties are included obligations to produce intrinsically good consequences, so that a large place is left for the factors recognized by the utilitarians. But other prima-facie duties such as the duty to keep promises and the duty to make reparation for harm done by oneself depend on the past and not on the future. Where there is a conflict between two prima-facie duties (*e.g.*, between the obligation to keep promises and the obligation not to hurt others), no general laws are suggested for the solution of the difficulty, but it is left to our admittedly uncertain intuitive judgment. If the harm is very great the rule, it is held, should be broken and if very small it should be kept, but we have no rule for measuring how great. This view has an advantage over that which relies on intuition in each particular case insofar as it admits a priori general laws in ethics without having to admit that it is never legitimate to act against any one of them.

It would be widely felt that this fourth view gives a better account than the others of the way in which we actually think in settling ordinary moral questions but that it falls short of utilitarianism in not providing what could be regarded as an ultimate rational explanation. It is objected that we cannot accept as a final account of ethics one which just leaves us with a heap of disconnected prima-facie duties to be accepted as self-evidently obligatory without any reason that can be adduced for declaring them to be our duties and that the concept of prima-facie duty, being that of a tendency to be a duty, must be logically posterior and not prior to the concept of absolute duty.

Against utilitarianism, the following objections are brought. In the first place, any act has a vast number of consequences which we cannot possibly foresee. To this, however, the utilitarian may reply that it is quite right to ignore these in making our decisions since we cannot tell whether they will be good or bad. Secondly, it is plain that no good man will do such things as lie or break promises merely because there is a very slight good to be gained by doing so, and it is futile to appeal to the effect of the bad example since the action may be done secretly, or to its bad effects on the agent's character because this already presupposes that the act is wrong. But the utilitarian may meet this by replying either (*a*) that, provided that obedience to a certain rule generally produces good results, it is safest to act on the rule even in cases where it does not seem to do so, at least where the harm is not very obvious and large, because there is so much risk of going wrong if we try to decide each case on its merits; or (*b*) that it is intrinsically good, for example, to keep promises or intrinsically bad to break them and this goodness or badness must be set against the goodness or badness of the results. Thirdly, if utilitarianism were true, we should be under as much obligation to produce pleasure for ourselves as for another man (pleasure being at any rate a good, if not the only good). Fourthly, utilitarianism leaves hardly any scope for actions which are not obligatory and yet permissible.

ANALYSIS OF ETHICAL CONCEPTS

There seem to be four alternatives here: an ethical concept may be analyzed in terms of other ethical concepts, or naturalistically (in terms of concepts of a natural science, including psychology), or metaphysically, or it may be declared unanalyzable. Obviously the first alternative cannot be fulfilled for all ethical concepts. A naturalistic analysis makes ethics a mere branch of a natural science, a conclusion which some regard as a strong argument for, others as a strong argument against, such an analysis. Examples are the definitions of "good" as pleasant, as what is desired by all, by most, or by some people, as what would ultimately satisfy desire, or as what is in accord with the main trend of evolution, and of "right" as that of which most people who look at it impartially (in a sense in which this term does not already presuppose

ethical concepts) would feel approval, or as that course of action which is conducive to social development of a certain kind or to social stability. Other naturalistic theories define "good" or "right" in terms of the feelings or desires of the speaker. In addition to the objections to specific definitions, which there is not space to give here, they all seem liable to the objection that they do not leave any room for the concept of "ought" as distinct from what is in fact the case. Not one of the characteristics equated by various naturalists with the concept of goodness seems to carry with it any real obligation. Nor, it is objected, can ethical conclusions be established in the way in which they should be if a naturalistic analysis were correct. No set of statistics about human psychological attitudes or of introspective psychological observations will prove that something is good or right. A naturalistic account seems to take away altogether the specific character of ethics. Naturalists have tried to support their view by theories which explain the origin of our ethical beliefs and attitudes from nonethical factors such as fear of punishment; but it must be remembered (1) that the psychology of young children and of the primitive savages among whom ethical beliefs originated is one of the most speculative of studies because we cannot obtain reliable introspective reports from young children and because there are no peoples now surviving who are at such a primitive stage as that in which men first developed ethical ideas; and (2) that though no doubt ethical ideas developed from a state in which there were no ethical ideas and so must have developed from nonethical factors, the effect need not be identical with its cause. Nobody would accept a similar argument about mathematical ideas. Metaphysical definitions of ethical concepts are not popular now and are liable to the same kind of objections as naturalistic ones in that they take away the unique character of ethics and try to reduce it to something else. They are also liable to objections of their own arising from the doubtful character of metaphysical theories, with the difficulty of making any specific deduction from them as to what is right. The alternative to accepting a definition of either sort is to regard at least one ethical term as indefinable in the sense of being irreducible to anything else. Such a view was put forward by Moore in regard to good (in the sense in which this stands for intrinsic goodness, good-in-itself). He called good a "nonnatural quality," meaning chiefly that it was quite different from any of the qualities recognized by natural science or given in nonethical experience. If good is indefinable, its nature and presence can only be discerned by immediate awareness, but there are many qualities in this position. Moore instanced yellowness, which can be recognized immediately, but not explained, to anyone who has not had the experience of seeing it, by analyzing it in terms of anything else. But many thinkers are reluctant to admit any immediate awareness which cannot as in the case of yellow be that of sensation, but must on a nonnaturalistic theory of ethics assume rather the character of an a priori intuition. Further, even philosophers sympathetic to Moore's general point of view are usually by no means sure that they are immediately aware of the simple indefinable quality of which he speaks when they recognize something as good. It seems more plausible to maintain that we are immediately aware of an indefinable relation of obligation.

This brings us to the question of the relations between different ethical concepts, of which the principal are "good" (in the sense of good-in-itself) and "ought." "Ought" has been defined in terms of "good," in which case "this ought to be done" becomes something like "this is most liable to produce good"; and "good" has been defined in terms of "ought," as being that which we ought to desire, pursue, welcome, approve, admire or, in general, regard with some favourable attitude. Alternatively, "good" and "ought" have both been thought to be indefinable. If "ought" is defined in terms of "good," utilitarianism must be accepted; but if "good" is defined in terms of "ought," utilitarianism may be accused of a vicious circle, explaining the "ought" by means of a good which itself has to be defined in terms of "ought." The concepts of "right" and "duty" are obviously essentially linked with that of "ought."

The Nature of Ethical Judgment.—Many philosophers think that both the naturalist and the nonnaturalist commit the same error of assimilating ethical judgments to assertions of scientific facts. The naturalist does so quite definitely, the nonnaturalist appears not to do so, but though he speaks of goodness as a nonnatural property it still is for him a property. He treats it as if it were just some extra characteristic of a peculiar kind which some things possess over and above the characteristics recognized by natural science, and the same kind of complaint may be made about the nonnaturalist view of obligation. It is assumed on these views that the point of ethical judgment lies in giving us information as to which things have this peculiar quality or stand in this peculiar relation. Against this is put forward the view that the function of ethical language is not to give information or to assert anything at all, but to express emotional and practical attitudes and to lead others to adopt such attitudes. Ethical expressions can, on that view, neither be true nor false, at least qua specifically ethical. They may indeed, it is admitted, include in their meaning some assertion about the state of mind of the person who utters them, but except insofar as they go beyond this they are not ethical. It follows indeed that, strictly, there are no ethical *judgments*.

To the objection that, if so, when A calls an action right and B calls the same action wrong, they are not contradicting each other, the reply is made that they are not saying what is logically contradictory but advocating conflicting policies and that an objective view of ethics confuses these two different senses of contradiction. To the objection that, if there is no possibility of true ethical judgments, there can be no scope for argument in ethics, it is replied that ethical arguments do not prove or even make probable any conclusion but perform the function of bringing people into a state of mind in which they will react in the way desired by the person who uses the argument. This, however, involves the admission that there is no such thing as a good reason in ethics, but only successful ways of persuasion; and all successful means adopted to induce a person to change his ethical attitude certainly would not commend themselves equally to moral philosophers.

Further, our ethical attitude seems essentially to involve the belief that what we approve or disapprove is objectively good or bad, at least in the sense of being a fitting object for approval or disapproval and not only an actual object of such. If the element of judgment, claiming objectivity, were really altogether removed, would not our ethical expressions be mere abuse like swearwords used in anger without any claim to rational justification or shouts or threats intended to browbeat people into doing what we want? The view fails to distinguish between actual and right approval. This is the case for an objective theory of ethics, at least as an analysis of our ordinary psychological attitude when we are dealing with ethical questions; *i.e.*, we do claim cognitive objectivity. It may still be said that the claim is unjustified, but in that case one has to make the very paradoxical admission that we are mistaken when we utter the most apparently obvious ethical judgments, for we are supposing them to be true when they cannot really be true.

The main reason why many thinkers are inclined to reject an objective view of ethics lies in the divergence of ethical opinions and the supposed impossibility of settling ethical disputes. A very large proportion of disputes as to what is right or wrong are indeed due merely to differences of opinion as to fact, especially regarding the consequences of proposed actions, and therefore cannot be used as arguments against the objectivity of ethics. But there remains a sufficient residuum of intractable purely ethical disputes. These may however be explained by differences in the experiences and psychology of different people so that the same external situation does not produce the same inward state; by misapplication of general principles or intellectual confusions of some sort; by one-sidedness (*i.e.*, failure to attend adequately to all relevant aspects of a situation); by the kind of causes with which psychoanalysts deal; or by sheer unwillingness to seek or admit the truth. All these circumstances can also lead to error in judgments as to matters of fact, but what has made ethi-

cal disputes a special source of perplexity is that they often seem to turn on appeals to rival intuitions, and it has been objected that there is no possible method of settling these differences. But if what presents itself as an intuition may arise either from genuine insight or from any of the causes of error just mentioned, it follows that there are on the contrary as many methods as there are means of removing these causes. Further, there is always the coherence test, which consists in asking whether an intuition can be fitted into a coherent system of ethics, or, less ambitiously, whether it is consistent with what the man who has it cannot help believing in other similar cases. When suitable methods have been applied, a formerly intuitive conviction may disappear or may be amended.

The difference between ethics and science may indeed turn out to be much more one of degree and less one of kind than is often thought. We cannot prove that all ethical disputes are not soluble in principle (could not be settled if the above methods were properly used) and we cannot claim that all scientific disputes are soluble in practice. Attention is naturally focused on the disputable points in ethics, but there are a vast number of ethical judgments, at least about the preferability of one action to another under given circumstances, concerning which no reasonable dispute is possible. And the difference from ethics becomes much less marked when we think not of a science like physics but of history or of astronomy, for example.

The appeal to "intuition" is a stumbling block, but if there are valid reasons for action at all, it must be possible sometimes to know or rationally think something to be a good reason for acting in a certain way without having to give a further reason why it is a reason, and this is all that need be meant by "intuition" in ethics. But intuition must be closely linked with and supplemented by reasoning and cannot be pronounced infallible. To it may be ascribed two different functions: (1) it may be held to tell us that some kind of thing (presumably an experience or state of mind) is good in itself or that some kind of action is a prima-facie duty, this in either case being taken to give an ultimate reason in favour of some action; (2) it may be held to tell us that a particular act is right or that a particular result is on the whole better or worse than another. The second type of intuition at least must be regarded as the terminus of a process of reasoning, explicit or implicit, even if it goes beyond what the latter proves.

A view intermediate between subjective and objective theories denies that we can make true or false ethical judgments but admits that we can still distinguish between ethically right and wrong attitudes and policies without analyzing the distinction naturalistically or merely in terms of people's actual attitudes. On that view what we describe as ethical judgments are not statements of fact of any sort (naturalistic or nonnaturalistic), nor are they merely expressions of emotion or exhortations. What they are is something unique, the nature of which may be illuminated by analogies with the product of other mental activities (*e.g.*, scientific knowledge, commands) but is not reducible to any of them. Ethics has criteria of its own distinct from but not inferior to those of science, so that good reasons can be given for ethical judgments, only these criteria will determine not the truth or falsity of judgments but the rightness or wrongness of attitudes and actions. On behalf of this view it has been urged that ethical judgments cannot be true in the ordinary sense of the word since they are not scientifically verifiable, but certainly "true" is constantly used more widely in ordinary speech than the argument suggests. It is the person who insists that the word be so limited and not his opponent who is introducing an innovation in the use of language. Even the scientist regards a priori mathematical propositions as true, though the criteria for their truth are quite as different from the criteria of empirical propositions as either are from the criteria of ethics. The distinction between different kinds of true judgment is not the same as that between different senses of "true."

MOTIVES AND SANCTIONS

Problems which moral philosophers discuss under this head include the conditions on which praiseworthiness or blameworthiness depends; the relative value of motives other than the specifically moral one (viz., the urge to do one's duty for duty's sake); one's responsibility for subconscious desires; the question whether the motive is relevant to the rightness or wrongness or only to the praiseworthiness or blameworthiness of an act; and the question in what sense it is true that a person always ought to do what he thinks he ought. The fact that blameworthy actions are held to "deserve" punishment raises further problems, as also does the fact that blameworthiness implies responsibility, this being hard to reconcile with our natural belief that everything is caused (*see* FREE WILL).

A question which used to be thought extremely important is "What are the 'sanctions' of morality?" If this means "Why should we do our duty?" it can only be asked through a misunderstanding, since it amounts to asking why we ought to do what we ought (a tautology); but it may be interpreted as a practical question, in which case it means "Because people are not always sufficiently inclined to do their duty just because it is their duty, what additional motives can we find for inducing them to do it?"

See also references under "Ethics" in the Index volume.

BIBLIOGRAPHY.—Plato, *Republic*, Eng. trans. by F. M. Cornford (1941); Aristotle, *Nicomachean Ethics*, Eng. trans. by Sir David Ross (1954); Bishop Butler, prefaces to Sermons i, ii, iii, and xi, in *Fifteen Sermons* (1726 and later editions); D. Hume, *Enquiry Concerning the Principles of Morals* (1751); Immanuel Kant, *Grundlegung zur Metaphysik der Sitten* (1785; Eng. trans. by H. J. Paton, *The Moral Law*, 1948); J. S. Mill, *Utilitarianism* (1863; ed. by J. Plamenatz, 1949); G. E. Moore, *Principia Ethica* (1903), *Ethics* (1912); E. F. Carritt, *The Theory of Morals: an Introduction to Ethical Philosophy* (1928); W. D. Ross, *The Right and the Good* (1930), *Foundations of Ethics* (1939); R. B. Perry, *A General Theory of Value* (1926); R. M. Hare, *The Language of Morals* (1952). (A. C. E.)

ETHICS, COMPARATIVE. What actions are right or wrong, good or evil? At different times and places, man's behaviour has been ruled by various answers to this query, and these are the province of ethics. While philosophical ethics studies questions of validity, comparative ethics examines the rules and ideals as in fact found at various times in different societies, and the part that they play in human affairs.

In advanced societies there are various types of social norm; *e.g.*, legal, religious, moral and conventional. But such distinctions develop slowly, and in earlier phases they are not evident. The raw material of comparative ethics comes from the working codes found in all societies as oral or written traditions, or as tacit assumptions underlying a framework of law and institutions.

The history of conduct cannot be studied directly, since there are no criteria for assessing how far conduct follows the rules. We can merely ask what is enjoined on the individual and what ideals, unattainable though they be, he is expected to pursue. The gap between ideal and working codes may be wide or narrow.

Such codes exist long before men begin to think about them. When reflection does set in, its findings are embodied in the maxims of the worldly wise, the sayings of religious critics and prophets, or the analyses of philosophers. All these may affect codes and ideals, and thus concern us here.

The Special Problems of Comparative Ethics.—Three questions arise: How do the various codes resemble or differ from one another? Is there ethical development, and if so, how does it link with changes in other fields, for instance in religion and law? Is there moral progress?

As to the first question, all stable societies have rules, upheld by the impartial onlooker and the general judgment of society, prescribing or forbidding definite classes of action and visiting a breach with reprobation if not with specific penalty. The rules cover the principal relations of social life. They deal with persons and property and with the relations between the sexes. Within the primary groups, extending beyond the simple family of parents and children, there is everywhere what E. B. Tylor called a "natural solidarity"—a measure of mutual forbearance, helpfulness and trust—giving rise to elementary duties supporting it in turn. All known codes insist on some measure of internal peace and order and tend to sanction whatever helps to satisfy social needs and to

ensure the group's survival, while censuring whatever disturbs the social peace.

Moral Elements Common to All Known Societies.—In considering more specifically how ethical codes resemble one another, the following elements common to all known societies may be distinguished:

Customs forbidding homicide within a more or less extended but well-defined sphere are universal. These may be enforced either by a system of public justice, or by blood revenge considered not only as a right but as a duty.

Likewise, there are rules forbidding bodily violence or injury to the person. In early stages these rules are upheld by the law of talion, an eye for an eye and a tooth for a tooth; or by compensation, graded according to the injury and the rank or status of the parties concerned.

Insults or attacks on honour and reputation are widely condemned though this too is influenced by status.

Property rights in some form exist everywhere. A detected thief suffers punishment or revenge, or at least must give back his loot. With regard to property rights the difference between friend and foe in war survives even today.

Some elementary duties of doing good and furthering the well-being of others are insisted upon in all societies. Within the family mothers must look after their children, men must protect and support their dependents. In turn, grown-up children are expected to help their parents when necessary, though in some extreme and exceptional cases, decrepit parents may be abandoned or killed. Helping more distant relatives is also a duty, the range depending on the extent of ties of kinship, which usually reach further in early society (*see* KINSHIP). Mutual aid is widespread and generosity everywhere praised. The higher religions extend the duty of helping the needy and distressed beyond clan or country to strangers. Making charity a duty shows moral advance, though the link between almsgiving and sacrifice sometimes distorts deeper moral insight. If almsgiving is the "medicine of sin," it may become detached from charity; we may thus lose sight of deeper needs and ways of permanently relieving them. Modern society is gradually achieving revolutionary changes here; it accepts a duty not just to deal with ills as they arise, but to forestall them. Man is beginning to develop an international conscience, responding not only to the occasional cry of distress, but bending his efforts toward abolishing disease and poverty, ignorance and neglect, while helping those who are building up new societies.

Telling the truth and keeping promises are widely recognized as duties, though they are sometimes withheld from strangers.

Pride, boastfulness, arrogance are widely condemned, while their opposites, modesty and humility, are praised; though differences of status also enter, and the line between self-respect and weakness is variously drawn. Indeed, religious influences sometimes make pride not just a vice but a sin, while stressing humility to the limit of self-effacement.

As to the "self-regarding" virtues, Charles Darwin (*Descent of Man*, ch. iv, 1871) thought that they were on the whole not much esteemed by primitive peoples. Joseph Butler remarked more generally that "Nature has not given us so sensible a disapprobation of imprudence and folly as of falsehood, injustice and cruelty" (Dissertation, "Of the Nature of Virtue," sec. 9, 1736). The attitude to temperance is variable and it is not always easy to say how far self-control and self-denial are valued for their own sake or as a means to broader ends.

The Variability of Moral Judgments.—Diversity in this field is proverbial, for several main reasons.

Variations due to difference in moral import of the "same" acts in different social situations and institutional frameworks. The exacting of payment for money on loan is "usury" at one level of economic development, but "legitimate interest" at another. Self-redress is condemned where a system of public justice exists, but otherwise allowed as normal. Conflicting duties are balanced in various ways, depending on circumstances. Killing the aged or newborn infants, which Locke used to illustrate the diversity of moral judgments, may be thus explained. Under the hardship of nomadic life affection and care for the helpless may have to yield to survival of the group.

Variations arising from changes of knowledge or belief regarding the nonmoral qualities of acts and their consequences. In our own time the moral attitude toward birth control changes as we learn more about infant and maternal health, human heredity and the economic consequences of population policy. The attitude toward abortion has depended in part on when the fetus was held to have become a separate being. As psychological knowledge grows, views on motives, intentions and responsibility change and may alter the attitude toward the ethical basis of the criminal law. Ignorance and superstition long affected the treatment of the insane, many of whom were burned as witches or heretics, or else treated as criminals. As the writings of E. A. Westermarck and W. E. H. Lecky show, countless variations in moral attitude are caused by differing religious and magical beliefs. Some of these beliefs condemn murder and suicide, while others advocate human sacrifice and certain forms of voluntary self-destruction; some inculcate kindliness and universal benevolence, others justify cruelty and persecution. They may enjoin marriage or celibacy, chastity or temple prostitution. By insisting on inwardness and sincerity they have helped the growth of morality, yet by multiplying regulations they have tended to replace moral discrimination by observance.

Variations in the importance attached to different virtues or in their balance. Sometimes the military virtues of courage, discipline and heroic self-sacrifice dominate; sometimes charity, gentleness, self-abnegation are more important; sometimes devotional habits, purity in thought and deed are given prominence. In periods of industrialization thrift, steadiness in work, punctuality, forethought are stressed. Where the scientific spirit is strong, value is placed upon the intellectual virtues, love of truth for its own sake, faithfulness to evidence and impartiality.

Great differences exist in the way in which religious duties are squared with those of daily life. But the difference between the active ethics of self-fulfillment and the passive ethics of the extinction of self-assertion, whether irreconcilable or not, does not divide west from east. In China, the Taoist outlook is balanced by the more practical Confucian doctrine of conduct. In Christianity, quietist and activist elements are variously balanced at different times. Even Hinduism has some activist elements.

Variations due to different ways of meeting primary needs. In modern western societies, great value is placed on linking sex with lasting companionship, tenderness and the enhancing of parenthood through a common interest in the offspring's welfare. It is thought that children can best acquire a sense of stability and security by living with both parents. There arises the notion of the individual as an end in himself and therefore of a union based on reciprocity free from possessiveness. In other societies many of these values may be recognized but in different clusters. Thus marriage may be based on economic and sexual relationships, but may not be expected to meet the need for companionship. The care of children may be entrusted not to the narrow unit of the western family, but to the extended family, and it may even be claimed that in this way children are better prepared to face the loss of parents or their separation. (*See* also FAMILY.)

Variations due to differences in the range and strength of sympathy, and in moral insight. In Westermarck's view, men extend the scope of moral principles and their impartial application when growth of the social unit, or greater contact with other societies, widens their outlook and range of sympathy. But sympathy is not a feeling only; it also requires imagination, enabling one to see himself in another's place and thus to recognize him as human. As we extend our moral rules to wider groups, so we refine our notion of the human person. Moral universalism goes beyond kindliness, for this always discriminates. We cannot love all and sundry equally but we may come to see that men deserve certain things as men, whether we love them or not. This moral intuition draws both on thought and feeling, and both are deepened by growing experience and insight.

The uniformity of basic moral principle and the variety of practical rules are well illustrated by the ethics of the family. To the philosopher, the function of sex morals is to ensure the kind of

family life best designed to safeguard the well-being and the future of mankind, while allowing full scope to the affections and emotions most vital to happiness and self-fulfillment. The rules actually found are gropings to attain these ends, shaped by a vast array of forces, psychological and social, interacting with one another in ways which differ greatly in different societies.

Family ethics deals with highly complex psychological forces. There are, first, impulses and feelings surrounding sex relations; secondly, the conjugal and parental impulses and affections attendant on relations of dependence and protection; and, thirdly, the possessiveness and self-assertion of the male, which leads him to hold on to what he has appropriated and to fear the contempt and ridicule that he might arouse if he failed to maintain his claims. Accompanying all these are the emotions of rivalry, jealousy and envy, which infect all relations tinged with possessiveness, and the hatred and antagonism that they engender. Finally there is the feeling of mystery about procreation and birth, which partly accounts for the medley of feeling concerning sexual purity and perhaps also for the recurrent horror of sex in the history of morals. All these impulses and emotions interact with social forces embodied in institutions—the forms of property, the regulations governing inheritance and the legitimization of children, the economic organization and its division of labour, the religious and other beliefs acknowledged in the society. The ethics of marriage and the family seek to deal with these complex situations.

Most nearly universal are the regulations governing sex relations or marriage between near kin. The degree of consanguinity which rules out such relations varies; but the closer the relationships (*e.g.*, mother and child, brother and sister), the greater the uniformity. The rules have been variously explained. Some hold that they prevent sexual jealousies and rivalries within the family circle by enjoining males to seek their sexual mates elsewhere (Jeremy Bentham, *The Theory of Legislation*, Eng. trans., 1864). Others appeal to a deeply rooted feeling that sexual intimacy is incompatible with family affection and respect (Montesquieu, L. T. Hobhouse, B. Malinowski and others). However this may be, the sociological outcome is to widen the social unit by binding distinct groups through intermarriage. Psychoanalysis holds that these rules help the individual to break away from his own family, which he can do only if marriage takes him into another.

The rules enjoining conjugal fidelity vary considerably. Among many preliterate peoples it is a stringent duty for the wife, but not generally for the husband, though there are some exceptions. In the ancient civilizations the law generally required faithfulness from the wife only. The Roman jurists condemned adultery in the wife because of the danger of introducing strange children to the husband. Jewish law requires chastity from both parties. Christianity makes no distinction between husband and wife. Popular morals and the laws relating to divorce and separation show, however, that standards are not the same for the two.

In attitudes toward sexual relations before marriage peoples differ greatly. Among preliterate peoples, as many condone as condemn them; and, if the latter, the seducer is censured as well. More advanced peoples insist on female chastity; in men it may be considered as an ideal, but its violation often meets with little censure. Christian morals regards all extramarital intercourse as mortal sin, though this theoretical standard is not firmly upheld in practice. In many parts of Europe pregnancy is a usual preliminary to and occasion for marriage.

It has been suggested that checks on early sex relations unconsciously aim at the ideal of prolonging the period of growth, in harmony with the general rule that the longer the youth of a species the higher its organization (*cf.* A. E. Crawley, "Chastity," in J. Hastings *et al.*, *Encyclopaedia of Religion and Ethics*). But there is little reliable evidence here. Sigmund Freud argued that without repression of sexual relations in childhood it would be impossible to control sex in adults. It is often said that early sexual relations, especially if promiscuous, would tend to make sex experience trivial and so hinder the growth of the deeper feelings which, particularly in western societies, are greatly valued. The working codes in the regulation of sex morals have been influenced by a confused mass of feelings and ideas, by magical and religious beliefs, by traditions which may have outlived their usefulness and by various social forces often working in opposite directions. Nevertheless certain underlying principles slowly emerge. There are fundamental rules that tend to maintain the stability of the family; *i.e.*, the union of the sexes sufficiently prolonged to ensure the proper rearing and training of children and to define their place in the group into which they are born. Extramarital intercourse is condemned, for it deprives men of adequate motives for shouldering the burdens of marriage. Premarital relations are forbidden, in order to prepare the young eventually to follow the same path. At a later stage there emerges the ideal of love as satisfying not only sexual needs but also the deeply rooted need to seek and give response; an ideal that grows with the respect for the individual as an end in himself. Along with this go norms of fidelity, trust and reciprocity, affecting the ethics of the monogamous family. The evidence, though incomplete, shows that this provides the best conditions for the free personal development of women and the fair apportioning of rights and duties between men and women. As to the bearing of monogamy on nonmarital sexual relations, both law and the working code reflect great differences in views.

Trends in the Development of Morals.—As social units grow larger and wider contacts are established, the range of persons within which rights and obligations apply is steadily extended. However, group morality persists: the limitation of universalism survives in race discrimination, in war and in the precarious rules of conduct supposed to govern it.

As for moral teaching, great changes divide the attitude of primitive peoples from that found in all the higher religions. For example, in primitive societies strangers are usually without rights; in Judaism, by contrast, "The stranger who sojourns with you shall be to you as the native among you, and you shall love him as yourself; for you were strangers in the land of Egypt" (Lev. xix, 34). The command against hatred (Lev. xix, 17) is interpreted in the Talmud as having universal significance: "Whoever hateth any man hateth him who spoke and the world came into existence." The doctrine of universal love is common to Christianity, Buddhism and Confucianism. The Stoics base their cosmopolitan ideals on the identity of reason in all men. "If our reason is common, there is a common law," says Marcus Aurelius; according to Seneca "it is a crime to injure any man, for he is your fellow citizen in a larger state." It is difficult to assess how far these religious and philosophical ideas influence actual behaviour. Still, universalism has grown; the scope of common principles and the impartiality shown in their application have steadily expanded.

Another trend in the history of morality shows conscience growing more internal and individual. Increasing stress is laid on inward sanctions. In the customary morality of the simpler peoples the sanctions are mainly external and prudential. The transition stands out clearly in the history of Hebrew morals. In Deuteronomy the emphasis is on rewards and punishments. In the Priestly Code the motives appealed to are rather the honour and glory of God. In the rabbinical teaching the thought emerges that the essence of observance lies in itself and not in external reward. In Christianity this trend continues. The desire for reward and the fear of punishment are not repudiated as motives, but the basis of morals in love is strongly insisted upon. Chinese ethics, as taught by Confucius and his commentators, rests morals on the intrinsic desirability of the ideal. Buddhism, while not quite abolishing reward and punishment, emphasizes the inherent character and inevitable consequences of acts.

At the same time there arises the notion of individual responsibility in the criminal law. In primitive societies responsibility is often collective and vicarious. The family, or even a whole tribe, may be held responsible for a murder committed by one member. In Anglo-Saxon law a family could be enslaved for a theft committed by the father (F. Pollock and F. W. Maitland, *History of English Law*, vol. i, p. 56, 1895). In Judaism the transition emerges: "In those days they shall no longer say: 'The fathers have eaten sour grapes, and the children's teeth are set on edge.' But every one shall die for his own sin" (Jer. xxxi, 29–30). Purity

of motive gradually is seen as distinct from outward performance. Already the Babylonian gods exacted from their followers "perfection of the heart" and, as to truthfulness, distinguished between saying "yes" with one's mouth and "no" with one's heart. In the Psalms the demand is made for "clean hands and a pure heart" (Ps. xxiv, 4). In rabbinical teaching a distinction is drawn between legal and moral obligation. The former is justified by the latter and both are based on the principle of righteousness, that human perfection can be attained only by following "the ways of God" and imitating his attributes (*cf.* M. Gaster, "Conscience [Jewish]" in *Encyclopaedia of Religion and Ethics*). It is often claimed that inwardness or rightness of heart and spirit is the peculiar mark of Christian ethics. Yet similar views exist not only in later Judaism but in the doctrines of all the great moralists. Confucian ethics insists that in all moral relations sincerity is essential. Rightness of purpose, preference of virtue for its own sake, suppression of vicious desires are considered essential alike by the Aristotelians, who were not inclined to belittle outward circumstances, and by the Stoics to whom external things were indifferent (*cf.* H. Sidgwick, *Outlines of the History of Ethics,* ch. iii, sec. 3, 1886). Pagan and Christian ethics differ not in the value attached to "inwardness," but in their views on its conditions: for the Greek thinkers what mattered was knowledge or wisdom; for the Christian teachers faith or love. Christianity, in contrast with Judaism, relies on a saviour, seeing man as radically imperfect and unable to attain goodness unless given supernatural aid or grace. The central idea in Christian ethics is love of God, and of all men as objects of his love and as sharers in the humanity ennobled by the Incarnation—or love of God and love of neighbour, as it is put in the well-known "summary of the law" (Matt. xxii, 35–40). The word "love" cannot mean the same in the two cases, however: in the sense of benevolence toward men it can hardly be a mere derivative from the love of God. "If it were, there would not be much of it even in the Christian world. As Dean Inge says: 'love of God' is talked about more often than felt" (E. A. Westermarck, *Christianity and Morals,* p. 79, Routledge & Kegan Paul, London, 1939).

A third trend in the history of morals is the growing rationalization of the moral judgment. Magical elements gradually disappear, a distinction comes to be drawn between ritual or ceremonial rules and moral rules proper while customary morality and social institutions are subjected to persistent criticism in order to reconstruct them on a rationally defensible basis.

In early morality tabu plays a large part. Tabus on intercourse universally influence the rules of marriage. In many savage societies tabu is used to protect property. Paternal authority is often upheld by the power of the father's curse or blessing. The tabu is contagious; whoever breaks it may suffer exclusion from the community. Tabus being unconditional, they resemble Kant's categorical imperative in that the forbidden acts are deemed bad as such. Nevertheless, the operation of the tabu is mechanical—it takes no account of character or intention. It is paramoral: it works alongside of moral principles, but not always on the same lines (L. T. Hobhouse, *Morals in Evolution,* 7th ed., p. 421, 1951). Morality progressed when tabus were challenged: when rules came to be obeyed for reasons other than fear in breaking them or the opprobrium thus provoked.

What further helped morals to grow more rational was the distinction that came to be drawn between ritual rules and moral rules: the former prescribe acts that a deity expects as due to himself, the latter such as are thought necessary to human well-being. The history of the ensuing struggle, familiar from Judaism and Christianity, is paralleled in other religions. In India, religious ceremonials have for long loomed large in the daily life of the classes subject to metempsychosis, especially of the Brahmans. But many ancient rites and restrictions are now falling into disuse. Islam lays great stress on ceremonial observance, though in the Koran the Prophet does not call merely ritual performances good: "Piety is not that ye shall turn your faces toward the east or the west, but the pious man is he who believes in God . . . and gives his wealth for his love of kinsmen, orphans, the poor and the wayfarer." The formalism of Islamic orthodoxy has at times provoked protests; and on the view of modern exponents, Mohammed's chief mission was that of a moral reformer (*cf.* E. A. Westermarck, *Origin and Development of the Moral Ideas,* vol. ii, p. 727, 1906).

Above all, morals become more rational through the recurrent protests of prophets and other innovators, the movements which they inspire or from which they emanate, the efforts of jurists to systematize the law and, in modern times, to elicit its ethical principles. The direct influence of philosophical ethics, more difficult to estimate, was important in some cases: in the impact of Stoic teaching on Roman law and of utilitarianism on the reform of the common law of England. The rational approach in ethics is to elicit the underlying principles of the working code of morals, to bring out any discrepancies and to seek to resolve them in the light of wider principles. Philosophers contribute to this task. But the work of rationalization reaches much further and in more or less articulate form inspires many popular movements. An instance of this is the socialist criticism of the rights of property. These rights were shown to be based on the principles that a person is entitled to the produce of his labour, to what results from the products of his labour and to what he acquires by inheritance. All these rights are alleged to be deeply rooted in human nature and in the conditions of social life. Societies also recognize, though insufficiently and in varying forms, the right of all members of the community to subsistence, that is, to the commodities and services needed to support life. In simple societies these rights and principles can be made to work in harmony. There are no large accumulations of "unearned" incomes, and the system of mutual aid deals effectively with those who are in want. In large-scale, complex societies, on the other hand, the principles conflict. For the accumulation of property made possible by the right to the returns on capital and the right of inheritance has been accompanied by modes of distribution that prevent many workers from keeping enough of what they produce to meet their needs. The socialist reconstruction tries to reconcile the conflict of rights and principles thus engendered by a fairer sharing-out of the produce of labour and the means to well-being. Whatever the force of the account, this nicely illustrates rational procedure in the analysis of the ethical basis of social institutions. Efforts made in this century to elicit the principles of punishment and to amend the criminal law accordingly provide other examples.

The three trends of moral development here briefly sketched are not independent. The last is connected with both of the former. These developments are neither regular nor irreversible, but they have occurred in many cultures and with some persistence.

In all known societies there is some sense of mutual obligation upheld by custom and some demarcation of spheres of conduct which come under social control. In simpler societies where retaliation or compensation prevail, the notion of an impartial rule impartially applied has not yet emerged. The notions of responsibility and purposiveness essential to the fully formed ethical judgment are not yet fully grasped. The sanctions lying behind custom at this stage are essentially nonmoral, namely tabus or other magical terrors, and the fear of vindictive or resentful spirits.

The more developed societies have systems of public justice and a recognized authority which administers the law more or less impartially and enforces definite sets of rights and duties while repressing retaliation. The sanctions, still egoistic and prudential, are based mainly on the fear of punishment either in this or in another world. Yet they are ethical insofar as they are general and impartial. Moreover, at this stage religious ideas tend to become more ethical, and there arises the conception of a just deity, who in this or in a future life deals with acts according to their deserts. As the groups grow in size, distinctions within them arise. Class morality is added to group morality and loyalties based on hierarchical subordination replace primitive equality. In this respect economic and political advance has been accompanied by moral retrogression.

In principle, group morality is overcome in the teaching of the spiritual religions. All of them, though in different forms, preach good will to all mankind. Hobhouse rightly says that they engendered a new species of ethical truth, the conception of an order

based on self-sacrifice rising above an order based on self-assertion. The contrast, however, was overdrawn, as by Christianity, Hinduism and Buddhism in their tendency to exaggerate the ascetic elements in morality, to concentrate on personal salvation rather than on social organization and to seek goodness in forms of life remote from the ordinary interests of men. Greek ethics, on the contrary, sought to bridge the gap between the individual and the social by the doctrine that personal well-being lay in contributing to the civic life: active fulfillment of human striving, rather than its suppression, was here the end of conduct. Stoicism provided a new basis for universalism in the notion of a law of nature applicable to man as man everywhere.

In modern times it has become far more difficult to reconcile self-fulfillment with self-surrender, to harmonize individual with social welfare. As religious struggles have deepened the conception of the human person, the difficulties of freedom have been intensified. With many more ties, modern man must face the problem of conflicting loyalties; the vast growth of social organization has brought great inequalities of political and economic power and made it harder to reconcile efficiency with freedom. With these problems neither the primitive codes of reciprocity, nor medieval status ethics, nor yet the universal benevolence taught by the spiritual religions can cope. Personal ethics has to issue in social ethics; there has to be a turn from commutative justice, or equivalence of exchange between individuals, to distributive justice, which provides the social conditions of well-being in terms of a communal standard sustained by a spirit of social responsibility. Hence the efforts to extend the notions of human rights from the political to the social and economic spheres and to universalize them. Both liberal-democratic and communist societies recognize these ends, but interpret and pursue them differently.

Moral and Religious Development.—In their developed forms religion and morality are closely interwoven. It has indeed been maintained that the truly religious life and the truly moral life are one and the same. Yet historically the two orders have not always been in sympathy, and they have frequently been in conflict. Religious conservatism has often hampered the moral innovator, while religious teachers have at times pointed to a way of life far in advance of that sanctioned by the morality of the day.

There is first the question how religion and morals have interacted in history; and secondly, the logical question, how far ethics needs religion to justify its claims, and whether moral beliefs warrant the belief in God. Comparative ethics is mainly concerned with the historical and sociological question. In the animistic phase the links between morality and religion are slender, "ethical laws stand on their own ground of tradition and public opinion, comparatively independent of the animistic beliefs and rites which exist beside them. The lower animism is not immoral, it is unmoral" (E. B. Tylor, *Primitive Culture*, vol. ii, p. 360, John Murray [Publishers] Ltd., London, new issue 1929). The gods of polytheism are not necessarily moral. They are rather allies to be invoked or enemies to be dreaded and they share in the resentments and approvals common to their believers. For many preliterate societies there is evidence of belief in a kindly god, who, however, is often too remote for worship. There are many rites to propitiate or repel evil spirits, who sometimes may have risen to the status of a deity without discarding their evil nature. The belief in vindictive, jealous and irritable spirits may tarnish the attributes of the higher divinities. The high gods may be credited with mercy and benevolence, but a cruel ritual may persist. On the other hand some evidence has been adduced to revive a theory of primitive ethical monotheism: among certain of the peoples who are still in the food-gathering stage (*e.g.*, the Semang, the Fuegians and some African pygmies) there is a belief in a power who takes a certain interest in a group of offenses. These are mostly morally indifferent, but they include adultery, practices which might give rise to incest and, in one case, murder. How far the belief is original or borrowed is in dispute. In general the high gods or supreme beings of preliterate peoples do not prescribe or enforce the tribal laws. When they are considered as creators of the world they are often too remote to be concerned with the affairs of men and are rarely the objects of a cult.

The later history of religion shows that the gods are slowly moralized. Values, it would seem, must be found in experience before they can be attributed to another world. According to this view, it is because men admire justice that they regard God as just, because they value love that they consider God as loving. Consider the moral growth in the notion of the sacred or holy. This is first conceived magically, as a superhuman quality, inspiring dread and to be met with care. As dread turns into solemn awe, the qualities which cause it change, though even in the higher religions the holy retains magic elements. The notion of sin is of special interest. Starting as a sort of physical pollution whose consequences may be magically averted as, for example, by burning it away, it came to be considered as an offense against God. The link between religion and morality is closest in the higher world religions. In Buddhism the two almost merge; in its purest form it advocates a way of life and salvation. Confucianism is primarily an ethic, at first quite free from theology. In both cases the later phases contain theistic elements, but the ethical teaching retains a measure of independence (and in practice codes of behaviour are not tied to beliefs in metaphysical doctrines or religious dogmas).

In the case of Christianity, Hobhouse has pointed out a reversal in our view of how morals and religion are linked: religion, no longer the basis of ethics, now must pass the ethical test. Westermarck holds that "the moral ideas of uncivilized men are more affected by magic than by religion and that the religious influence reached greatest extension at certain stages of culture, which, though comparatively advanced, do not include the highest stage" (*Moral Ideas*, vol. ii, p. 746, Macmillan & Co., Ltd., London, 1906). It might be, however, that in the higher stages religious influence has changed rather than slackened. In modern times both religion and morals appeal to direct and widespread experiences; historical forms and creeds become secondary and authority ceases to be the final arbiter. In this way religions may well provide a body of moral convictions at the very time when their dogmas lose in importance.

Still, many may form moral judgments outside any religious faith and let these guide their actions as firmly as might believers whose morality is rooted in religion.

The varying relations of religion and morals also emerge in the belief in immortality and in reward and punishment in an afterlife. Tylor holds that the two need not be connected and that the belief in retribution is, on the whole, later than that in survival. Hobhouse (*Morals in Evolution*, pp. 432 ff.) makes the following points in regard to the matter: often there is belief in continuance without retribution being mentioned; the conception of a future life is very vague as soon as it abandons the notion that the dead continue the present mode of life; the well-being of the dead is thought to depend mainly not on their conduct in this world but on the manner of death and on care by relatives after death; belief in punishment, where it occurs, is partial and unsystematic, certain offenses being chosen and not others; and, finally, belief in retribution is not necessarily linked with the notion of a god who judges.

However, in some cases there is judgment designed to discover whether the person is fit to enter paradise; this may also be tested by ordeal. In the higher polytheisms the notion of retribution is found more often. Egypt shows a gradual shift from the view that funerary rites and knowledge of potent formulas are necessary to future bliss to the weighing of the heart before Osiris.

Ideas of personal immortality and reward and punishment in an afterlife have at times been very crude. Lecky relates the views of many writers from the 7th to the 12th century on the condition of the damned. "It is impossible," he concludes, "to conceive more ghastly, grotesque and material conceptions of the future world than they evince, or more hideous calumnies against that Being who was supposed to inflict upon His creatures such unspeakable misery" (*History of European Morals*, vol. ii, p. 235, 1869). The idea that without belief in a future life there can be no morality has been and is widely held. Martin Luther is quoted as saying: "If you believe in no future life I would not give a mushroom for your God. . . . Do then as you like. For, if no God, then no Devil, no hell. . . . Then plunge into lechery, ras-

cality, robbery and murder." Such an attitude would now be thought not only immoral, but irreligious. Theological defenders of the belief in personal immortality hesitate to connect it with the idea of retribution or compensation. They tend to accept the view that the notion of a reward for virtuous conduct is incompatible with the intrinsic value of virtue. Dean Inge quotes Plotinus: "If any man seeks in the good life anything beyond itself, it is not the good life that he is seeking." He adds: "The substitution of a gross pleasure and pain compensation doctrine for this noble faith has been punished, in Christian times, by a decay of the belief in human immortality. It has withered because its roots have been cut ("Philosophy and Religion," in *Contemporary British Philosophy*, ed. by J. H. Muirhead, 1st series, p. 205, George Allen & Unwin, Ltd., London, 1924). Jewish and Christian thinkers have maintained that the belief in a future life has lessened the fear of death and made it easier to endure the calamities and sufferings of life on earth. On the moral side, it has been further claimed that without the belief in reward and punishment "whole nations could not rise to an austere morality or even maintain a decent standard of conduct" (*cf.* W. McDougall, *Body and Mind*, preface, Methuen & Co., Ltd., London, 1911). However, the belief in divine justice could and did arise without reference to any life beyond the grave to explain or justify its workings. Non-believers in personal immortality can be as tenacious of principle and as fearless of death as anyone.

The following general points emerge: First, all world religions lay great stress on the importance for morality of inward sanctions and sincerity of motives, but while in the monotheistic religions this emphasis is linked with the belief that God sees into the hearts of men, in Buddhism and Confucianism it is based rather more clearly on direct moral experience and the intrinsic desirability of the ideal. Secondly, the world religions all teach universal benevolence and all stress the unity and interdependence of the whole world, but some as an emanation or the creation of one God and others as the expression of one law governing the universe. With this root the idea of kindliness to all creatures is closely connected.

As an operative belief universalism is variously limited: in Christianity and Islam the unity of mankind has long meant the unity of the faithful; Confucian writers speak of mankind in general, but in view of China's history it is doubtful whether they had the conception of humanity reached in western thought. Thirdly, among the important contributions of the religions to human development are the numerous heresies to which they gave rise, and the struggles between the spiritual and political powers. In the long run, though at a terrible cost, they have enriched the human mind and raised the level of morality over a large part of the world. How much this process owes to changes in religion and how much to rational reflection at variance with religion is hard to say.

On the negative side there is, first, the tendency to exaggerate the ascetic elements in morality. In this respect the world religions differ greatly. As compared with Hinduism, Buddhism or Christianity, Judaism and Confucianism have little of asceticism in them. There is great need for a study of what makes for asceticism, such as conditions that encourage world-weariness, the heightening of a sense of sin and consequent desires to escape from sin and to inflict self-punishment. Of special interest is the problem of the origin of the belief that the body is the source of all evil and that sexual relations are "impure," incompatible with a holy life or only admissible under religious rites.

Closely connected with the emphasis on self-surrender or self-negation is the tendency of the world religions to concentrate on personal salvation rather than on the shaping of social institutions. Here too they differ greatly from one another. In Judaism the individual good and the social good were not opposed, but regarded as bound up with one another. So in Confucianism the ethical code is on the whole directed to the needs of social life, peaceful industry, a happy family life and the checking of official corruption and military ambition. Even within any one religion the emphasis on personal salvation as compared with social ethics changes greatly in the course of time. Thus in Christianity no single formula will describe adequately the medieval notion of "my station and its duties," the Puritan desire to control every detail of social life and the efforts of the Methodists and Evangelicals to make simple Christianity the touchstone of political and social issues. In modern times all religions have been transformed by the growth of science and the infusion of the humanist spirit. Under their influence the old antitheses of the flesh and the spirit, the claims of the inner life and of life in society, of self-negation and self-fulfillment have lost much force or greatly changed in character.

Morality and Law.—The relations between law and morals have varied greatly. Men make and remake their laws as their notions of justice change; but laws in turn may have a great influence in shaping, developing or perverting men's sense of justice. No systematic study exists of the complicated interactions thus arising. Systems of law which have grown up slowly in response to social needs embody relations between law and morals different from those in systems imposed by a conqueror. Systems in which ethics, religion and law are interwoven, as in the law of Islam, differ from secularized systems; in these, again, the impact of morals on law varies with the type of political structure and the power of public opinion. We must distinguish between what the law is, and what philosophers think it ought to be. Law as it stands, though in the long run affected by changes in moral outlook, is never a simple record of clearly thought-out ethical principles. A rational ethic might state that the law should aim to curb arbitrary power and assure a just distribution of the community's resources to further general well-being. In practice, conceptions of what constitutes arbitrary power and general well-being are shaped by existing inequalities of power, political or economic.

Still, the law gradually grows more just, while legislation comes more and more to rely on fact-finding inquiries and on notions of public welfare.

In democratic societies, as the law grows more moral, so men have tried to define its limits, to mark off spheres which are best left to individual choice and initiative. Certain values or goods can only flourish in freedom: such are religion, learning, research, science, philosophy and art. The state can use the machinery of the law to provide some of the external conditions which these ends require, but it seeks to avoid direct control. In authoritarian societies, by contrast, the distinction between law and morals tends to be blurred. These societies have reverted to the fusion of spiritual and temporal power by subjecting art and science to political control. In seeking to regulate all areas of life they tend to legalize morality and thus to give it the flavour of the jural morality of medieval casuistry, or the all-pervading discipline of the age of Puritanism.

Moral Progress.—Despite widely felt disillusion the belief in moral progress remains strong. It raises two questions. First, are there any principles of appraisal for evaluating moral development? Secondly, if, as might be conceded, there has been improvement in ethical conceptions, has there been improvement in social institutions, as judged by their ability to provide the conditions most nearly in accord with ethical ideals?

As to the first question this article assumes that there are such principles. They may be stated in terms of the following criteria: (1) *differentiation*, whereby morals become distinct from custom, law and religion, and a moral attitude arises, recognizing values and obligations as self-sustained and independent of external sanctions; (2) *universalization*, whereby common moral rules become applied more widely and more impartially, with a better grasp of principles and a greater range of sympathy and altruism; (3) *comprehensiveness*, whereby morality comes to cope with a wider range of needs and values, dealing with them more sensitively, while growing more alive to new values and more tolerant of divergent ones; (4) *coherence*, whereby moral judgments become more systematic, bringing out discrepancies and disentangling underlying assumptions; and (5) *self-criticism and self-direction*, whereby moral codes come to be examined more critically and public policy is increasingly shaped by fact-finding and evaluation of actual or possible ends.

Moral development does not proceed in a straight line, but suffers setbacks and retreats. Moreover, it goes on in many distinct

centres each with its own tradition. Yet, in later phases, they often converge and interact. Peoples learn increasingly from each other and see that in essentials they are engaged in a common task —the unfolding and fulfillment of human faculty. Different systems satisfy the five criteria in varying degrees; over long stretches of time, it may be difficult to balance gains and losses. Still, certain lines of advance are evident. The later phases of all systems show more or less clearly a moral attitude insisting on inward sanctions and sincerity of motives. As to universalism, limitations not only persist but are even approved by the operative code. Still, many societies have greatly advanced, group morality has been variously transcended, and arbitrary exclusion has come increasingly under attack. In systematizing codes the western peoples have made great advances, aided by new psychological insights and philosophical analysis. These methods are steadily bringing to light the discrepancies and contradictions in western morals. They will no doubt do likewise when applied more fully to other systems.

The notion that men can and ought to do something to direct the course of future development has spread the world over. The most impressive fact in the history of social development is the growth of legality. Even among the most advanced nations of the modern world justice was, a mere thousand years ago, in the hands of irresponsible, warring feudatories. Today more than half of the world's population lives under systems derived from Roman law or the common law and there are strong tendencies toward further unification.

Legal systems themselves have become more systematic, self-critical and more open to changes required by new concepts of general well-being; they have become more rational and have in turn helped the general trend toward rational morality. International law, too, has grown; for the most part states honour their obligations, and in cases of dispute they accept the decisions of courts even when these go against them. We are apt to think that international law is more frequently defied than the internal law of states, but we cannot well compare states with individuals, since states are few. The law under feudalism provides a fairer term of comparison: the obedience given by states to international law is no less regular or more reluctant than that given by powerful feudatories to the common law.

There have also been failures. Thus religious development has often hindered the growth of knowledge; the power conferred by knowledge has led to situations that moral wisdom, such as it is, cannot encompass; political and economic changes have often undermined or compromised the code of moral obligations. In any case morality is only one and by no means the most powerful of the forces shaping the life of societies. We cannot assume that the ethical elements will always triumph; but we may reasonably hope that, as moral ideas gain in rational coherence, through greater insight into human nature and its scope, they will also gain in their power to inspire and sustain social action.

Because morality has grown unevenly in different places and traditions, many conflicts and suspicions remain. To overcome this unbalance is the great problem of the 20th century. The spiritual religions and rational humanitarianism have all failed so far to make their doctrine of universalism effective. Yet it is a sign of ethical advance that we can now think of the relations between peoples at different levels of development, not in terms of subordination, or even of enlightened administration from above, but of equality in freedom. It has been suggested that, though a change in institutions may alter men's behaviour, it may leave them essentially the same. For instance, the rules of the road may leave the fundamental stock of human selfishness unaltered, while reducing the occasions that bring it out. On this view, improvements in institutions may well have the effect of making men appear better than they are. Likewise, bad institutions tend to make men appear worse than they are, though this conclusion is not usually drawn.

Underlying all this is a belief in "original" sin, which sees in human nature evil impulses which institutions may control, but not eradicate. By contrast, those who think that human nature is somehow better than its performance believe in the fundamental goodness of man. Both beliefs rest on too abstract a view of human nature, separating it too sharply from its manifestations in behaviour. The concept of an "original" human nature has little meaning.

With collective action it is difficult to allocate responsibility. Much here is unintended and beyond the control of individuals. This has led to the view that social forces are "impersonal" and moral factors of little account. On the theory followed in this article this view is wrong. In advanced societies it is recognized that responsibility for public policy is unequally divided, but that in some measure it is shared by all. Men acting in concert cannot on that account disclaim responsibility. Its precise incidence depends of course on the structure of the group, on the extent to which members can exercise control, on the opportunities for active consultation, effective protest or repudiation and on what information men have to assess the outcome of their actions. In democratic societies different degrees of responsibility attach to direct policy-makers (*e.g.*, legislators and politicians), to those who mold public opinion (*e.g.*, publicists and journalists) and to the ordinary citizens, according to how far each can take part in shaping policy. Social progress then partly consists in the growth of common purposes and common responsibility. As societies develop, more men become willing and able actively to influence policy, while growing impatient of being told what they want or ought to want. Morality is still far from dominant in human affairs, but on the evidence it is gaining in strength.

See ETHICS, HISTORY OF for ethical theory in western philosophy; for the various eastern philosophies *see* BUDDHISM; CHINESE PHILOSOPHY; CONFUCIANISM; INDIAN PHILOSOPHY; TAOISM.

BIBLIOGRAPHY.—L. T. Hobhouse, *Morals in Evolution*, 7th ed. (1951); E. A. Westermarck, *Origin and Development of the Moral Ideas* (1906); C. Read, *Natural and Social Morals* (1909); A. Sutherland, *The Origin and Growth of the Moral Instinct* (1898); W. E. H. Lecky, *History of European Morals from Augustus to Charlemagne*, 2 vol. (1869); C. Brinton, *A History of Western Morals* (1959); A. MacBeath, *Experiments in Living* (1952); M. Ginsberg, *Essays in Sociology and Social Philosophy*, vol. i, *"On the Diversity of Morals"* (1957). On the ethics of the simpler peoples *see* R. Redfield, *The Primitive World and Its Transformations* (1953). For ethical systems of higher civilizations *see* M. Lazarus, *The Ethics of Judaism*, Eng. trans. by Henrietta Szold (1901); I. Mattuck, *Jewish Ethics* (1953); E. Westermarck, *Christianity and Morals* (1939); C. C. J. Webb, *The Contribution of Christianity to Ethics* (1932); W. R. Inge, *Christian Ethics and Modern Problems* (1931); R. Levy, *An Introduction to the Sociology of Islam* (1933); D. W. Donaldson, *Studies in Muslim Ethics* (1953); E. W. Hopkins, *Ethics of India* (1924); J. K. Fairbank (ed.), *Chinese Thought and Institutions* (1957). (M. GI.)

ETHICS, HISTORY OF. The following historical sketch traces the main lines of development of ethical theory in western philosophy. Fuller details will be found in separate articles on special subjects. There are also separate articles on the most important thinkers whose names occur in this survey.

The present article is outlined as follows:

I. Greek and Greco-Roman Ethics
  1. Socrates
  2. Plato
  3. Aristotle
  4. Cynics and Cyrenaics
  5. Stoicism
  6. Epicureanism
  7. Roman Stoicism and the Academy
  8. Neoplatonism

II. Medieval Ethics
  1. Augustine
  2. Abelard
  3. Bonaventura
  4. Anselm
  5. Thomas Aquinas
  6. Roger Bacon
  7. Duns Scotus
  8. Ockham and the Averroists
  9. Maimonides
  10. Humanism and the 16th-Century Scholastics

III. Modern Ethics
  A. 17th and 18th Centuries
    1. Grotius

2. British Ethics
3. Utilitarianism
4. Continental European Ethics
5. The British Idealists
B. Later Theories
1. Nietzsche
2. The Developments of Utilitarianism and Its Critics
3. Ethical Relativity
4. Pragmatism
5. Existentialism
6. Logical Positivism and Analysis

I. GREEK AND GRECO-ROMAN ETHICS

The ethical speculation of Greece and therefore of Europe had no abrupt and absolute beginning. In Greek philosophy before Socrates there are ideas which go beyond the proverbial gnomic wisdom of the poets or the "seven sages," but they are only the rudimentary and crude beginnings of a reflection on moral life. The Pythagoreans applied their general conception of the opposition between limit or number and the unlimited or indefinite to human conduct, in regarding justice or right action as the hitting of a definite mark in an indefinite range of possibilities. Others, including Heraclitus, saw particular objects as mixtures of opposing elements and this idea too was applied to human conduct. The sophists undertook to educate men to be good citizens. But they regarded being a good citizen as a technical achievement like being a good flute player; success in life, as in flute playing, required a technical education. The sophists therefore claimed to equip their pupils to get on in the world. So far as they considered morality as such at all, they identified it with the rules or customs prevailing in each city, and this gave rise to a primitive form of social relativism in moral values. Some (for whom Thrasymachus in Plato's *Republic* stands as an example) looked farther and asked why these rules were laid down; and answered that they were the rules which brought benefit to the governing group or individual in each state, thus anticipating the Marxist analysis of political power and "bourgeois morality."

1. Socrates.—No system of ethics could be constructed until full attention had been directed to the special features of moral experience and to the vagueness and contradictions of ordinary moral opinions. This required a philosophic intellect of the first rank concentrated on the problems of conduct. In Socrates these requirements were met.

It is difficult to set exact bounds between the achievements of Socrates and Plato because Socrates is the principal speaker in almost all the dialogues of Plato, from which we have to discover Plato's views also. Apart from these dialogues the only evidence we have about Socrates are the rather simple-minded *Memorabilia* of Xenophon and the caricature in the *Clouds* of Aristophanes. It is assumed in this article that Socrates is historically portrayed in Plato's early dialogues, but that in the middle and later dialogues "Socrates" gives Plato's own views. This division is, however, relatively unimportant, because what matters here is the development of the ethical views, not the name labels attached to them.

Socrates went about showing that all those who claimed to teach men how to live, poets, orators and sophists, were unable to defend their views by argument or even to give adequate definitions of what they were claiming to explain. They were both inconsistent and ignorant. Socrates looked for release from this ignorance first to the clear understanding of the words that we use, that is, to definitions. But he refused to put forward such definitions himself. He alone knew better than the rest, he said, because he knew his own ignorance. His function was that of the electric eel, to galvanize his hearers into intellectual activity, or that of the midwife, to try to bring the thought-children of others to the birth. While therefore the specific results of his discussions with his victims were negative, they implied and indeed gave birth to the famous Socratic maxim that "virtue is knowledge." No man, said Socrates, could base a satisfactory life on the shifting, vague and contradictory opinions which were all that ordinary men had and (as he showed) all that their most reputed teachers could offer them. We must know what courage and justice and law and government are before we can be effectively just or brave or good citizens or good rulers. More particularly, everyone wishes for his own good and would get it if he could; and no one would deny that justice and virtue are of all goods the best. The problem is thus to find what things are good and how they are related to each other. If we knew that, we should inevitably lead good lives. Men err by pursuing things which they think good but which are really inferior to other greater goods. This involves the paradoxes that no man errs willingly, that all sin is ignorance and that no man knowing good would do evil; and these paradoxes Socrates seems willingly to have accepted.

2. Plato.—The ethics of Plato cannot be treated as a finished product but rather as a continuous movement from the position of Socrates toward the more fully articulated system of Aristotle. The first development from the Socratic position carried the problems from morals into metaphysics. Socrates had initiated the search for definitions. Plato held that if this search succeeded we should know what justice, courage, etc., really were. Since the word "know" implies that what is known is independent of the individual mind of the knower, he held that there must be such independent realities as justice, courage, etc. At the same time he had learned from the Heracliteans that particular objects in the world around us are always in a state of flux and change and that there could be no knowledge of them; and from the Eleatics that the apprehension of particular objects by the senses is subjective and relative. He concluded that such realities as justice, courage, etc. (which he called "forms"), were permanent and unchanging entities existing independently of particular things and actions. He held that these realities could not be apprehended by perception or empirical observation but must be pursued by pure reason, by the special activity that he called "philosophy," whose goal was certain, final and infallible knowledge. Since the "forms" alone are real, he held man's reason, which leads him to knowledge of them, is the highest side of his nature and that most akin to the eternal and the divine. Since such knowledge is the only firm guide to action, reason must be trained and directed to achieve it; and, in the state, philosophers must be kings.

These conclusions had paradoxical results. If the forms alone are real and if it is only when he is in pursuit of them that man achieves his highest activity, then the practice of moral virtues such as courage and justice would seem to be inferior to the practice of thought and philosophy. The Socratic maxim that "virtue is knowledge" takes on a new and esoteric significance. How does Plato re-establish the links between truth and goodness, reality and value, philosophy and morals? He holds that all the various forms are themselves made intelligible by their relation to one supreme form, the Form of the Good. (Later followers of Plato called this supreme form "the One" and criticized the term "good" in this context.) Plato was probably moved to use the term "good" by his general theory that anything is called "good" if it performs its function in a wider unity (the eye in the body, the citizen in the state). Hence the forms could be understood as good only by their relation to a higher unity which would be good underivatively.

Further corollaries of Plato's theory of forms were that supreme virtue (knowledge) was possible only for a few selected and trained people; and that here on earth their supreme activity (philosophy) could occupy only part of their time. For they would have also to rule their cities and to control and satisfy within reasonable limits their own desires (or "lower natures"). In these subordinate activities the virtues of practice and action—justice, temperance, etc.—would be required. As for those who were completely incapable of philosophy, their lives would be wholly practical and would have to be governed by belief or opinion, not knowledge. But, in a good state, their opinions would be right opinions, the results of a discipline imposed by the philosopher-kings. Plato recognized, in addition to reason and the bodily desires, the existence of a third psychological element, called by him "spirit" (*thymus*), which is the spring of action of aggression, pugnacity, competition and ambition. He then related these three psychological elements to the traditional four cardinal virtues. Wisdom is the virtue of the rational part of the soul, courage the subordination of "spirit" to reason (compare our use of "moral courage"), temperance the subordination of bodily

desires to reason and justice the harmonious development of the whole self.

It remains to ask what position Plato gave to pleasure. Is it an essential element in the good life? Are all pleasures good? Plato's arguments on this in the *Republic* are confused. He maintains sometimes that a pleasure which accompanies the satisfaction of a felt want is unreal, at other times that only the pleasures which accompany the satisfaction of the desire for knowledge are real, since only the objects of knowledge are real; but on the whole he seems to treat pleasure as a by-product of effective achievement and not as an element in the good.

In the *Philebus* these confusions are cleared up. It is clearly stated that pleasure is an essential element in the good for man, though an element subordinate in value to knowledge. Plato also seems to maintain throughout that it is necessary to "justify" the highest life for man in terms of the pleasure that it produces; but in the *Laws* he says this is because we are "discoursing to men not gods," as if these pleasure arguments were a concession to human weakness.

3. Aristotle.—Aristotle was a pupil of Plato's for 20 years and his views particularly on ethics were determined both in their resemblances and in their reactions to those of Plato. He distinguishes between those matters which admit of exact and infallible knowledge and those in which only probable and approximate conclusions can be drawn. Mathematics and theology belong to the first class, moral and political theory to the second. But he retained Plato's view that knowledge of the first kind is the highest possible human achievement and the life of contemplation of these eternal truths the highest life. Moral principles, however, have to be discovered inductively by examining the opinions and actions of men, particularly of good men. Men become good by education, but education does not mean literary or technical or intellectual education but the character training that a man receives who is brought up in a good family and a good city.

Aristotle also grappled with the Socratic paradox that no man can know what is right and do what is wrong. He rejected Plato's solution of it, namely that when this seems to occur it is opinion and not knowledge of the good which is conquered by desire. Yet in the end he too accepts the paradox. He points out that a man may be said to "have" knowledge which he is not using at the time (as when he is asleep). He must not only know the right principles of action but also know the particular facts of the case which make a particular principle applicable. This latter knowledge, since it concerns particulars, is less truly knowledge than is the grasp of principles (again the Platonic influence appears). The solution finally reached is that, when a man is said to know what is right but not to do it, this knowledge of particular facts is rendered latent and inoperative by desire. We still say that he knows because (1) he knows the right principles of action and (2) he has previously appreciated and, when his desire or anger dies, will again appreciate that this is the kind of situation which makes such and such a principle relevant, but at the moment desire or anger blind him to this fact. But this means that Aristotle is still accepting the paradox and holding that a man cannot fully appreciate that an act is wrong at the moment at which he does it.

Aristotle found a general formula for right action in his doctrine of the mean. This has been much misunderstood, as if it recommended mediocrity or setting one's mark not too high. What he maintained is that each kind of situation stimulated a certain kind of emotion or action: fear, anger, expenditure, punishment. In any situation there is one degree of the naturally appropriate emotion or action which is correct. Vice consists in exhibiting too much or too little of it. Fear is the reaction appropriate to danger: too much is cowardice; too little is foolhardiness; the right amount is courage. Spending is the action appropriate for the possession of great wealth: too much is vulgarity; too little is meanness; the right amount is liberality. If this seems to reduce vice and virtue too much to quantitative terms, Aristotle reminds us that, besides the right amount, there is also the right time and place, the right people to whom to respond, and so on.

Aristotle also tackled the problem of pleasure, but with some inconsistency. Sometimes he described it as the unimpeded activity of a natural human function and says that there is no reason why it should not be the good. Elsewhere he says pleasure accompanies an impeded activity and sets the final seal on it. On this view the activity must be judged independently and the pleasure that it gives is incidental or additional to its own intrinsic value.

Much misunderstanding of Aristotle has been due to a mistranslation of his word *eudaimonia*, which is usually rendered "happiness." This translation turns Aristotle into a utilitarian saying that happiness is the only thing worth having and that a life is to be judged good if it makes the man who lives it happy. But the word really means "good life" or "well-being" or "blessedness." The question "What is *eudaimonia?*" is the same as the question "What life is best for man?"

It is noticeable that both in Plato and in Aristotle there is no recognition of what we should call "duty" or of the correlative notions of sin, guilt and free will which we owe to the Jewish tradition and to Christianity. For Plato and Aristotle, each man necessarily does what his own character prompts. His character is the result of his natural gifts as developed by his moral education. The problem of ethics then becomes the question which natural or acquired characteristics are the best and how they are to be fostered or instilled.

Thus the Greeks found it difficult to draw a line between moral excellence and other forms of excellence—beauty, charm, health, intelligence. This supported the aristocratic tendency to restrict the best life to a few people especially favoured by talents, birth, wealth and education. We think a man may be a good man and do his duty in the highest degree no matter what his profession, whether he is rich, handsome, healthy and intelligent, or poor, ugly, ill and slow witted. Plato referred to manual work as degraded and Aristotle said that there were some jobs in which it was impossible for a man to be virtuous. The Christian tradition finds moral goodness in doing one's best with whatever natural endowment and in whatever circumstances one may find oneself. It links this moral worth with freedom of the will. Unlike the Greeks, Christians do not hold that what a man does is wholly determined by heredity and environment. A man can choose between right and wrong and do wrong knowing it to be so. He can feel remorse and allot blame, whereas in the Greek view he should be able only to feel regret and express distaste.

4. Cynics and Cyrenaics.—We have traced the development of the line which ran from Socrates through Plato to Aristotle. We must now turn back and follow other threads which also originated in the Socratic circle. Both Aristippus of Cyrene and Antisthenes the Cynic had learned from Socrates that the secret of good conduct was rational control of impulse and rational criticism of the dictates of custom and tradition. They both insisted on the need to control conduct by a theory and to maintain a calm, firm and critical spirit in the face of the vicissitudes of life.

Aristippus (*see* CYRENAICS; HEDONISM) argued that all that is beautiful or admirable in conduct is so because it is calculated to achieve some end. We must therefore look first for something which is good in itself and not for its further consequences. This supreme end he found in pleasure. He did not restrict the meaning of this term to bodily pleasures, though he admitted that they were the keenest. He accepted the existence of purely mental pleasures such as joy in the prosperity of one's native land. He fully recognized that pleasure as a goal must consist in a series of successive pleasurable experiences; and he even insisted on the importance of concentrating on the immediate pleasure and not troubling too much about a future which is dubious and incalculable. It was in the calm, resolute and skillful achievement of such pleasures as circumstances might afford from moment to moment, undisturbed by passion, prejudice or superstition, that he conceived the quality of wisdom to be exhibited. He thought that a wise man would not conform to the rules of current morality further than was needed to avoid the penalties of their violation, though he held with Socrates that these penalties rendered conformity reasonable in most cases. Thus early appeared the most extreme form of hedonism.

Antisthenes and the Cynics (*q.v.*) took a far different view. They held that Socratic wisdom was to be exhibited not in the

skilful pursuit but in the studied disregard of pleasure. Pleasure, declared Antisthenes, was evil. He extended this condemnation to most other objects of ordinary men's desires. By insight and self-mastery, he thought, man might achieve a complete spiritual independence which left nothing wanting for perfect well-being (*see* DIOGENES). The things that men call evil—poverty, painful toil, disrepute—he regarded as positively useful in attaining spiritual freedom and virtue. Men should find their ideal in freedom from human desires, their superiority in rising above them, their goal in detachment and independence. The Cynics had, however, no positive criterion of virtue and no standard of action to be achieved by knowledge, only a mere negation of pleasure and desire. Their freedom was an empty and undirected freedom; their knowledge was of evil rather than of good.

5. Stoicism.—It was noted above that Plato and Aristotle found it hard to draw a clear line between moral excellence and other excellences and to disentangle morality from a dependence on gifts and talents, on wealth, birth and prosperity. The Stoics were the first to draw a sharp line between the moral and the worldly, between virtue of character and of other forms of goodness, in a way which caught the imagination and compelled respect for their ideals. They regarded practical goodness, not theoretical contemplation or enlightenment, as the highest manifestation of human wisdom; and they bound the common notions of morality together in a coherent system, by a formula which comprehended the whole of human life and exhibited its relation to the ordered processes of the universe. In their practical ideals they owed most to the Cynics, particularly the fundamental tenet that the practical knowledge which is virtue and the character which is inseparable from it are alone to be accounted good. They seem generally to have regarded the eccentricities of Cynicism as an emphatic way of expressing the essential antithesis between philosophy and the world, an attitude not normal or necessary but one which might advantageously be adopted when the contrast required especially violent emphasis.

The Stoics rejected the ideal of virtue which we saw emerging in Plato and Aristotle, the ideal of a life of controlled and harmonized desires—desires limited but satisfied within these limits by reason. Passion or desire they held to be a morbid condition of the soul, involving erroneous beliefs about what is good. The truly wise man will be immune from these errors; he will of course be conscious of physical appetite, but he will not suppose that such appetites can in any degree or under any limitation be good or for a good. He will undergo pleasure and pain but these experiences will not cause him joy or grief or upset his calm conviction that they are matters indifferent to a reasonable will.

That this impassive and detached sage was not to be found on earth the later Stoics at least were fully aware. They hinted that one or two heroes of bygone days might have realized the ideal, but they admitted that even the best of living men were only progressing toward it. But this admission did not diminish the rigour of their demand for absolute loyalty to the claims of wisdom. Whatever was not of knowledge was sin; and since the distinction between right and wrong was absolute, admitting of no degrees, all sins were equally sinful. Similarly all wisdom was somehow equally involved in any manifestation of wisdom commonly called a particular virtue; though whether these virtues were specifically distinct, or only the same knowledge applied in different circumstances, the Stoics do not seem to have agreed.

It has been noted above that Aristotle failed to escape the Socratic paradox that no man can know the right and do the wrong. To the Stoics, with their enthronement of wisdom as the supreme good and its identification with the whole of virtue, this problem was even more acute. If vice is ignorance, then either vice is involuntary or ignorance is voluntary. The Stoics accepted the latter horn of the dilemma as the less dangerous to morality. But they did little to justify this acceptance and it was in sharp conflict with their other views. For if all ignorance is voluntary, this involves a vast extension of the field of human freedom, yet their view of the physical universe was thoroughly deterministic. The Stoics held that a man could avoid error if he chose to exercise his reason. His choice would be determined by his own force and firmness of soul. Moral responsibility was adequately safeguarded if vice was the result of a man's own nature and not due to any external cause.

It remains to consider how the Stoics avoided the wholly negative results of the Cynic views. How could virtue be knowledge unless we can say what this knowledge apprehends? "Knowledge of the good," said Plato, but this was an objective transcendent good, other than goodness of human character. Moreover, how could there be any connection between this view of wisdom and the particular duties incumbent on man? The Stoics answered these questions by their striking doctrine of "nature." To live by wisdom is to follow "nature." But this cannot bring us into any relation with the world around us unless it too is an expression or embodiment of wisdom. So to the Stoics it was. The world was organized and filled by divine thought (which they conceived materialistically as a pure and primary substance, a subtle fiery ether). This pantheistic doctrine had a double effect on Stoic ethics. First, it showed how the status of wisdom in human life was only a special case of the status of wisdom in the universe. It gave to ethics a religious emotion. Reason was the same in Zeus the lord of the universe and in man—his divine spark, the God within him. Second, it brought morals back into contact with the world. For even the lower nature of man is divinely designed and naturally fitted for its place in the world. Thus when reason develops it will recognize and prefer certain desires and certain outward objects, though it is still emphasized that in themselves all these desires and objects are "indifferent" and concern the wise man only because reason approves them. His real good lies in the exercise of reason's choice not in the things chosen.

The same argument was extended from the physical to the social sphere. Reason exhibits the natural basis of justice in the evidence that man was born not for himself but for mankind. All rational beings form a single community under a single law, the law of nature or reason. That the members of this "city of Zeus" should observe their contracts, that they should abstain from mutual harm and that they should combine to protect each other from injury were obvious points of natural law. Again it is clearly necessary for the survival of human society that its members should form sexual unions, produce children and bestow care on their rearing and training.

In the strict theory no government or law would be binding save that of the sage. He alone is the true king. But the revolutionary implications of this doctrine were not usually drawn out, and the rational law of an ideal community was seldom distinguished from the traditions and customs of actual societies. It was a fundamental maxim that the sage was to take part in public life, but it does not appear that his political action was to be regulated by any principles other than those current in his community. Similarly, in the view taken by Stoics of the duties of social decorum and in their attitude to popular religion, we find a fluctuating compromise between the disposition to insist on remote and purely rational standards condemnatory of what is conventional and the disposition to revere what is established, each tendency reflecting one interpretation or the other of the principle of "conforming to nature."

Among the ends natural to man and therefore to be preferred by reason the Stoics included freedom from bodily pain; but they refused to find, even in this outer court of wisdom, a place for pleasure. They held that pleasure was not an object of uncorrupted natural impulse, but an "aftergrowth." Nor did they restrict this status to bodily pleasures. Even the "joy and gladness" which accompany the exercise of virtue were regarded by them as only an "inseparable accident" or by-product and not as the essential constituent of well-being. It is only by a later modification of Stoicism that cheerfulness and peace of mind are taken as the ultimate end to which the practice of virtue is merely a means. Yet it is probable that the serene joys of virtue which the sage was supposed to experience among the worst tortures and misfortunes formed the main attraction of Stoicism for ordinary minds.

6. Epicureanism.—The Stoic view of pleasure was an attempt to resist hedonism on its strongest ground, namely in its appeal to the love of pleasure natural to all living things. It is therefore

interesting to see how near together in the end the two great rival systems came. The Epicurean sage no less than the Stoic is to be happy even on the rack. The Stoic found his perfect state in the exercise of wisdom which is independent of time (since, being perfect, it cannot be increased by duration). The Epicurean, too, could find perfection even in a happiness limited in duration when he realized the natural limits of life. The Epicurean, like the Stoic, attempted to present an ideal free of all the imperfections of human existence. Here is the key to the difference between him and the more naïve hedonist of the school of Aristippus. Hedonism gave the simplest and most obvious answer to the question of the good for man. But, besides offending the moral consciousness, it failed to provide that "completeness and security" which, as Aristotle said, "one divines to belong to man's true good." Philosophy, in the Greek view, should be the art as well as the science of the good life; hedonism would seem to be only an uncertain and bungling art of pleasure. Hence later Cyrenaics felt compelled to change their fundamental notion: Theodorus defined the good as "gladness" depending on wisdom, as distinct from mere pleasure; and Hegesias proclaimed that happiness was unattainable and that the chief function of wisdom was to render life painless by producing indifference to all those things which give pleasure. But by such changes their system lost the support that it had had in the pleasure-seeking tendencies of ordinary men. It was clear that if philosophic hedonism was to be established on a firm basis it must combine what the plain man naturally sought with what philosophy could plausibly offer. Such a combination was effected, with some little violence, by Epicurus, whose system with all its defects showed a remarkable power of standing the test of time by attracting the firm adherence of generations of disciples over a period of 600 years.

Epicurus returned to the basic tenets of Aristippus. Pleasure is the sole good and pain the sole evil. No pleasure is to be rejected except for its painful consequences, no pain chosen except as a means to greater pleasure. The stringency of all laws and customs depends solely on the legal and social penalties attached to their violation. All virtuous conduct and all speculative activity find their justification only in contributing to the pleasantness of the agent's life. Epicurus insists that by pleasure he means what the plain man means; if the gratifications of sense and appetite are discarded the notion becomes empty of all significance. The originality of Epicurus lay in his theory that the highest point of pleasure is attained by the mere removal of pain or disturbance, after which pleasure admits of variation only, not of increase; thus the utmost gratification of which the body is capable can be provided by the simplest means; and "natural wealth" is no more than any man can earn. Further he opens an ample area for the exercise of the philosophic intellect when he teaches that the attainment of happiness depends almost entirely on insight and right calculation, fortune having very little to do with it; that the pleasures and pains of the mind are far more important than those of the body because of the accumulation of feeling caused by memory and imagination; and that an indispensable condition of mental happiness lies in relieving the mind of all superstitions through a thorough knowledge of the physical universe. Again, when he stresses the misery that the most secret wrongdoing must cause from fear of discovery and when he exalts to the skies the value of disinterested friendship, he shows a sincere if not wholly consistent intention to avoid the offense that egoistic hedonism is apt to give to ordinary human feeling. Epicurus himself was a man of the widest human sympathies. The genial fellowship of the philosophic community which he gathered round him in his garden remained a leading tradition of his school; and certainly the ideal which Stoics and Epicureans equally cherished of a brotherhood of sages was most easily realizable on the Epicurean plan of withdrawing from political conflict to simple living and serene leisure in imitation of the gods, living aloof from the fortuitous concourse of the world. No doubt it was the practical rather than the theoretical side of Epicureanism which gave it so strong a hold and so long a history.

7. Roman Stoicism and the Academy.—The two systems which have just been described divided between them the adherence of most of the ethical thinking of the ancient western world from the time of their almost simultaneous origin to the end of the 2nd century A.D., when Stoicism almost entirely vanished from view. But beside them the schools of Plato and Aristotle maintained a continuity of tradition and a more or less vigorous life. Philosophy as a recognized element in Greco-Roman culture was understood to be divided among these four branches. Aristotelianism (*see* PERIPATETICS) showed no significant changes, and the Epicureans became a sect rather than a school. But both Stoicism and the Platonic (or Academic) tradition developed new emphases in this period. In Stoicism we can follow these changes easily as almost the only Stoic writings we possess are those of the Roman Stoics of this period. They reflect both a natural inner development of the system and also the influence of Roman ideals upon it. The figure of the Stoic sage already had affinities with the Roman standard of manliness, but emphasis came to be laid more on the moral virtues than on intellectual or scientific interest, in accordance with the Roman preference for practice over theory. The gap between the ideal sage and the actual philosopher also impressed the Romans and led them to rely less on the power of individual human reason and to lean on the thought of God. Pious resignation replaced self-poised indifference; the idea of the world and the flesh as a field for the exercise of reason yielded to a positive aversion from the body as an alien element imprisoning the spirit; the body is "a corpse which the soul sustains" (Epictetus), and life "a sojourn in a strange land" (Marcus Aurelius).

These echoes of Platonism in the Roman Stoics had a parallel in the development of Platonism itself. For a time the Academy continued to defend the ideal of a harmony in the whole human soul under the guidance of reason, the ideal that Antiochus had laid down as the true Platonic doctrine. But in Plutarch emphasis on the imperfections of the world reappears, along with a revival of the value attached to supernatural communications in dreams, oracles and warning voices, communications for which the soul must be prepared by tranquil repose and the subjugation of the senses.

8. Neoplatonism.—The system of Plotinus (A.D. 205–270) is a striking development of that side of Platonism which has had most fascination for the medieval and even for the modern mind but which had almost vanished from sight in the controversies of the post-Aristotelian schools. At the same time the Neoplatonists maintained the most reverent adherence to Plato and showed, in their commentaries on his works, that their own doctrine truly followed that of the master. Plato had held that the real was the intelligible. The things of the everyday world, whose imperfections failed to embody the eternal forms, were unintelligible unrealities. Plato had therefore no word for "matter," the Aristotelian counterpart of form. As we saw above, Platonism involved a progressive turning away from the world and the body. But these were not yet in Plato the objects of positive moral aversion. In Neoplatonism, however, the inferiority of the position in which the embodied soul finds itself is more intensely and painfully felt. Matter is recognized as the "first evil," and from it is derived body, that "second evil" to whose influence is due all that is bad in human life. The good for man now becomes the pure existence of the soul, which apart from the contagion of the body is perfect and free from error or defect: if it can be restored to this untrammeled activity of its original being, nothing can impair its perfect well-being. It is only the lowest form of virtue—the civic virtue or "justice" of Plato's *Republic*—that is engaged in regulating those animal impulses whose presence in the soul is due to its mixture with the body. The highest virtues are purifications from this contagion until finally the goal is reached and the soul has no bodily communications left and is turned wholly to the true and the eternal. Plotinus himself is still too Platonic to hold that the absolute mortification of the natural bodily appetites is essential for the purification of the soul, but this positive ascetic inference was relentlessly drawn by his disciple Porphyry.

There is, however, a yet higher point to be reached in the upward ascent from matter, and here the divergence from Plato is the more striking because it is based on indubitable elements in Plato's own teaching. For Plato the real is the intelligible: the

further the mind advances the clearer its thought and understanding become. But Plato had spoken of the Form of the Good as "beyond reality"; and at least one plausible interpretation of his dialogue the *Parmenides* is that which the Neoplatonists adopted, namely that there must be a unity which is beyond all plurality and relation. Plotinus pointed out also that knowledge involves a duality of soul and form, of subject and object. Indeed the position of the soul in Plato's metaphysic is very paradoxical. A soul is not a form and only forms are real: a soul is a particular individual being, and particulars are not fully real. Yet the soul is immortal and, especially in its function of reason, is divine and akin to the eternal. Still more difficult for Plato is the idea of God. For God too is not a form but a self or soul. Plotinus therefore urged that the final reality, the One, must be a unity beyond and above all distinctions of subject and object, of soul and form, of one soul from another, of souls from God. For Plato plurality everywhere involves a unifying principle more real than the plurality that it unifies. Plotinus drives this doctrine right home. Thus the highest human experience is one in which the soul apprehends this unity, or rather (since "apprehends" still suggests the subject-object relation) one in which the soul loses its individual nature in union with the divine. Thought and reason vanish and the final experience is ineffable and unique. Porphyry tells us that his master Plotinus attained this state of mystic "ecstasy" four times in the six years that he spent with him.

Neoplatonism, originally Alexandrian, is often regarded as a Hellenistic rather than a Hellenic product, the offspring of the mingling of Greek with oriental civilization. But, however great the similarity between Neoplatonic views and certain forms of oriental theosophy, the processes by which these views were reached were thoroughly Greek; it was by a rigorously reasoned development of certain essential features of Platonism that the Hellenic pursuit of knowledge came to end in an ecstasy beyond knowledge and that the Hellenic idealization of man's natural life led to an antipathy to the body and the world. At the same time we must not overlook the affinity between the doctrine of Plotinus and the remarkable combination of Greek and Hebrew thought which Philo Judaeus had expounded, two centuries before Neoplatonism too was developed, in conscious antagonism to the new religion which had spread from Judaea and was already threatening to conquer the Greco-Roman world.

It was moreover this Neoplatonism that furnished finally the chief theoretical support for the last desperate struggle under Julian to retain the old worship. (H. St; J. D. M.)

II. MEDIEVAL ETHICS

The beginnings of medieval ethical speculation must be sought in Christian theology and in the system of St. Augustine. The chief medieval thinkers, Abelard, Anselm, Bonaventura, Thomas Aquinas, Duns Scotus, William of Ockham, Maimonides and the thinkers of the early Renaissance can then be considered, as nearly as possible in chronological order, together with some of the leading topics of debate which they raised. While a certain continuity of development can be traced, some effort has been made in the ensuing sections of this article to show the variety of opinion that prevailed; and the appearance of innovations has been duly noted.

The sublime moral teaching of Jesus Christ, transmitted to mankind in the Sermon on the Mount and in the apostolic teaching, made it inevitable that Christian concern with morality should have soon become a matter of deciding what moral conduct was inadmissible in a candidate for baptism; and the lists of forbidden occupations given in the various versions of the *Traditio apostolica* of Hippolytus give sufficient proof of the care that was exercised to preserve the Christian body as a third race, set apart from Jews and pagans. Those who after baptism returned to pagan ways of cruelty, lust or avarice were made the subject of legislation at ecclesiastical councils; and gradually there was formed a body of Christian law on moral matters, law that was codified in the canonical collections which begin to be formed after A.D. 500 and in the many penitentials or tariff books of the public penance that was required to be undergone for various sins. That this process of forming a Christian law did not go to the lengths of the Jewish Talmud was due to occasional reactions provoked by the memory of the origins of Christianity and its entire disapproval of Jewish legalism. At the same time the greater part of Christian writing on moral topics down to the 12th century is occupied with prŏblems of casuistry arising out of the application of the new Christian law to varying situations.

1. Augustine.—The few great exceptions among early Christian writers who do discuss the roots of law and the nature of moral activity are soon enumerated. St. Ambrose (340?–397) wrote a work *De officiis* on the Ciceronian model in which he combines Stoic ideals of the life of virtuous harmony with the Christian idea of happiness in the vision of God; he also introduces into Christian thought the idea of the cardinal virtues. St. Augustine (354–430), his great convert, was so penetrated by the Christian doctrine of original sin that he could not help being much preoccupied with theories of morality. Taking from Plotinus (and also from some patristic writers) the idea that evil was not a substantive reality but an absence or "privation" of right order, he was able to overcome the Manichaean arguments which had so long held him in thrall. It is now generally held that Augustine received most of his Plotinian teaching through Ambrose, who was already hard at work Christianizing Plotinus. Evil being a lack of order, the eternal law is consequently defined as the divine reason, that is, the will of God, bidding the natural order to be preserved and forbidding it to be disturbed. This eternal law is imparted to the minds of men by a divine illumination, but Augustine is not quite clear or consistent in making of this illumination a freely given grace or a strictly natural activity.

Augustine takes from Plotinus the three stages of the ascent of the mind to God (purgation, illumination and union); but whereas for Plotinus the final ecstasy is a blind and passive contact with God, for Augustine it is vision and activity, free from disturbance but not from action: it overflows upon the body that has risen from the dead and its effects extend to other men, for no one who is perfectly happy in loving God could entirely forget the other great precept of loving one's neighbour. For Augustine as for Plotinus the cardinal virtues help a man toward this goal, first of all at the civic level and then at the levels of purgation and illumination, where they have to be acquired again in a transposed sense. But Augustine has in addition the three virtues of faith, hope and charity, divinely infused at baptism, to supplement them; and, especially against Pelagius, he is not inclined to stress the human part in the gaining of this goal of happiness. He rejects the Plotinian transmigration of souls; but elsewhere he is conditioned by his time, as when he holds that slavery is according to the natural law in a sinful world, though in a world without sin it would not be. In spite of this it can be seen how the doctrines of Christianity were working as a leaven in the mind of so great a thinker as Augustine; the disputes about nature and person in Christ pointed out a distinction, unknown to ancient philosophy, between nature and person in man and thus prepared the way for the idea that the slave too was a person and not merely an animate tool. Similarly, the Mariology of Augustine and his defense of virginity were leading him to a higher ideal of womanhood than antiquity knew. Above all, the example of Christ's sacrifice and His greater love set a standard to which, as Augustine says, Christians could look but could not reach.

Augustine has been accused of abandoning his belief in the freedom of the will, under stress of the Pelagian controversy, by admitting that God can give irresistible graces. This is a misconception due to a reading of some of his words (*De correptione et gratia*, 38) in the context of Reformation controversies to which they do not directly apply. In his *Retractationes* (i, 23) at the end of his life he still upholds free will; and in his *De civitate Dei* (xxii, 1) of the same period he gives as his answer to the problem of moral evil the principle that it is better for God to have given men freedom and to draw good out of evil than to have made men automata who could not sin at all. To Augustine the will is its own cause in acting, and he illustrates this paradox by citing another one, that the intellect in knowing an object should at the same time know itself. Concupiscence

(or the revolt of the appetites), which is produced by original sin, has left the will in a state of ill health but has not taken away liberty. The will can be borne along by grace like a ship under sail, but there is still the human possibility of furling sails and dropping anchor.

There are unfinished places in the ethical system of Augustine, such as the confusion that he leaves about the problem of altruism and his distinction between wisdom and knowledge, which has to be more fully elaborated by the medieval philosophers; but by the range and variety of his writings and by the depth of his psychological introspection he provided endless matter for the use of later speculative thinkers. The Pseudo-Dionysius (*see* DIONYSIUS [the Areopagite]), who influenced the west through versions made by Johannes Scotus Erigena (*c.* 858) and later translators, did not effect such changes in philosophy as he did in spirituality, for the amount of Neoplatonism that had entered the stream of thought by way of Augustine was such that it was harder for a second dose to attract notice. A word should be given to a work later called *The Theology of Aristotle,* which came from Greek to Latin by way of Syriac and Arabic and was in reality written by a Christian Neoplatonist who made large borrowings from Plotinus. This work was known in the west just in time to convince St. Thomas that he was on the right lines in giving a Christian twist to the interpretation of doubtful places in Aristotle, as from this esoteric doctrine of his it seemed clear that he must have been a Christian at heart.

2. Abelard.—Abelard (1079–1142) really inaugurates the medieval period of moral speculation in the west, taking up a loose end from Augustine and saying that all that matters about moral activity is the intention with which acts are undertaken. This challenged the axiom of the Pseudo-Dionysius that the goodness of an act depended on the goodness of all its circumstances. Abelard says: "We call the intention good which is right in itself, but the action is good not because it contains within it some good but because it issues from a good intention. The same act may be done by the same man at different times: according to the diversity of his intention the act may be at one time good, another time bad. . . . All sins are of the mind only, for there alone can be the crime and the contempt of God where is the seat of the knowledge of Him and where reason resides." To an age so completely extrovert as the early middle ages, which had been brought up to regard as wrong so many external acts listed in the penitentials, this teaching was more than startling. Abelard was condemned at the Council of Sens (1140) at the instance of St. Bernard, but his advocacy of the importance of intention was so far admitted that an idea of inculpable ignorance (as distinct from various culpable kinds) was elaborated by the schoolmen which sometimes exempted from guilt an externally wrong act. Peter Lombard, author of the manual of theology known as the *Sententiarum libri quatuor,* gives a representative 12th-century reaction to Abelard's views when he lays down that a man's actions are judged good or bad according to their cause and intention, except such acts as are of their nature evil. This supposes in all men a spontaneous knowledge of some at least of the precepts of the natural law, and the problem has been carried a stage further. St. Thomas will restore the maxim of the pseudo-Dionysius that moral goodness requires both good intention and good external act, but not until many refinements have been made in the problem.

3. Bonaventura.—The arrival of Aristotle in the west came at an opportune moment (in Paris, *c.* 1205; in Oxford, 1206–09) for the development of ethical theory. What he had to say stimulated the schoolmen, who were just finding an appetite for the rational discussion of matters which had hitherto been treated incidentally to points of Scripture and canon law.

St. Bonaventura (1221–74) sees much more of the contrast between nature and grace than the harmony; before Adam's sin it was sufficient that he should read the mysteries of God in the book of Creation, but afterward men had need of the book of the Scriptures. Philosophers went astray by supposing that the condition of dislocation and imperfection that was observable about the nature of man was something that had belonged to it from the beginning: it was only the Christian revelation that let men see that this was a dislocation. Hence there is an urgency and a spiritual fervour about the writing of Bonaventura which is not so noticeable in the other schoolmen. When he does linger on a problem of pure moral philosophy, it is on those of the origin of moral ideas and of the way in which the virtues can be intensified. While admitting that by practice a habit of virtue can be strengthened so as to pass from the stage of civic virtue to that fitted for the way of purgation, even here he brings in the fact that grace will do much more for a man, making his actions no longer merely morally good but meritorious also; and he describes how the illumination of grace will even draw out a man's virtues from latency where there has been no habit-forming practice at all. Bonaventura distinguishes from the four cardinal and three theological virtues the gifts of the Holy Spirit. He was a man of spiritual experience, and to him the mystics turned when they wished to find their bearings in the labyrinthine ways of the spiritual life. (Duns Scotus, however, later identified gifts and virtues.)

4. Anselm.—The progress of scholastic thought was not confined to the infiltration of ideas from the realm of the theological. Much speculative work had been done on the nature of free will. Anselm (1033–1109) was the first to put the problem (in his *De libero arbitrio*) that if freedom means the power to do right or wrong, God cannot be free. He answers this by a carefully drawn definition of free will which makes it the power to preserve rightness of will (thus omitting the notion of sin or wrong entirely), and rightness is defined as the will to do what God wills. With his usual neatness of dialectic Anselm proves that not even God can take away this liberty from man.

5. Thomas Aquinas.—St. Thomas Aquinas (1225–74) does not adopt the whole of this speculation of Anselm's, which, had it gained general acceptance, might have led to a medieval debate on the right and the good. Thomas seizes upon the Aristotelian distinction of wish and choice, which are directed to end and to means respectively, and applies it to God, who can thus be said not to be free in directing His will upon His own nature but free in His choice of all outside Himself. Man likewise resembles God in this, that he directs his will of necessity upon his end of happiness, which he cannot help desiring, while the means thereto fall under his free choice. The fact that God by a unique act of will is concerned with Himself and with creatures leads to further speculation about the nature of this act with its varying objects. But at this point human analogies begin to fall away.

Commenting upon the *Nicomachean Ethics* of Aristotle in 1271, Thomas adopts with approval the statement that among human operations the most pleasurable is the contemplation of truth, and this he holds to be accepted by all men. He enlarges on what Aristotle has to say about the reasons for this, explaining that the purity and enduring nature of this contemplation arise from its objects, which are free from matter and from change. Its pleasure comes from the very fact that it is so unlooked for by the common run of men. He takes over Aristotle's distinction between the pleasure of search and the pleasure of possession in contemplation, but whereas Aristotle rejected the pleasure of search, Thomas makes it out to be different only in degree and not in kind from that of possession.

So far the discussion has been about what is possible in this temporal life. But now comes the great philosophical advance made by Thomas in the light of the Christian revelation. Aristotle had called upon his contemplative to play the immortal as far as he could, rather than to be ever bearing in mind his mortality. Thomas alters the emphasis: his contemplative is to direct his aim upon immortality to the utmost of his power (with the implication that he can always do this to some extent and must now try to do it to the full extent of his ability); he must strive to live an intellectual life, with his sensitive appetites subject to reason, for intellect is the best of his powers, being immortal and godlike. In fact the happy contemplative who lives an intellectual life and makes the good of his intellect his chief care is also most beloved by God. Aristotle would have found this conclusion near to blasphemy.

The problem that now arises for Thomas is whether he has to admit in man a natural desire for the vision of God. If he does, he would seem to destroy the gratuity of the supernatural elevation of man to that vision by Christ's merits (for God must be expected to implement any natural desires that He implants in man), while if he does not, he must sacrifice his whole reconstruction of Aristotle. Three of the solutions drawn out of his writings by his disciples in the debate on the topic must be cited here: some, following Cajetan, say that it is not a natural desire that men have for the vision of God but one that is supernaturally infused (with this they jettison the reconstruction of Aristotle); others, with Francisco Suárez, say that it is not a desire but a velleity, which God is not bound to fulfill; and a third group, with Francis of Ferrara (Francesco Silvestri), says that it is not a desire to see God in Trinity but to know more about His nature in an abstract way. Another consideration is that Thomas' distinction between perfect happiness for man (which consists in the supernatural vision) and imperfect happiness fits onto another idea, which he had earlier exploited (*In Boethium de Trinitate*, ii, 1, ad 7; vi, 4, ad 5), that though man is at an infinite distance from God, his natural destiny is to make progress Godward even if no one can tell him without revelation that there is any prospect of his arriving. Man is like Zeno's tortoise; he will have much pleasure of search and some partial pleasure of possession, but it is God's secret that He is going to meet man more than halfway.

The next ethical topic that Thomas had to handle was the meaning of natural law. Three definitions of it had descended to him in the tradition: the jurist Ulpian made it what nature taught all animals; Isidore (after Gaius and Justinian) said that it was what was common to all societies of men; Gratian identified it with what was taught as law in the Old and New Testaments, such as the Golden Rule. Bonaventura held to Ulpian's meaning, but Thomas broke away from the tradition to define it as a sharing by rational creatures in the eternal law of God, or a natural light in men's minds enabling them to tell good from bad. This return to Platonic and Augustinian doctrine bespoke Thomas' conviction of the ultimate rationality of the universe. Gratian is left to one side, as revelation is not called upon for this definition, while Isidore's position is turned by a distinction (not original to Thomas) between the primary precepts of the natural law, which none can fail to know, and the secondary, which are in fact unknown to vast numbers.

St. Jerome had bequeathed to the scholastics an interpretation of the four living creatures in Ezekiel's vision as reason, spirit and passion (the Platonic triad) together with a fourth element in man named in Greek *synteresis*, which was latinized as synderesis (a conservatory of moral principles) and approximated to conscience. St. Thomas made of this synderesis a natural disposition to recognize the first principles of the natural law. It could not err and it was incorruptible in man, but the application of those principles to a given situation made by conscience was far from infallible. The mistake was thought to occur always in the minor premiss of the moral syllogism: everyone knew that murder was wrong, but it was not always easy to say whether a particular act was murder. This treatment of synderesis means that Thomas would have given only a modified assent to the intuitionist position in ethics. One peculiarity of the Thomist view was the idea that man's first duty when he comes to the use of reason is to direct his life toward God. This conclusion did not mean that every man was bound to take the most effective means to carry out his obligation. Certain means were looked upon as necessary (at least in a negative way) to that end, and these were the precepts of the Decalogue, while other means were good but not necessary, such as the counsels of poverty, chastity and obedience, which enabled the end to be reached with more expedition and more sureness. In many of these positions Thomas was innovating, or else adopting innovations of his master, Albertus Magnus, while the traditional view was upheld in the Franciscan school by Bonaventura, who placed synderesis in the will and not in the intellect.

6. Roger Bacon.—Roger Bacon (*c.* 1220–*c.* 1292) affords a sufficient proof that not all Franciscans were conservative. His *Moral Philosophy*, which he sent to Pope Clement IV and of which the manuscript is preserved in the Vatican, follows Aristotle's *Ethics* for some distance, with Avicenna as a guide, but concludes with an innovation which might seem to foreshadow the Renaissance. Bacon says roundly that, since ethics is a practical science, its argumentation should be rhetorical rather than scientific, for the student needs something to help him to overcome the difficulties of following out the argument to practical conclusions; one rhetorical argument is better than a thousand syllogisms. This leads him to praise Seneca exceedingly, and a large part of his work is given up to quotations from the moral treatises of the younger Seneca. Another harbinger of humanism much used by Bacon is the *Corpus hermeticum*, a pagan (or Jewish) work which he acclaims as showing what the power of unaided reason could achieve in the direction of Christianity. Similarly, Aristotle is interpreted as saying that the desire of the human mind cannot be satisfied save with the Infinite, which must therefore be enjoyed forever in eternal bliss. From Aristotle too, though by a slight twist in the interpretation, comes the idea that knowledge is of little or no help toward virtue. Bacon thus makes a sharp rejection of the Platonic paradox that virtue is knowledge and emphasizes all the more the importance of rhetorical argument.

7. Duns Scotus.—John Duns Scotus (*c.* 1265–1308) might be regarded as a moderate progressive beside his fellow-Franciscan Roger Bacon. He took up many of the innovations of St. Thomas and refined upon them with his psychological subtlety. Thus he kept synderesis in its Thomist meaning, while allowing to the will a tendency toward good which he compared with gravity and called *pondus naturae*, distinguishing thus the will-as-nature from the will choosing freely. Though often accused of divorcing the moral law from reason and basing it upon the arbitrary will of God, Duns Scotus was not guilty of this step. He accepted the Thomist notion of natural law, but somewhat limited its scope. God can will whatever does not imply a contradiction; hence He could have created another universe in which morality might differ somewhat from the morality of this universe, but once granted that He has created man in this particular universe, it would imply a contradiction in God if He ordered men to hate each other. It would be equally impossible for God in this universe to save Judas, who is already damned, or to raise a mountain peak to the enjoyment of the beatific vision. What gives the natural law stability is the nature of God and of the creatures to whom it applies; there is an arbitrary element consisting of God's free choice of this world as the object of His creative activity. This element Duns Scotus tends to insist upon, in order to save the contingent character of the created universe against the Arabian theories of its necessary emanation from God. The distinction of God's absolute power from His regulated power (*potestas ordinata*) is found in Albertus Magnus and Thomas, but it is more fully worked out by Duns Scotus and will be exploited in a very different way by William of Ockham. If Duns Scotus seems to question somewhat the necessary connection of the Decalogue in all its parts with the eternal law of God, that is because (as is known from the Old Testament) He had at times dispensed certain men from the observance of some of its precepts. A distinction therefore seemed to Duns Scotus to be required between what was connected with man's last end by absolute necessity and what was not so connected. He regards it as not quite certain that, according to the natural law, God would show love toward every intelligent creature who loved Him, even though their duty to love God followed immediately from the nature of God. In this he shows deference to Aristotle's view that there can be no friendship between God and men.

It is easy to see the moderation of Duns Scotus when one examines the history of the disputes about the nature of conscience and synderesis which followed upon the strongly intellectualist stand made by St. Thomas. Walter of Bruges (*c.* 1225–1307), who commented in detail on the position of Thomas, was inclined to bring back a voluntary element into the picture of moral judgment and to effect a compromise between the Thomist intellectualism and the position of Bonaventura. Later Henry of Ghent (d. 1293) set aside the whole syllogistic apparatus of Thomist moral

judgment from the problem of what conscience was and declared that it was entirely in the will, being a choice of the will that never disagreed with right reason. The psychological machinery behind such a declaration and the whole system of Henry of Ghent have been much neglected by historians because of the general inaccessibility of his works, but what is significant for the development of ethical theory in the middle ages is the fact that Duns Scotus returned to the position of Walter of Bruges and devoted much of his energy to answering the arguments of Henry.

8. Ockham and the Averroists.—It can be seen from the list of questions publicly disputed at Oxford in the later part of the 13th century that the question of God's arbitrary power was always within range of the discussion. When William of Ockham (*c.* 1300–1349) came to take it up, he declared that God was not in any way limited in His treatment of man. The two powers, absolute and regulated, were one in God. Just as the pope or any other lawgiver could sometimes act according to his own laws but might set them aside on occasion, so God might be regarded by men as sometimes using His power according to His own regulations and sometimes quite arbitrarily. The fact that natural law cannot be dispensed with in the way that positive law can is not considered by Ockham when he uses this comparison. Gradually the close articulation of created nature with God was being loosened: the Latin Averroists (Siger of Brabant and Boethius of Dacia) had from 1260 onward maintained that God could not be the author of man's happiness, since the First Cause could not now produce any new effects. Hence the beatific vision was not merely beyond the scope of reason but was against reason: one might hold it to be true in theology but not in philosophy, and this "double truth" theory became the mainstay of Latin Averroism. Ockham could have appealed to it to protect his position when confronted with the fact that in some places he accepted the idea of natural law (which is based on the unchangeableness of created natures, once they are created), while on the whole he preferred to speak of the absolute power of God to disregard all laws whatsoever. That he did not do so (as far as can be ascertained from his writings) was probably due to his having come, through his nominalism, to attach very little importance to the specific natures of created things.

The struggle against Latin Averroism was waged largely through innumerable commentaries on the *Ethics* of Aristotle, for there were few texts more suitable for discussing the interplay of natural and supernatural than that work. Many of these commentaries remain unpublished; *e.g.*, those by the Englishmen John Peckham and (in part) Walter Burley (1275–1337). One can see how the Aristotelian theory of the virtues as means could clash with Christian standards: thus Siger defines humility as a restraint upon the desire of glory that keeps it within the bounds of right reason; and he denies that it can be humility to hide one's gifts and hold oneself cheap, whatever may be the teaching of the saints.

9. Maimonides.—Some of this polemic against Christian asceticism came from Jewish writings. Maimonides (1135–1204) had condemned what he called the sin of asceticism on the grounds that the Jewish law taught a man to aim just a little beyond what was strictly right according to the natural law, with the result that right conduct ensued, whereas the Christians went to the opposite extreme with their renunciations and forsook the mean of right reason altogether. On the question of the ceremonial law Maimonides is content to say that there is a distinction between laws the reason for which men can see and laws which they regard as impositions by God. The impositions are not really arbitrary, but the reason for them is a mystery.

Maimonides was a true follower of Aristotle in his not accepting any personal immortality as the goal of human life. The unique active intelligence, which operates in man during his life but is not part of him, is made somehow different for the fact of the individual's having lived a life of intellectual virtue; but it is immortal and he is not. It is true that Maimonides wrote a *Treatise on the Resurrection of the Body*, to put himself right with the synagogue; but it is not mentioned in his *Guide*, and the philosophy of the *Guide* leaves no place for it. This early example of acceptance of the "double truth" theory may have influenced the Latin Averroists. As matter is for Maimonides a privation of being, it is evil and is the source of all sin and imperfection. The four kinds of perfection at which men aim are the acquiring of external goods, the health of the body, the moral virtues (such as generosity, courage, humility, moderation) and the intellectual virtues. These last are the only fitting object of such aim, and a man should even accustom himself to imitate God in seeking the good for its own sake, regardless of his own personal survival.

10. Humanism and the 16th-Century Scholastics.—Giovanni Pico della Mirandola (1463–94) went back from the wilderness of logical distinctions in nominalism to the Platonic pastures and, putting God, as the Neoplatonists had done, above and beyond being, set man in the centre of the universe. He named Albertus, Thomas, Duns Scotus, Henry of Ghent and Giles of Rome as his masters in scholastic philosophy, but proposed to gather his honey from every flower. Of the 900 *conclusiones* which he offered to defend publicly in 1487, only 107 come from these writers. Yet he does not suppose that man can be perfectly happy unless God elevates him, nor does he regard human freedom as an end in itself. A like restraint is shown by Pico's great admirer, Thomas More (1478–1535), whose *Utopia* is a philosophical abstraction designed to show what knowledge and practice of morality might be possible without the Christian revelation. Apart from these two, most of the new humanists chose either to go from Plato to the pseudo-Dionysius and mysticism or, with Pietro Pomponazzi (1462–1525), to strip off from Aristotle all Christian accretions and present him in his full pagan rigour. The "double truth" theory now found little support, and many thinkers, such as Cajetan and Martin Luther, accepted theology as the only science giving truth about the ways of God and jettisoned philosophy (Luther) or at least the philosophical proofs of immortality and all that depended on them (Cajetan). Petrus Ramus in Paris defended the thesis that everything taught by Aristotle was false. The humanists jettisoned theology and devoted themselves to an ethic of this world. Rather than accept this dilemma a few thinkers revived the Thomist harmony of nature and grace, notably the Dominican Francisco de Vitoria (1486–1546) and Gasparo Contarini (1483–1542) who broke away from the Averroism of Padua in 1516. The existence of these isolated revivals became more significant when the *Summa* of St. Thomas was chosen as a textbook in all the newly founded Jesuit universities of the Counter-Reformation from China (1648) to Peru (1572). The ensuing revival culminated in the restatement of Thomas' theory of natural law by Francisco Suárez (1548–1617) in his famous treatise *De legibus*. (J. H. Cn.)

III. MODERN ETHICS

The intellectual atmosphere in which the middle ages ended was dominated by three factors: the growth of rival Christian sects, the birth of the nation-state and the rise of science. Each of these affected ethics in its own way. The Reformation withdrew large bodies of Christians from the moral authority of Rome, and they came to look instead to Scripture for their moral guidance. But this soon stimulated such a variety of selection, application and interpretation as almost to involve a principle of private moral judgment. There arose a demand for a moral standard independent both of ecclesiastical authority and of biblical texts. The disappearance of the single society which in theory Europe had constituted, with the pope as its spiritual and the emperor as its temporal head, coincided with a period of religious wars. The independent states which began to take over political authority were left face to face with one another. What were their relations to be? Naked power politics? Again there was a demand for principles of morality which should not be the monopoly of any one nation or even of Christianity itself. Finally, science had begun its work of subjecting all natural phenomena to free rational inquiry. It was inevitable that social and moral phenomena should be subjected to the same critical review, and the result was a demand for a rational basis for morality.

A. 17th and 18th Centuries

1. Grotius.—The man who first met these demands in a way

which could be called "modern" was Hugo Grotius (1583–1645). Echoing the Stoics, he saw the whole universe as dominated by a rational law of nature. For man, the law of nature is the collection of those rules which follow inevitably from his essential nature. Natural law is therefore intelligible and immutable. Man is distinguished from the animals by his "appetite" for tranquil association with his fellows and by his power of acting on general principles. If men are to live at peace with their fellows there are certain rules which must be observed: the keeping of promises, the recognition of human equality, the principles of equity and justice, of parental responsibility, of marital fidelity. Here then was that set of moral standards, independent of church and of Scripture, which the new Europe demanded. Here were rules tied to no narrow national tradition because based on the universal nature of man, rules to which princes and governments no less than private individuals should be subject.

The law of nature had long ago been identified with the *jus gentium*. In one sense this had meant that these rules could be discovered not only by deduction from the nature of man but also inductively by noting their universal acceptance in all human societies. But, for Grotius, *jus gentium* had the deeper meaning of a rule governing the relations between peoples considered as moral agents. Some of the natural laws cited above (marital fidelity, parental responsibility) could not apply to the behaviour of sovereign states; and indeed Grotius was inclined to emphasize one single rule as relevant here—the rule that promises are to be kept. What was important for ethics, however, was the conception of an absolute and objective moral code applicable even to infidels: Grotius insisted, against the medieval tradition, that the Turks had equal rights with Christian peoples under the rules of international law. Here was a code which was determined by reason on the basis of an understanding of human nature; a code which could be called scientific, though its model was the a priori science of mathematics rather than the empirical sciences now rising to view. It was the rapid and resounding success of Grotius' treatise *De jure belli et pacis* which brought the whole notion of natural law into prominence and set men's minds on a new path of inquiry in moral matters.

2. British Ethics.—Grotius had made his basic tenet the social nature of man and had asserted the power of human reason to determine moral law. It was on these two linked issues that British ethics concentrated for nearly 200 years. Is human nature social? What is the constitution of the human mind, and is it reason which lays down moral standards or must we find their source elsewhere in the psychology of man?

Hobbes.—The answers given by Thomas Hobbes (1588–1679) to these two questions marked the beginning of independent ethical theory in England. Hobbes's view of the world was materialistic. He held that all that really occurred in a man's history was a bodily process of which the mental or conscious feelings and thoughts are only appearances. He was also a thoroughgoing determinist. The process that we call deliberation is simply the occurrence in the mind of changing desires, hopes and fears which mirror the mechanical course of changes in the body. The will is the name that we give to the desire or appetite which occurs in this series immediately before an action and mirrors its immediate cause. Pleasure is a motion helping vital action and pain a motion hindering it. Hobbes shows some signs of accepting hedonism and making pleasure and the avoidance of pain the determining motives of all human action. But elsewhere he speaks of human desires as directed to various specific objects of which the chief is self-preservation. What is, however, continuously clear is his denial that human nature is social. All man's natural instincts and passions are self-regarding. Hobbes carries this doctrine through relentlessly: pity arises from the imagination that a like calamity might befall oneself; charity in general is a proof to a man's self of his own power and a delight in its exercise; the charity that helps a stranger is a contract to purchase service or a bribe to purchase peace. From the pursuit of private ends arises the desire for power, and this derivative desire tends to occupy more and more of man's attention as he becomes more and more able to foresee the future and to pursue his ends intelligently. Hobbes appeals to direct introspection in support of his views. But he adds that, if anyone is still inclined to believe that man is a social animal, he need only look around him. Bees and ants are social animals; they do not compete for honour and dignity and show envy and malice to one another as men do; they do not set private above public good, criticize and malign each other as men do. A man shows his opinion of other men by locking his doors when he goes to sleep, by carrying arms when he travels, thereby "accusing mankind by his actions" as Hobbes does by his words. What then of morality, of good and evil? Whatever a man desires he calls "good," whatever he dislikes he calls "evil." "The object of every voluntary act is some good to himself." "Moral philosophy is nothing but the science of good and evil . . . good and evil are names that signify our appetites and aversions."

Yet Hobbes also frequently refers to the law of nature which he sometimes calls the law of God; and his chapter on it in *Leviathan* (ch. 15) gives a list of moral rules identical with the Christian code and summed up, like it, in the maxim "Do as you would be done by." How is this to be reconciled with his other views sketched above? The originality of Hobbes lay in his answer to this question. The unbridled exercise of selfish desire and particularly of the desire for power would lead to conflict and a war of all against all. This condition would thwart the most powerful of all desires, that for self-preservation, besides presenting obstacles to the fulfillment of most other desires as well. Man therefore hates war and desires peace. But peace can be sustained only if men act justly, help their neighbours, keep their promises (and generally do as they would be done by). These "laws of nature" are therefore "articles of peace" or corollaries of the principle of self-preservation. But, says Hobbes, as long as there is no guarantee that other people will obey these rules, I cannot be expected to obey them myself. My knowledge of human selfishness and human passions tells me that every man will be ready to break such rules for his own advantage if he sees no indirect disadvantage accruing from the breakage. Thus men are driven to set up a sovereign, a man or body of men, which will enforce these rules. Until this is done, since I have no assurance that anyone else will keep the rules, I have no obligation (and indeed no motive at all) to keep them myself. They are means to my peace and self-preservation only if they are generally observed, not if they are observed by me alone. When, however, the sovereign is set up, whatever it lays down as the rule becomes the standard of justice and of injustice. For a rule that the sovereign does not enforce is one on which I can place no reliance and one which therefore I am not bound to obey. To set up my own view of justice against the sovereign's would be a move toward that anarchy which was the reason for the sovereign's appointment and which is the worst condition of all. (With the political implications of this argument we are not here concerned.) Hobbes's most telling argument in support of his view is his appeal to international relations. The behaviour of independent states is a model of what the relations between individuals would be like if there were no sovereign over them, no common power to fear. The dominating motive is self-preservation. In the absence of a "common superior" (namely, a world government), this leads to "war," which in Hobbes's use does not mean "battle" or continuous fighting, but rather "cold war"—a state of fear and suspicion and restless desire of power after power breaking out now and again into "battle" as one side or the other sees a chance of gaining an advantage. It would be folly in such circumstances to adopt the Christian rule, to disarm, to cultivate peaceful avocations, to do as you would be done by, when there is no guarantee that other powers will do likewise.

There is sufficient truth in Hobbes's conception of human nature, in his account of what life would be like if there were no government and in his analysis of international relations to make his theory attractive. The power of his style, the acuteness of his mind, the coherence and clarity of his presentation all added their weight to the inherent force of his argument. The result was that almost all moral philosophy in England for the next 100 years was explicitly or implicitly a reply to Hobbes. For his theory, with its linking of egoistic morality to absolutist politics,

was regarded with natural horror by traditional moralists and theologians everywhere.

Cudworth and Locke.—Ralph Cudworth (1617–88) was the most distinguished of the little group of thinkers at Cambridge in the 17th century commonly known as the Cambridge Platonists. In his treatise on *Eternal and Immutable Morality*, his main aim is the Platonic one of showing that "eternal and immutable distinctions of good and evil" are independent of mere will or desire whether human or divine. These distinctions have an objective reality, as readily cognizable by reason as are the relations of space or number. Nothing can be made just by being commanded by an authority; for it makes sense to ask whether any given command is just or not. Cudworth endeavours to refute the doctrine of Hobbes, which he calls a "novantique philosophy"—a mere revival of the subjective relativism of Protagoras—by the following *argumentum ad hominem*. Hobbes's materialism involves the conception of an objective material world revealed not to the senses, whose deliverances vary from man to man and from moment to moment, but to that active intellect which is the same in all. It is inconsistent to refuse to admit a similar exercise of reason in moral matters (which are clearly not decided by the senses) and thus an objective order of right and wrong which the mind by its power of reason clearly apprehends as such.

John Locke (1632–1704) too held that moral rules are binding independently of the commands of governments or of any relation to political society and that they are capable of being intuitively known as are the truths of mathematics. Yet he also held, inconsistently with this, that Hobbes was right in basing human conduct on self-interest and in interpreting "good" and "evil" as equivalent to "pleasant" and "painful" respectively. He reconciles these two positions by maintaining that we call those actions "morally good" or "morally bad" which bring pleasure or pain to us through the rewards and punishments attached to them by the supreme lawmaker, God.

Cumberland.—Richard Cumberland (1631–1718), in his treatise *De legibus naturae*, was the first to lay down that "regard for the common good of all" is the supreme rule of morality. He is thus the precursor of utilitarianism in its universal rather than its egoistic form. His supreme law of nature is thus stated: "The greatest possible benevolence of every rational creature towards all the rest constitutes the happiest state of each and all, so far as depends on their own power and is necessarily required for their happiness; accordingly Common Good will be the Supreme Good."

Shaftesbury and Hutcheson.—The 3rd earl of Shaftesbury (1671–1713) was the first English moralist to turn his attention to the structure of the human self and to examine the interplay of its various impulses and sentiments. In his *Inquiry Concerning Virtue or Merit* he begins by attacking the egoism of Hobbes. This egoism would be convincing if man were a wholly unrelated and solitary creature. But "if eating and drinking be natural, herding is so too . . . an isolated being cannot be happy." We can understand an eye or a cog only in relation to the organism or machine to which it belongs. So a man can be understood only as a part of a wider whole, the human species. We can call his nature "good" only if it is so adjusted as to contribute to the good of the whole to which he belongs. But we must consider not only how his actions affect the whole but also what their motives are. A sick man who cannot promote the public good is not vicious; a villain bound in chains who can do no public harm is still a villain. A good man, then, is one whose impulses carry him primarily and immediately and not indirectly or accidentally to the good of mankind. Shaftesbury recognizes that some of these direct public affections, such as love or pity, may be inordinate and thus wrong. But he argues that this is because they are partial and limited only to one group or person as when love of children leads to favouring them or spoiling them and so works against the good of the species as a whole. How does man come to know what is right and wrong? Which are those motives which directly concern the good of the whole human race? Shaftesbury says man has a power of looking into his own heart and inspecting his own motives and impulses. He calls this a "reflected sense": "reflected," because it is turned not on outward things or actions but back on a man's own soul so that he sees himself as if reflected in a mirror; a "sense," because it detects harmony, fairness, beauty, fitness, in his own nature as the aesthetic sense detects these features in musical numbers or visible forms. If a being cannot thus reflect on his own motives and impulses (and by imagination on those of others) he can have no awareness of right and wrong and so cannot be virtuous. Shaftesbury's use of "sense" here suggests that he is rejecting the view that moral right and wrong are ascertained by reason or intellect. But he admits that custom or superstition may derange this sense of right and wrong and he seems to believe that under these circumstances reason may be called in to correct it. Shaftesbury concludes by arguing that to have in the highest degree the sense of right and wrong, this steady tendency toward the good of the whole species, in fact leads a man to private happiness. The selfish man is morose, rancorous and malignant, ill-humoured, sour and disquieted, suspicious and jealous. No bodily satisfaction can make up for these mental ills. To be in harmony within, one must be in harmony with one's environment and especially with one's social environment. Thus it is to the private interest and good to work toward the general good; and virtue is the good and vice the ill of every individual.

Francis Hutcheson (1694–1746) developed these views of Shaftesbury into an elaborate system of moral philosophy, clarifying and defending them effectively. He held that our judgments of right and wrong are direct and immediate and are therefore to be ascribed to a moral sense and not to reason. They are reflex, that is, they express our admiration and pleasure in actions of ourselves and others independently of any advantage these actions may bring to us. But this moral sense is posterior to actions and passes judgment upon them. It presupposes that an action has already occurred (actually or in imagination) and approves or disapproves it. The moral sense is thus not a motive to action. The motive to moral action is benevolence and moral actions are those directed to the common good. Hutcheson notices that our benevolence is usually limited to those near us. But cases of a wider benevolence are to be found (as when we rejoice to hear that distant peoples are governed well or pity them in disaster); and near-sighted benevolence may sometimes (as in nepotism or jingoism) lead to actions which the moral sense cannot approve. Thus the widest benevolence is the best motive. The reason why benevolence normally acts in a limited field and is, like gravity, universal but weak at a distance, is that we can do most good near at hand, and attempts to consider the whole species would usually be distracting and incapacitating.

Butler and Price.—Joseph Butler (1692–1752) is probably the greatest British moral philosopher; not that his work is strikingly original, but that it combines weight and clarity of argument with caution and fairness in a way unequaled in any other author. His *Sermons* (1726) and *Dissertation on Virtue* (appendix to the *Analogy*, 1736) do not constitute a complete system of ethics, but the points that they make have a finality which compels acceptance. In the first place, Butler provides an unanswerable refutation both of hedonism and of the self-interest theory of human conduct. Pleasure, he points out, is the by-product which attends the attainment of ends other than pleasure. Men seldom pursue pleasure for its own sake, and the best way to attain it is to pursue other ends wholeheartedly. Moreover, among the ends that men pursue are quite certainly benefits for other people. Thus the only sense in which my desires are all self-interested is that they are all *my* desires. Butler sharply and rightly distinguishes the operation of "particular passions," such as hunger or maternal affection, from the operation of "self-love" (the rational pursuit of one's own happiness) and from "benevolence" (the rational pursuit of the general happiness). Above all these three kinds of motive, he places "conscience." With him, as with Shaftesbury and Hutcheson, conscience is a cognitive critical faculty which passes judgment on men's motives and intentions. But he takes the vital step of making it also an authoritative practical power, or motive, determining action as well as passing judgment upon it. It is the highest element in our nature. Self-love can control the passions but may itself be inordinate or

untimely; conscience lays down its proper limits. Benevolence might seem (as it did to Hutcheson) to be an adequate guide to moral action; but on Butler's view even benevolence is not the whole of virtue; conscience again has the last word on what a man ought to do. Butler points out that if benevolence were our principle we should not have duties to bestow benefits on some persons rather than on others, or to condemn falsehood, injustice and unprovoked violence except so far as we saw that they diminish genuine happiness. Butler inclines to the view that conscience is a faculty of immediate and direct awareness of right and wrong, more akin to a sense than to reason. Those who argue that he is a rationalist because he calls conscience a "principle of reflection" have missed the sense of the word "reflection" (explained above in connection with Shaftesbury). "Reflection" now tends to mean "cogitation," but its 18th-century meaning was "introspection."

Richard Price (1723–91), in his *Review of the Principal Questions and Difficulties in Morals* (1757), pointed out that a faculty of direct and immediate perception need not be a "sense"; the senses are notoriously arbitrary, subjective, variable and liable to error. Reason itself can apprehend directly and immediately truths such as the axioms of logic or mathematics, and the apprehension of right and wrong is of this rational intuitive type. Ideas of right and wrong are simple and are not definable in terms of anything else. As in any other field a rational argument must have premisses which do not require demonstration, or there will be an infinite regress. These ultimate premisses in morals are those characteristics of actions (equity, justice, veracity) of which every rational being naturally and necessarily approves.

Hume.—David Hume (1711–76) made his contributions to ethics in his *Treatise of Human Nature*, book iii (1740), and in his *Enquiry Concerning the Principles of Morals* (1751). He drew a clear distinction between two questions which it is important to consider separately: (1) What do we mean by calling anything "good"? and (2) What states of affairs are good? He held that to say anything is good means that most men approve of it. So to find what is good we must ask what most men do approve. The answer is that they approve acts so far as these bring pleasure directly or indirectly to all those affected by them. Hume's definition of "good" is odd for two reasons. In the first place, to decide whether anything is good would require a statistical inquiry into human attitudes, presumably including those of men at all times in history and everywhere in the world; with the result that any actual conclusion that an individual might make would have to be inductive and to rest on the very small sample known to him. Secondly, it would be possible to say "I strongly disapprove of so-and-so, yet I have good reason to believe it is good" (*i.e.*, that most men approve of it). Hume is not wholly consistent here: sometimes he says that for *you* to say anything is a crime, *you* must feel indignation at it; but at other times he quotes the parallel of aesthetics and suggests that to say something is good is to mean (as it does when we say something is beautiful) that it arouses approval in an *intelligent* and *sensitive* beholder (not necessarily in me, and not necessarily in most men).

On the question whether this approval is the exercise of a rational faculty Hume is explicit and consistent. Reason can only reveal truth, facts, connections; and these themselves have no power to move the mind to action or feeling. We say that a man's reason stops him from drinking polluted water. But what reason does is to tell him that polluted water will probably make him ill. Unless he has an aversion from illness, this information remains mere information with no power to move him. Reason may ask why something should be done, but somewhere it must stop and its stopping place must reveal a desire for which no reason can be given. Why do we take exercise? To avoid illness. Why do we shun illness? To escape pain. Why do we shun pain? No reason can be given. We just do. Similarly in morals: no proposition containing a moral term—"good" or "ought" or "right"—can be derived by reason from a proposition not containing a moral term, from a statement of fact or of necessary connection; and again, unless there is to be an infinite regress there must be at least one moral premiss for which no reason can be given. Reason is thus in nonmoral matters the slave of the passions; in moral decisions the slave of our approval feelings.

3. Utilitarianism.—Jeremy Bentham (1748–1832) worked out a complete system of ethics based on egoistic hedonism. Every man pursues his own maximum pleasure and nothing else. The moral standard is the greatest pleasure of the greatest number. Bentham is not clearly aware of the divergence between these two aims, but he appears to try to unite them by his theory of sanctions. How can a man (whose only possible aim is his own greatest pleasure) be persuaded to ask for the general happiness? If he does not, says Bentham, the law and public opinion will descend on him and he will be sorry for it. Or, if he escapes these penalties by not being found out, God will punish him. These solutions are unsatisfactory because they presuppose that the authorities—the legislator, the member of the public, God—have an interest in the general welfare as such, which, at least for the first two, Bentham has implicitly denied. John Stuart Mill (1806–73) attempted to help things out by invoking further sanctions—the internal sanctions of social feeling: I cannot be happy if my children are suffering. But all these sanction arguments presuppose that the moral law (general utility) has in itself no power to influence the will of men. Mill also undermined the simplicity of Bentham's system by taking the view that pleasures differ in quality as well as in quantity; so that a man may often prefer less of a higher to more of a lower pleasure. He did not notice that these terms "higher" and "lower" introduce a nonhedonistic standard of value.

4. Continental European Ethics.—In the 17th and 18th centuries continental ethics had no influence on British moralists. Hence the preceding sketch could deal with the latter in isolation from the former. Moreover the British moralists whom we have considered did not treat ethics as an integral part of a systematic philosophy of the universe. Most of them had no such systematic philosophy, and those who had kept their ethical theories independent of it. The continental philosophers of this period, however, all had metaphysical theories with which their ethical views were closely bound up. The following sketches must therefore do them less than justice in isolating their ethical theories for separate summary; and the reader is referred to the fuller articles dealing with them individually elsewhere in this encyclopaedia.

Spinoza.—Benedictus de Spinoza (1632–77) is above all the metaphysician of ethics. His *Ethica ordine geometrica demonstrata* is a treatise of the whole nature of the universe, yet he recognized the ethical problem as the most important and subordinated to it all other philosophical issues. For Spinoza, philosophy was a completely rational religion, enriched by elements of mysticism. The development of human enlightenment is one with that of human freedom. As reason comes to play an ever greater part in controlling man's life, he approaches nearer and nearer to the human ideal. Each individual has as his essential nature a self-preserving impulse (*conatus*) of which all human feelings and emotions are the expression. The lowest level of knowledge is sense perception and uncritical belief founded on instinct, association and hearsay. Connected with this, on the noncognitive side, are the "passive emotions" which are due to external stimuli or to internal feelings which express blind instincts of self-preservation. These are all efforts of the self to preserve its balance and harmony; but they are uncontrolled and unco-ordinated. This is the condition of human bondage. As soon as a man begins to understand his own nature he begins to be able to control it. This control is correlated therefore with the second stage of intellectual development, that of rational insight into necessary connections. When we understand human nature—our own and other peoples'—we no longer hate and fear impulsively and blindly; we cease to feel regret or resentment and acquire peace of mind. The "active emotions" are those which arise wholly from within a man and express this rational insight into human nature. For example, sensuous pleasure (*titillatio*) is the passive reaction to local stimulation of an organ; joy (*hilaritas*) is the positive vital expression of the whole personality. At the highest stage of intellectual development, that of intuitive knowledge, man con-

templates the whole universe as a unity, identifies himself with the Absolute Being and is filled with the intellectual love of God.

Kant.—Like Spinoza, Immanuel Kant (1724–1804) made ethics an integral part of a complete system of philosophy, a system which nevertheless was all developed with the main purpose of accounting for and making possible a belief in moral values in the face of the materialism of science and the irrationalism of Hume. Kant held that the morality of an act depended entirely on its motive and that the only unconditionally good motive was a sense of duty. Other motives, such as pity or maternal affection, however charming and attractive they might seem, were morally good only if the actions that they prompted were approved by the sense of duty. For without this approval pity may inspire sentimentality and mother love, nepotism. Talents such as intelligence are good only when in the service of a sense of duty. When they serve a bad will they make things worse, not better. Pleasure itself and happiness are worthy of acceptance as the reward of virtue; but pleasure as such is not good. For when a man takes pleasure in malice, his pleasure makes the situation worse than it would otherwise be. A moral action must be done not only in accordance with duty but from duty. Duty expresses itself in "categorical imperatives" such as "thou shalt not steal" as distinct from the hypothetical imperatives associated with desire ("do not steal if you want to be popular") where reason is, as Hume had said, the slave of a "passion." These categorical imperatives are general rules valid for all moral agents and are therefore attributed by Kant to "practical reason." The criterion of the morality of an imperative is the test of universalization: "Act only on such a principle as you can will should be a universal law."

No action done from desire or inclination is moral. It is true that the doing of duty may give pleasure and its neglect pain, but these pleasures and pains cannot be the motives of moral actions since they depend on prior recognition of the rightness of an act as the sole motive for doing it. Happiness is the due reward of virtue. Since this linkage cannot be established by human contrivance or observed in life on earth, Kant regards the moral necessity for it as his moral argument for the existence of God and of immortality.

Kant also links his view of morality with his doctrine of freedom. Duty is the expression of a man's rational nature, completely independent of his desires and inclinations. Man, said Kant, is a denizen of two worlds: the world of the senses and the body, in which all events must be regarded as causally determined through their relationship to other natural phenomena; and the world of mind and reason, in which action is subject to rational moral law. Some interpreters present Kant's view as being one reminiscent of Spinoza's: that when I am driven to act by external stimuli or by my own desires or inclinations I am in the bondage of causal necessity, but that whenever I am motivated by a rational concept of duty I am free; freedom then consists in action determined by moral law. But this interpretation would mean that no wrong action is free and thus that no one can be blamed or feel justifiable remorse. There is some evidence for another interpretation: that I myself decide whether to submit to the bondage of passion or to lay down for myself the rule of duty and that this choice, whichever way it goes, is the basic exercise of freedom. The rule of duty can never be a rule of the kind that necessitates action: rather, like a law of the land (and unlike a physical law), it must be a rule to which a man subjects himself and which he can therefore reject at will. Kant sometimes appears to hold that reason itself in the practical field can motivate action, that simple recognition of duty can stimulate a man to obey the law. But he also refers to a special kind of attitude or emotion which he calls *Achtung* ("reverence") for the law and suggests that it is this "reverence" which impels me to do my duty. Thus after all he seems to come back to Hume's view that reason alone cannot move to action. There must be some emotion or passion (even if it be an emotion "self-wrought by a rational being" and different in kind from all other feelings) which supplies the motive power.

Hegel.—The ethical doctrine of G. W. F. Hegel (1770–1831), expounded chiefly in his *Grundlinien der Philosophie des Rechts* (1821), shows a close affinity with Kant's and yet also a striking contrast to it. Kant had left two difficult questions unanswered. His test of universalization was a criterion by which to judge whether any proposed principle of action was moral; it did not explain whence these principles were derived and how they were related to each other; it was not a premiss from which particular moral principles could be deduced. Hence Hegel calls the principle of universalization empty and negative. Moreover, by making morality a matter of motive, Kant failed to face the difficulty that the sense of duty in different people may point in divergent directions. The patriot and the conscientious objector are both prepared to will that the maxims of their actions should become universal law. Both act morally and from "reverence" for duty. Hegel agreed with Kant that it is not enough that an action should be in accordance with duty; the subjective motive must be that of duty; and many moralists would agree with Kant in asking no more than this. Hegel, however, points out that the individual's moral judgment claims to be objectively valid for all rational agents and therefore must be judged not only by its motive and its form (*i.e.*; whether the agent would be willing to universalize it) but also by its content, by what in particular he holds it right to do. The content of particular moral judgments, according to Hegel, comes from the laws, institutions and customs of the community of which the agent is a member. The only adequate moral standard is harmony with the objective social relations (kinship, economic, political) in which the individual finds himself placed. It is only when this social setting is itself collapsing (as Greek democracy was at the time of Socrates) that an individual can set up his own private judgment and standards in rivalry to the demands of his objective social environment and its law and tradition. Even if he has to do this he cannot rest in a simple act of rebellion or feeling that he must go his own way; he must work out his dissent (as Plato did) into an objective criticism and reconstruction, and this will mean the creation of a new ideal community. Yet this new ideal state will be seen (by later eyes though probably not by its creator) to be in all essentials the same as the actual state within which the critic lives. For no one can transcend his own historical environment. Except for such extreme cases, Hegel explicitly condemns as morally evil the man who sets the judgment of his own conscience against the demands made on him by the law and traditions of his society.

Hegel finds the highest expression of rational objective morality in the state. Ideal morality is the union of the subjective (conscience) with the objective (law and tradition), when a man from a sense of duty upholds and conforms to rational political institutions and laws.

5. The British Idealists.—We may here notice (though in date they belong to the next section) the school of philosophers in Great Britain who broke with the English tradition of insularity and antimetaphysical empiricism, through their susceptibility to the influences of Kant and Hegel. The two most influential of these were T. H. Green (1836–82) and F. H. Bradley (1846–1924). Green in his *Prolegomena to Ethics* (1883) defended ethics against attempts such as those of the evolutionists (*see* below, under the heading *Nietzsche*) to base morality on natural science. He held on metaphysical grounds that the self can never become an object of scientific understanding: it has a timeless and unitary nature which scientific analysis must fail to grasp. A desire in man is not (as it is in an animal) a separately operating impulse moving the organism to action: it is an expression of the whole self, consciously felt and accepted as such. "I move in the desire; it does not move me." My aim is always satisfaction of the whole self, not satisfaction of the isolated desire. In applying this theory of satisfaction of the self to moral conduct, Green ran into great difficulties, and it is impossible to reduce his views to a single coherent answer. He saw morality as the direction of the will to a common or social good, yet his common-sense English realism rejected the ideas of a common or corporate self, whose good this is. He spoke of my good as a good in which the welfare of others is "included," as a good that all "share," as a good in which "the distinction between self and others does not arise." Yet he per-

sists in maintaining that the good for any man is the ideal of a perfect self-conscious life for himself and that others come in only "so far as he finds the thought of their being perfect necessary to his own satisfaction." The difficulty is increased because he admits that the main element in such a conscious perfect life is virtue itself; yet virtue is the one kind of well-being that I cannot bring about in anyone except myself. In fact Green's metaphysical views, which gave primary importance to the individual self, were at continual war with his moral and social consciousness, which drove him to value most highly work of public welfare and utility. Green saw progress in moral ideals as the continued expansion of the limits of the group of men "to whose good I recognize that I should contribute," an expansion which ends only when the Christian duty to all mankind is accepted.

Bradley also pursued the ideal of self-realization under the same influences, though Hegel is the stronger influence in his case (whereas Kant is in Green's). Bradley's *Ethical Studies* (1876) is a difficult and inconclusive work. He appears to accept the Hegelian moral standard of "my station and its duties" and the extension of the self to include other selves in the corporate self of a society; yet his book ends with doubts about this. He points out that Green's "expansion" theory ending with the "great society" of all mankind is self-defeating, for "mankind" is not the name of a society or association. He points out that there are values or types of perfect self-conscious experience—truth and beauty—which are in no sense social. The ideal self has thus social and nonsocial aspects and no single intelligible principle can be found which will unite them. Morality is thus essentially incomplete and unintelligible by itself. He hints that no rational argument can complete it and that it can only remedy its own internal stresses by leading on to religion or to some form of mystical experience.

B. Later Theories

Ethical theory after the middle decades of the 19th century, both in Great Britain and on the continent of Europe and also in the United States, broke up into a number of conflicting schools. Only one of these preserves a link with the historic development of ethical theories. The others are new developments with no roots in the past.

1. Nietzsche.—The first of these new developments was a direct result of the Darwinian theory of evolution. In England Herbert Spencer (1820–1903) and Leslie Stephen (1832–1904) attempted to apply the evolutionary principle to the understanding of morals. But Spencer never succeeded in disentangling himself from utilitarianism and from a belief (inconsistent with all his other beliefs) in justice as a self-evident moral principle. Stephen, a more consistent evolutionist, maintained that society and not the individual is the unit whose health and survival justify moral conduct; but he never found any adequate criterion of social health, nor could he explain how a factual inquiry into the origins of moral conduct could explain moral obligation. The one thinker who seemed to carry evolutionary ideas to their logical conclusion in morals was Friedrich Nietzsche (1844–1900). Nietzsche accepted frankly the glorification of brute strength, superior cunning and the other qualities conducing to success in the struggle for existence, to which the ethics of evolution lead; though, as he himself realized, his theories were more an expression of his own lonely, thwarted, bitter, repressed personality, of his vanity and misanthropy, of his failure as soldier, as lover and as human being, than a cool deduction from biological principles. Indeed he was suspicious of Darwinian evolution as a field of chance in which the fittest did not always survive. Yet his philosophy owed much of its contemporary appeal to its apparent consonance with scientific belief, as well as to the brilliant if fevered style in which it was presented. Nietzsche accepted to the full the pessimism of his predecessor Arthur Schopenhauer (1788–1860) concerning human beings and the world around him. But instead of plunging him, as it plunged Schopenhauer, into final passivity and gloom, the spectacle of the weaknesses of humanity and of the decadence of western civilization roused Nietzsche to a Dionysiac passion expressing itself in wild laughter and poetic frenzy. Will and not reason, power and the will to power are the keys to life. War is the great purifier and ennobler, but men are so drugged that they worship peace. Weakened by the Christian virtues—the "slave virtues"—of humility, chastity, selflessness and obedience, dragooned by the "herd virtues" of industry, moderation, regularity and acceptance of one's station in society, the will of man is hamstrung. Man as he is can never rise above these handicaps. The future lies with a higher type, the superman, who shall have shed the inhibitions clamped on man by current morality and by his own cowardice. Cruelty, courage and stoical endurance are the virtues of the future. Civilization is justified only as a stage for the liberation of a few great spirits, men like Cesare Borgia and Napoleon. Aristocracy is the only tolerable social pattern. Monarchy and democracy alike clamp the social bonds too tightly. Aristocracy requires the existence of a slave class among whom the "slave" and "herd" virtues will rightly prevail. A society so organized will have great driving force and can be looked on as a kind of superman itself. It will go on its way by conquest, cruelty and terror, despising yet profiting by the weaknesses inevitable in other societies dominated and sapped by Christian morality.

2. The Developments of Utilitarianism and Its Critics.—The moralists mentioned above as standing in a line of continuous development from previous ethical theories are next to be considered. Henry Sidgwick (1838–1900) was a patient and careful thinker whose fairness and balance place his work (*Methods of Ethics*, 1874) high among 19th-century moralists. He held that "right" and "ought," used as moral terms, are unanalyzable and that an action recognized to be right is always so recognized because it belongs to a class or kind of actions. When "ought" or "right" is used of a particular action on a particular occasion, there is thus always implied in its use a rule or principle, or a general characteristic of the particular act which makes it right and accounts for the obligation to do it. Common sense seems to recognize many such general features making particular acts right (veracity, justice, return of services rendered, etc.). Sidgwick, however, holds that none of these features provides a final justification of a particular act. Only the characteristic of producing happiness is adequate to do this. He does add some general principles concerning the distribution of happiness, but these can be reduced to the principle that the happiness of all those affected by the act must be impartially taken into account; *i.e.*, to the principle of general utility, as contrasted with the principle of egoistic utility (which would make the agent's own happiness the ultimate criterion). Yet Sidgwick also thought that the agent was properly concerned with the quality of his own experiences in a way in which he was not concerned with the experiences of anyone else; and this led him toward egoistic utility. He attempts to reconcile the divergence between these two conclusions by the method of sanctions adopted by his utilitarian predecessors. But he sees clearly that Mill's sanctions of sympathy and social feeling are inadequate for this purpose. Only the hypothesis of a God (or some universal principle in the nature of things) ensuring that an individual will always achieve the maximum happiness for himself if he promotes the maximum general happiness will be adequate. But this is equivalent to accepting egoistic utility as the ultimate criterion for right and wrong, for it implies that in the absence of any such divine intervention the individual would be justified in pursuing his own maximum happiness. The moral rules of common sense have been found to be the rules most likely to bring about happiness; and it is therefore advisable for the ordinary man to act on them without attempting to calculate in every case the happiness consequences of his actions. But a moral philosopher must ask whether these rules are ultimate and, if not, how they are justified; and he can find no adequate justification for applying one of these rules other than the production of happiness.

G. E. Moore (1873–1958), in his *Principia Ethica* (1903) and *Ethics* (1912), further develops the utilitarian position. He holds that "good" is an indefinable quality and attacks the fallacy of identifying it with any natural quality such as "pleasant." This identification would make the proposition "pleasure is good" a

tautology or analytic proposition, which it clearly is not. He therefore distinguishes sharply between the question what the word "good" means (which he thinks is unanswerable) and the question what kinds of states of affairs are good. Such states of affairs may be good as means or as ends. States of affairs are shown to be good as means by a causal argument showing that they lead to states of affairs which are good as ends or intrinsically good. There can be no proof that any state of affairs is intrinsically good. This must be recognized by direct inspection or intuition. Moore makes his first break with previous utilitarian theories by holding that there are characteristics other than pleasantness which make a state of affairs intrinsically good. Knowledge, virtue, beauty and personal affection are all intrinsically good. Moore also introduces a principle of "organic unity" to the effect that a state of affairs including several elements may have a total value greater than the sum of the intrinsic values of the component factors. For example, if a man takes pleasure in beauty this enjoyment of beauty has a greater value than the same amount of pleasure and the same degree of beauty would have if they existed unrelated to each other. Moore notices that there are two quite distinct problems in ethics: the problem what states of affairs are good and the problem what actions are right. He is true to the utilitarian tradition in maintaining that the right action is that one among all the alternative actions open to the agent which will bring into being the greatest amount of intrinsic good. He equates the right act with duty and "ought" (except in the special case where there are several alternative acts possible in a particular situation all of which would have equally good effects; for in that case it is correct to say that they are all equally right, but a man's duty is to do any one of them, not all of them). When one act would bring about more good than any alternative open to the agent, that act is right and is the agent's duty and is the act that he ought to do. It is to be noted that the propositions that beauty (or anything else) is good and that the right act is that which brings into being the most good are, in the language of logic, synthetic a priori propositions. They assert necessary connections between a moral characteristic and a nonmoral characteristic in the one case and between two moral characteristics in the other—connections which must be apprehended directly by intuition or rational insight with no possibility of further proof.

Moore makes a further advance on previous utilitarianism by firmly adopting general utilitarianism and rejecting egoistic utilitarianism, rejecting also, therefore, all the unsatisfactory sanctions —arguments by which his predecessors had tried to reconcile these two mutually incompatible positions. He holds clearly that it is irrelevant to whom the good produced by a right action accrues. Egoism's restriction of the good to that of the agent is indefensible; interest in one's own good is a natural but not a moral tendency. But he also rejects the standard formula of general utilitarianism, "the greatest happiness of the greatest number," because the number is irrelevant: if I can produce a greater total amount of good by an action bringing benefit to ten persons than by one bringing benefit to a hundred, the former action is my duty.

Moore draws another important distinction. He points out that many moralists have thought that the motive of an action had something to do with its rightness. But this is because they confused the notion of "right" with that of "morally praiseworthy." Motives concern the latter but not the former. If a man causes much suffering from a motive of benevolence we must say that he deserves praise but that he did the wrong act.

It is clear from the above that Moore, like all utilitarians, holds a thoroughly objective theory of morals. States of affairs are good or bad independently of the question whether men believe them to be so or prefer or desire them. Acts are right or wrong according to their actual consequences and to the existence of the good that they bring into being and are so independently of the motives or intentions of the agent. Moore frequently attacked and refuted subjective theories of "good" or "right" which maintain that to say that something is good or that some act is right is to say that someone has a feeling or volitional attitude toward it, that he approves or prefers or desires it. His main argument against these views is that they imply that when A says "X is good" and B says "X is bad" both propositions may be true and A is not contradicting B; for the propositions are asserting that A desires or approves X and that B dislikes or disapproves it. There is no contradiction here, any more than there was when Jack Sprat could eat no fat and his wife could eat no lean.

Moore's theory of ethics has been called "ideal utilitarianism" or "agathistic utilitarianism" to distinguish it from the hedonistic utilitarianism of his predecessors (with their belief that nothing except pleasure is intrinsically good).

Moore's utilitarian theory was subjected to a line of criticism first launched by H. A. Prichard (1871–1947) in his article "Does Moral Philosophy Rest on a Mistake?" (*Mind*, 1912) and fully worked out by W. D. Ross (1877–) in *The Right and the Good* (1930) and *Foundations of Ethics* (1939). Prichard argued that all previous moralists had made the mistake of asking why I should do my duty (*e.g.*, why I ought to pay a debt or keep a promise or tell the truth) and had inevitably tended to answer "because it will give me pleasure" or "because it will further the general happiness" or "because it will produce the greatest good." But all these answers are mistaken and indeed illogical, because they all imply that an "ought" proposition can be derived from a proposition not including an "ought." On Prichard's view the sense of obligation to do, or of the rightness of, an action of a particular kind is absolutely underivative and immediate. That there are many kinds of action which are obligatory is no more puzzling than that there are (as Moore had found) many kinds of states of affairs which are good. Ethics and moral decisions would be greatly simplified if there were only one kind of thing (pleasure) which was good, or if there were only one feature (productiveness of maximum good) which made acts right. But the facts of moral experience are against both of these simplifications, and it is more important that a theory should be true than that it should be tidy. Ross and Prichard both pointed out that the fact that a certain characteristic (*e.g.*, that it is the telling of truth) makes a particular act right does not mean that truth must always be told. It was Kant's mistake to think that moral rules must be thus absolute and have no exceptions. For there may be another characteristic of the act which is truthtelling which would make it wrong. For instance, if I have been told something in confidence and am later asked about it, telling the truth would be wrong insofar as it was the breaking of a promise. Rules cannot be absolute because any two of them may conflict in a particular situation. Thus each right-making characteristic (veracity, justice, etc.) is a characteristic which tends to make actions right and does in fact do so unless overborne by some stronger obligation arising from the presence of some other and incompatible right-making characteristic. This argument constitutes a refutation of utilitarianism. For the reason why we should pay a debt is that this kind of action is necessarily right as such and not that the payment would bring into the world more good than any alternative action. Indeed it is often right to pay a debt when we could do more good by some other use of our money. Utilitarians usually attempted to explain this difficult case by maintaining that the act of paying the debt is right here because it produces indirect as well as direct effects. It not only benefits my creditor but also helps to uphold the credit system; and if I were to default on my debt in order to help some deserving charity, I should not only damage my creditor but also make it less likely that other people would get credit in future. Ross points out that this indirect effect occurs only if my default becomes known. But I may owe a debt to a man who has forgotten about it or has died leaving no record of it; and in that case default would do no damage to the credit system. Yet I should not believe that under such circumstances all that need be considered is how I can do most good with the money. Nor, where there was any doubt about whether my default would leak out, should we regard this consideration as decisive. Ross objects in particular that the utilitarians make two mistakes. In the first place, they regard the relevant factors concerning any duty as factors wholly in the present or the future (the good brought into existence by the act either in itself or in its consequences); whereas some obligations are certainly due to a past act (*e.g.*, those of debt paying and promise keeping). Sec-

ondly, utilitarians consider other people only as possible beneficiaries of an action. Provided that I can do more good to A than to B, this is my duty regardless of who A and B are. "Everybody is to count for one and nobody for more than one," as Bentham said, and we have seen that Sidgwick and Moore follow him here. But Ross points out that we regard particular people (parents, children, colleagues, friends) as having a special claim on us; so that it may often be my duty to do less good (to A, my parent) than I might otherwise do (to B, who has no "claim" on me).

Ross is much nearer to Kant than to any other previous philosopher. We have seen one point, however, in which he diverges from Kant (when he refuses to make moral rules absolute). Kant had also maintained that utility must be wholly excluded from the concept of duty, whereas Ross allows to the utilitarians that productiveness of good is one of the many characteristics which tend to make acts right. If I have no special claims on me (debts to pay, promises to keep, parents to support, etc.) my duty will be to produce as much good as possible; but when one of these other special claims arises (*e.g.*, in the debt case above) I may have a duty (of debt paying) which outweighs the duty of producing the greatest good. There are, however, cases where the good done by not paying a debt (or telling a lie or breaking a promise) is so great that the duty of producing good outweighs the other special claim. It is again to be noted that for Ross and Prichard the proposition that the characteristic of being a payment of a debt (or the telling of the truth, or the keeping of a promise) tends to make an action right is a synthetic a priori proposition (compared by Ross himself to the self-evident and necessary propositions of arithmetic).

Ross admits that people differ widely in the kinds of acts which they think right and sees this as an objection to his theory that the fundamental propositions of ethics are necessary self-evident a priori truths. He explains this by saying that the moral rules which vary from people to people are not themselves basic or ultimate rules but are *media axiomata* resulting from the application (to particular circumstances or with limited knowledge of facts) of basic rules over which there is no divergence. For example, it would seem odd to say that our rule that we should honour and cherish our parents is a self-evident necessary proposition when we find that some savages believe it right to kill and eat their parents. But the savages really agree with us on the basic rule that parents should be kindly treated; it is only that in nomadic conditions and on the verge of subsistence the kindest thing is to kill them. As for eating them, the savage holds the old-fashioned dietetic theory that a man comes to resemble what he eats. Eating his parents is his way of trying to acquire their virtues. He agrees with us on the basic rule of honouring one's parents too.

Ross and Prichard hold (equally with Moore and Sidgwick) objective theories of "right" and "good." The rightness of an act and the goodness of a state of affairs are wholly independent of the beliefs, desires, feelings, emotions or preferences of the agent or the critic passing judgment on them and of the motives and intentions from which an act is done. But Prichard saw that paradoxical consequences arise from making "right" and "duty" objective terms. Ross had held that in any set of circumstances there is one act which is right and is my duty and what I ought to do, whether or not I am aware of it. For example, a man asks me the way to the station, which in fact leads north; but I have been misinformed and I believe that the way runs east. In this case my duty would be to tell him that the way runs north, and this is what I ought to do. Or again, if I am a surgeon and believe that an operation will cure a patient, when in fact, because of a heart condition unknown to me and undiscoverable by me, it will kill him, I shall fail in my duty and do what I ought not to do if I operate. It is clear that if in such circumstances I do my duty I shall do it by luck and not by intention and from an immoral or at best nonmoral motive. I shall never know what my duty is, nor whether I have a duty; nor shall I ever be able to do a duty because it is a duty. I shall frequently deserve the highest praise for failing to do my duty and the sharpest condemnation for doing it (since duty has nothing to do with my motives, beliefs or intentions, and praise and blame have to do with nothing but these). Hence Prichard was able to convert Ross to the view that my duty depends on what I believe about the situation and not upon the situation as it actually is. Moreover, Ross had held that my duty was to bring about the state of affairs which would satisfy the most stringent obligation upon me. Prichard argues that the production of such states of affairs may be beyond my knowledge or ability. "Ought" implies "can," and the most that I can do is to "set myself," or make an effort, or try, to produce the state of affairs which I believe fitting to the circumstances, as I believe them to be. If, through no fault of mine, the state of affairs does not come about, I may still claim to have done my duty and done what I ought.

3. Ethical Relativity.—The great variety in judgments concerning what is good or right has always tended to lead some moralists to the view that these judgments are not beliefs about objective facts, but that the terms "right" and "good" are relative terms and describe or express the desires or feelings or preferences of the speaker toward actions or states of affairs contemplated by him. The rise of scientific sociology and increasing knowledge of the behaviour of primitive peoples has heightened the general awareness of the variety of moral experience; and at the same time the breakdown of the rigid principles of the 19th-century morality has brought about the existence of conflicting moral judgments within most developed communities and has further shaken the belief that any moral principles can be self-evident necessary propositions as final and certain as the truths of mathematics.

Sociologists have observed how taboos have been inculcated and transmitted from generation to generation and how they have frequently been invested with religious sanctions which have tended to make them accepted as eternal and revealed truths. As a result some moralists, for example W. G. Sumner (1840–1910) in the United States, have held that "the right action" simply means the action which the custom of the society to which a man belongs enjoins upon him. The connection of "moral" with mores (Lat. "customs") reinforces this view that the mores can make anything right. In England E. A. Westermarck (1862–1939) in his books *The Origin and Development of the Moral Ideas* (1906) and *Ethical Relativity* (1932) works out in great detail a view of moral concepts which makes them relative both to the individual who uses them and to the society in which they operate. Words such as "right" and "good" express a specific emotion, that of disinterested and impartial approval of certain actions; "bad" and "wrong" express an equally disinterested and impartial indignation. These emotions are not natural or innate or instinctive; they are implanted in us by social pressures. Society teaches us to attach these moral terms to acts which are in the interests (or against the interests) of the society itself. A nomadic society will instill the approval of kinds of actions which assist its survival in the nomadic state and will foster indignation against acts that threaten this survival. The general theory that moral judgments are expressions of emotions and do not state beliefs about objective facts is dealt with later in this article. Here some comment may be made on the sociological aspects of the theory. The view that the mores can make anything right, that a man's moral judgments are determined by the teaching of the society in which he lives, is faced with the difficult if exceptional cases of the rebels, the heretics, the reformers, who reject the standards of the society in which they have been brought up. When the rules of a society are accounted for by reference to their survival value for that society this is really a different theory. But it may be invoked to explain at any rate some rebel opinions. The rebel may say "I know my society has done its best to instill in me moral feelings toward certain actions, as being appropriate to its needs. But I am convinced that these actions are inappropriate for my society in its present environment and will lead to its weakening or destruction. Accordingly I reject them." Sociological theories of ethics have indeed a general tendency to confuse these two views: the view that I think right whatever my society approves; and the view that I think right whatever benefits my society. Westermarck admits the difficulty raised by these nonconforming

consciences, and he also draws a distinction between enlightened and unenlightened moral judgments, a distinction impossible on the simple view that "right" means whatever my society approves. According to him, the man who first criticizes the view held by his society that a man should be punished for unintentional acts or that lunatics should be punished has appreciated more clearly than his fellows that indignation is felt not against acts but against agents and sees more clearly than they what "agency" involves. Or again, the man who first in a slave-owning society pleads for emancipation of slaves is widening the range of his impartiality. But these two types of explanation will not explain all moral criticism of current beliefs. And the latter involves a widening of the whole of "society" until it includes all mankind (as with the Stoics or the Christians), and then the meaning of the word "society" has altered and the relativity tends to disappear.

4. Pragmatism.—The founders of pragmatism, C. S. Peirce (1839–1914) and William James (1842–1910), were not primarily interested in ethics, and the main contribution to moral theory from within the movement that they founded was made by John Dewey (1859–1952). Pragmatism, or (as Dewey preferred to call his own theory) instrumentalism, holds that thought is an instrument of practice, a means of molding one's environment so as to satisfy one's needs and desires. If thought is aimed at such satisfaction, its truth will lie not in its correspondence with fact but in its success in surmounting the particular obstacle that it is invoked to meet. This account obtains of moral reflection as well as of thought in other fields. Moral problems arise out of a conflict of impulses or desires, and the aim of moral deliberation is to find a course of action which will bring this conflict to harmony. For Dewey, a tentative value is conferred on any object through being desired, but the typical moral problem arises when values conflict (*i.e.*, when the agent finds himself drawn in different directions by contrary desires). How is he to go about to appraise the competing values? Dewey answers that he must place each thing desired in its causal context; he must view it in the light of the means necessary to reach it and the consequences that would follow from it; he will then commonly find that his desire for it has been modified and its values accordingly transformed. Whether or not anything does have value for an agent can be verified empirically, Dewey holds; for the desire on which value depends is a bodily process of active seeking and can be observed like other physical processes; he described himself as a naturalist in ethics and, in his later years, was inclined to accept a behaviourist psychology. A choice is right if it leads to a solution of the conflict that gave rise to it; but there is no absolute right or good, since every successful solution gives rise to further conflicts which must be dealt with in the same way. The test of right conduct is the harmonious satisfaction, not of my own desires merely, but of those of all who may be affected by my conduct and, ultimately, of the human race in the long run. Dewey steadfastly refused to recognize any *summum bonum;* for, insisting that human nature was in course of evolutionary change, he held that any such specification of an end would tend to fix conduct prematurely in certain molds and thus to halt moral advance. He came most near to defining a *summum bonum* when he said that the end was continued growth or progress, this consisting apparently in an ever-widening variety and harmony in the satisfaction of desire. Similarly there are for Dewey no self-evident or universally valid rules of conduct; moral rules are hypotheses which have been found to work in many cases and hence offer helpful suggestions; they offer nothing more.

Dewey's influence, little felt in Europe, was very great in the United States, particularly in the field of education. The essentials of his final ethical theory are best stated in *The Quest for Certainty* (1929), ch. x, and in *The Theory of Valuation* (1939).

5. Existentialism.—The prime factor in the rise of existentialism was a revolt against the system of Hegel. The existentialists saw in such a system of pure thought an escape from reality, a magnificent palace in which men stifle and ultimately a mausoleum to immure the dead. We must escape thence into the street with its chaotic bustling inconsequent individual life. Hegelianism leads to a romantic optimism (the real is the rational) and to a conservative state worship that provides in the nation or the race a refuge to which the individual can flee and in which he can lose himself and shed the burden of his responsibility.

A second influence was the reaction against science. Henri Bergson (1859–1941) had argued that scientific method achieves its results by a process of abstraction, an addiction to laws and formulas which must fail to grasp that experience which is met with in the process of living, which is freedom. Worship of science leads to a liberal utopian optimism that regards men as material and believes in a progress to be achieved by blueprints and by planning.

Both of these tendencies treat individuals as cases or instances of laws and never as individuals; they traffic in "essences." "Existence" is met only in the individual; and the individual exists and creates himself only in free choice. This free choice is indubitably apparent only in "extreme" or "frontier" or "limit" situations, for which no rules prescribe. The creative choice is not the casual decision after which all goes on much as before, nor the choice "on principle" which subjects the self to an external rule or value and reduces it to an identity with other similarly subjected selves. The choice must be a critical one which commits the whole self and the whole life (like that of a wife or a career—or a death). A man must feel "care" (*Sorge*) and the load of responsibility in facing his own limitations and especially the ultimate limitation, death. The fact that he cannot escape death or fall back on other men or on rules or comfortable systems fills him with dread. The recognition of the insignificance of the natural world and of all human values overwhelms him with nausea. He has to commit himself to the deep waters, to the unknown; and the decision is anguish and despair.

Morality thus becomes something wholly beyond reason. There are no absolute values or principles, unless one is to count as the sole value the virtue of "authenticity" or "fidelity" or "sincerity." Such a theory is peculiarly appropriate to a time when civilization is breaking up and social bonds have vanished. It naturally expresses itself in myth and in metaphor and most effectively in literature rather than in argument, for it is in the novel and the play that abstractions melt into concrete living experience. It was the appropriate philosophy for the European "resistance" movements during World War II. The resistance worker had to reject all normal ethical bonds, to colleagues and relatives, to church and state. Moreover, he could not be like most martyrs, the voice of a cause, for he could not say what he was fighting for, except liberty, and a liberty which was made in his choice. The choice was a risk and a wager and his life was always the stake. Rules were useless, for the resistance situation was unique and unprecedented.

Some existentialists rest their position on a religious basis. At these crises of existence, man touches the infinite and the transcendent; but a transcendent which can never be formulated, verified, explained or proved. There is a gulf between man and God which no theory can bridge. Sören Kierkegaard (1813–55) repeatedly quotes Abraham's readiness to sacrifice Isaac as the supreme example of this. Every moral rule and every rational motive was against him; nothing but faith could hold him to his course. Abraham was no tragic hero; the tragic hero embodies a conflict of duties, or a conflict between duty and one of the deep human emotions. In Abraham there was no such conflict; all the rules of morality and the feelings of humanity united to protect Isaac, yet Abraham took him to the altar. The gulf between man and God is bridged for the Christian by the doctrine of the Incarnation. But this doctrine itself—that of the God-man—is rationally an absurdity and its absurdity is the essence of its challenge. Christianity is not a creed but a challenge to decision. The bond to Christ is this experience of personal decisions made and renewed. The response is the response to a person; not to a law or an ideal or a creed, for these have no living force; each is an "essence" or mere possibility unless and until individuals adopt or endorse them. But a person exists independently of our endorsement. Existentialism without such a religious basis faces the world with despair and anguish, a despair which is irremediable and an anguish without reward.

Existentialists differ in their attitude to the problem of human relationships. Some, like Kierkegaard and Jean Paul Sartre (1905–), insist on the loneliness of the human soul and on the need to preserve this loneliness from the encroaching, cramping, restricting invasions of other people. Kierkegaard's ferocious attack on his own Protestant church is reminiscent of Pascal's attacks on the Jesuits. Other existentialists, like Karl Jaspers (1883–) and Gabriel Marcel (1889–), see the final aim of life as "communication" between man and man (and not only between man and God, which is all that Kierkegaard will allow). But "communication" does not mean the use of the counters of language or the joint acceptance of some common goal or rule or the membership of societies or associations. For all these again tend to eliminate individuality. I must retain my uniqueness and my freedom and meet others on this basis. Communication becomes "loving conflict."

Existentialism had its main following on the continent of Europe. Its influence in philosophical circles in Great Britain and the United States was slight, though it had a following among writers and critics who were impressed by the novels and plays of Sartre, Simone de Beauvoir (1908–) and Albert Camus (1913–1960). In theology its links with such antirationalist and "personalist" theologians as Karl Barth (1886–), Martin Buber (1878–1965) and Emil Brunner (1889–) gave it a world-wide appeal. In the United States these tendencies, exemplified in the work of Paul Tillich (1886–) and, above all, in that of Reinhold Niebuhr (1892–), had a vast influence.

6. Logical Positivism and Analysis.—The philosophical movement known as logical positivism was the starting point for another new development, which linked up with the view of Westermarck, noted above, that moral judgments do not state facts about the world but express the emotional attitudes of those who utter them. Logical positivism began as a reaction against metaphysics; *i.e.*, against attempts to discover truths about the world by rational argument without any appeal to empirical evidence. It adopted Hume's principle that all meaningful statements are either analytical tautologies or synthetic propositions resting on empirical observation. This principle makes the propositions of metaphysics meaningless; and it followed that the propositions of theology and morals are meaningless also. We have noted above how the traditional theories had led Sidgwick and Moore and Ross to the view that the basic propositions of morals (that pleasure is good, that it is right to do what will produce most good, that truth ought to be told, etc.) are synthetic necessary propositions apprehended by intellectual intuition. The basic dogma of logical positivism required the rejection of such a priori synthetic propositions.

Most logical positivists, however, had accepted these conclusions concerning ethics and theology as a corollary of their basic premiss and did not base them on any examination of moral assertions or moral arguments on their merits, nor were they much exercised to give a positive theory of what moral judgments were if they were not meaningful propositions.

There were, however, good reasons other than the positivistic definition of "meaning" for doubting whether moral propositions state facts about the world and particularly facts involving necessary connections between independently recognizable characteristics such as "pleasant" and "good" or "truthful" and "right." For these principles are not universally recognized and the difference does not lie merely in their application. Nor can the difference be accounted for by holding that some people are not intellectually mature enough to recognize them (as a savage who cannot count beyond five cannot see the necessity that three threes must be nine). For the problem is not that one man sees duties which another cannot see but that two men see different duties. When this occurs there seems to be no criterion for discovering who is right. Argument can arise about the facts to which the moral principles are applied and can be settled by empirical examination of the facts. But when we have got it clear that we differ not on facts or application but on principle, we often find ourselves having to say "Well, if that is how you feel about it, there is no more to be said." Moreover, if recognition of a moral fact is an intellectual recognition of how certain characteristics are connected, it has to be explained how this intellectual recognition can lead to approval of a state of affairs recognized to be good, or to an obligation to perform an action recognized to be right. These connections cannot be empirical. It cannot be that I just happen to feel approval of what I think good or that I happen to have an obligation to do the act which if I did it I see to be right; or that it is logically possible to see that something is good and not approve it or to see that an act would be right and be under no obligation to do it. Further, on a theory of the type of Moore's or Ross's, these propositions cannot be analytic. To say that something is good is not the same as saying that I approve of it. To say that it is right to pay a debt is not a statement about my feelings at all. Here then are still more a priori synthetic propositions, and certainly this multiplication of such propositions is a cardinal difficulty in the traditional theories.

Accordingly, when recent critics of Moore began to consider moral judgments on their merits, they naturally began with the elimination of the last set of a priori synthetic propositions, by holding that "this is good" and "I approve of this" are identical in meaning and that "I approve of this because it is good" and "I ought to do this because I see that it is right" are indeed necessary a priori truths but are analytic, not synthetic. Moore had objected to any theory which makes "this is good" equivalent to "I approve of this" that on such a view when X says "this is good" and Y says "this is bad" they are not contradicting each other. Their two statements are perfectly compatible, as the first conveys the approval of X and the second the disapproval of Y. Yet most men do believe that such judgments contradict each other: when one says "Fighting is right," another says "*No*, it is wrong"; whereas when one man says "I like caviare," the other says, not "*No*, I do not," but simply "I do not."

The first full-length and careful attempt, by any of these critics, to analyze moral language on its merits is found in C. L. Stevenson's *Ethics and Language* (1945). Stevenson holds that the meaning of a moral word does include a reference to the speaker's attitude of approval or disapproval. But he meets Moore's objection by insisting that a moral word does more than communicate the speaker's attitude: it attempts to affect the attitude of the hearer; it has a persuasive or quasi-imperative force. "This is good" means "I approve of this, do thou approve of it also." This explains the apparent logical contradiction between "This is good" and "This is bad." For if A says to me "This is good" he is trying to persuade me to approve of it (as well as communicating his own approval); and when B says "No, it is bad" he is trying to make me disapprove of it. The word "no" becomes appropriate here just as it is if one man says "Shut the door" and the other says "No, leave it open." It is logically possible for conflicting commands or persuasions to be employed, though not for them to be successful. Thus a primary aim of the use of moral words is to influence conduct.

The majority of thinkers writing on ethics in Great Britain and the United States after the beginning of the 20th century would refuse to be called logical positivists. But they share with the positivists the belief that philosophical problems can be solved by considering how language is used; and they agree also that moral judgments are not propositions expressing belief or knowledge about matters of fact involving actions or states of affairs, but that they communicate psychological conditions or attitudes of feeling or will (or both) and not cognitive attitudes. Much of their ethical thinking is therefore an attempt to state clearly what these noncognitive attitudes are and how they resemble and differ from such cognate states or attitudes as liking, desiring, commanding. The emphasis is on the special purposes for which moral words are used and the special ways in which moral language is learned and taught. This interest in the special uses of language in different fields has led to another recent development. In earlier times the growth of logic had been arrested more than once by setting a standard pattern for rational argument and rejecting as irrational whatever could not be forced into the pattern. Thus all deduction had been assimilated to syllogism and

induction to deduction. So more recently Hume's influence was held to have dictated the dilemma "either scientific or irrational." But moral experience, like other nonscientific fields, may have its special type of rational discourse with its own standards of reason related to the structure of its language and subject matter.

BIBLIOGRAPHY.—H. Sidgwick, *Outlines of the History of Ethics,* 5th ed. (1902); L. Robin, *La Morale antique,* 2nd ed. (1947); E. Schwartz, *Ethik der Griechen* (1951); M. de Wulf, *History of Medieval Philosophy,* Eng. trans. by E. C. Messenger of 6th ed., 3 vol. (1952 *et seq.*); F. C. Copleston, *History of Philosophy,* vol. i–iii (1942–53); Odon Lottin, *Psychologie et morale aux XII^e et XIII^e siècles,* 6 vol. (1941–54); D. E. Sharp, *Franciscan Philosophy at Oxford in the Thirteenth Century* (1930); G. de Lagarde, *L'Individualisme ockhamiste: la morale et le droit* (1946), being vol. vi of *La Naissance de l'esprit laïque;* I. Husik, *History of Mediaeval Jewish Philosophy* (1946); J. Martineau, *Types of Ethical Theory,* 2 vol. (1891); C. D. Broad, *Five Types of Ethical Theory* (1930); T. E. Hill, *Contemporary Ethical Theories* (1950). *See* also the bibliographies attached to the articles on the philosophers and philosophical schools and movements mentioned in the text above. (J. D. M.)

ETHIOPIA (ABYSSINIA; Amharic YAITYOPYA NIGUSA NAGAST MANGUISTI), an independent empire of east Africa. The official name Ethiopia is developed from the tradition which asserts that Ethiops, the son of the biblical Cush, was the father of Aksumawi. The sons of Aksumawi are known as "the Fathers of Aksum," the ancient Sabaean empire established in the north of the country (*see* AKSUM). The terms Abyssinia and Abyssinians, in use outside the country, are derived from the Habastat, one of the two south Arabian tribal groups forming the Aksumite kingdom.

Ethiopia lies between 3½° and 18° N. latitude and 33° and 48° E. longitude. Before the federation between Ethiopia and Eritrea (*q.v.*) on Sept. 15, 1952, Ethiopia was an inland country cut off from the Red sea and the Indian ocean by the territories administered by Italy, France and the United Kingdom. The main outlet to the Red sea was through the French Somaliland port, Djibouti. Since federation, the 600 mi. coast line of Eritrea with the ports of Massawa and Assab has been incorporated in the Ethiopian empire.

The empire is bordered on the north and west by the Republic of the Sudan, on the south by Kenya, on the southeast by the Somali Republic and on the east by French Somaliland. It is roughly triangular in shape with its apex to the north and measures about 900 mi. from north to south and a similar distance from east to west, with its greatest width along latitude 8° N. Its total area (including Eritrea 47,876 sq.mi.) is 457,142 sq.mi. The boundaries in Ethiopia enclose almost all the high mountain country of this part of east Africa.

This article contains the following sections and subsections:

I. Physical Geography
  1. Geology
  2. Relief and Drainage
  3. Climate
  4. Vegetation
  5. Animal Life
II. The People
  1. Ethnology
  2. Languages
  3. Religion
  4. Customs
III. History
  1. Kingdom of Aksum
  2. Zagwe (Zagwe) and Solomonid Dynasties
  3. The Portuguese Interlude
  4. The Galla Invasion
  5. Roman Catholic Missions
  6. Theodore II (1855–68)
  7. John IV (1872–89)
  8. Menelik II (1889–1913)
  9. Lij Yasu (1913–16) and Zauditu
  10. Accession of Haile Selassie
  11. Italian Occupation
  12. Restoration of Haile Selassie
IV. Population
V. Administration and Social Conditions
  1. Constitution
  2. Local Government
  3. Living Conditions
  4. Health
  5. Education
  6. Justice
  7. Defense
VI. The Economy
  1. Agriculture and Livestock
  2. Minerals
  3. Industry and Power
  4. Trade and Finance
  5. Transport and Communications

I. PHYSICAL GEOGRAPHY

1. Geology.—Ethiopia forms part of the structural unit known as the Horn of Africa. The formations represented are:

Pre-Cambrian.—The underlying Basement complex is exposed in Eritrea, along the Sudan border, in Sidamo-Borana and in Harar province. It consists predominantly of metamorphic rocks, gneisses being most prevalent. Many varieties of ortho- and para-schists occur. Igneous intrusions varying from acid to ultrabasic occur with varying degrees of metamorphism. Pegmatites, aplites and dikes are very common. The foliation of the schists is predominantly north-south.

Triassic.—The Adigrat Sandstone unconformably overlays the Pre-Cambrian, attaining a maximum thickness of 2,500 ft. It represents Aeolian and near-shore marine sedimentation during the first Mesozoic transgression, and is of tentative Triassic age in Ogaden and Liassic–Mid-Jurassic age in Tigre and Eritrea.

Jurassic.—The Adigrat Sandstone grades up into fossiliferous limestones with frequent gypsum bands. Except in Harar, where ammonites occur, this group of sediments is undifferentiated as the Antalo Limestone. It ranges from Bathonian to Lower Kimmeridgian. The Jurassic transgression reached its maximum extent during the Kimmeridgian (almost to the Sudan border).

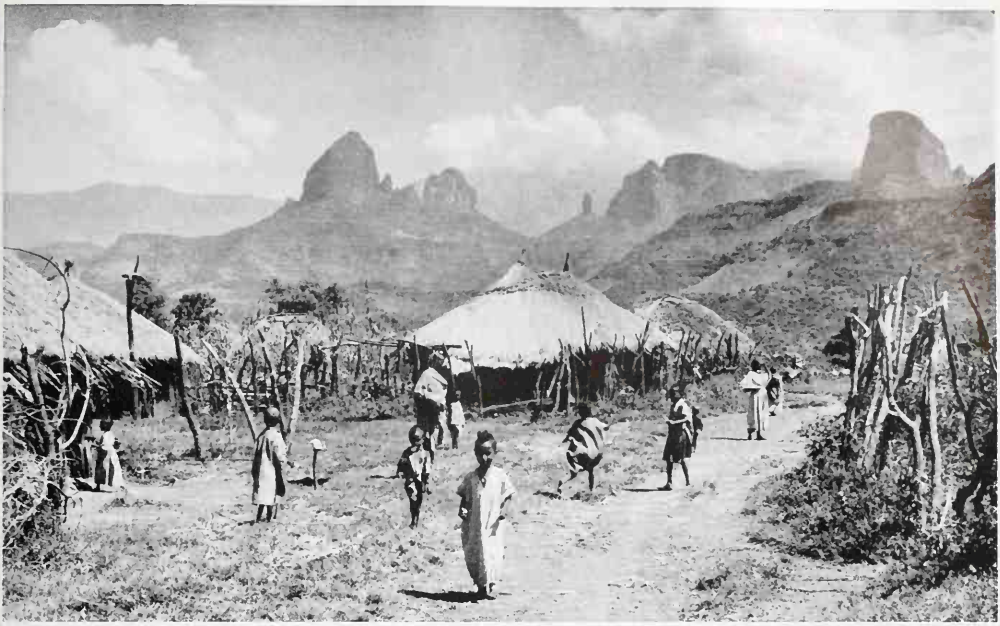
Cretaceous.—The regression of the Jurassic sea is represented by the Upper Sandstone facies of Tigre, joining with the Adigrat Sandstone facies in Eritrea and western Ethiopia, and wedging out the Antalo Limestone. The Upper Sandstone probably includes the Lower Cretaceous, a marine transgression from the southeast reached as far as Dire Dawa, depositing limestone and gypsum.

Tertiary.—Great outpourings of plateau basalts in the Oligocene or Upper Eocene continued throughout the Miocene and possibly Pliocene, forming two series of traps: the Lower Ashangi group, consisting of fine-grained basalt, often amygdaloidal; the Upper Magdala group, consisting of basalts, trachyte and rhyolite. These traps lie horizontal upon the Mesozoic sediments, overlapping them to the north and west, forming the highlands of Ethiopia. Laterite bands and tuffs are interbedded. Isolated volcanoes form the highest mountains. Rift valley faulting commenced in the Oligocene, becoming really active in the Late Tertiary, and affecting the trap series. High horsts overlook the Danakil or Afar depression.

Quaternary.—Mainly in the Great Rift valley, recent volcanics, vaguely termed the Aden Volcanic series, consist of scoriaceous, porphyritic basalts, alkaline rhyolites (*e.g.*, Zuquala, Fantale, Wachacha), with more recent basalts. Dormant volcanoes (Afdera, Dubbi, Erta-Ale) occur in the Danakil desert (DanKalia). Fumaroles and hot springs abound in the Rift valley (a part of the Great Rift valley) and adjacent areas of the plateau. Major Rift valley movements, accompanied by uplifting of the Ethiopian and Somali plateaus, occurred in the Mid-Pleistocene. The seismicity of the region today (an average of seven local earthquakes a month is recorded at Addis Ababa) indicates continued activity.

Extensive lacustrine deposits of three (?) Pluvial ages occur in the Rift valley. There is evidence of glaciation on the Simen (Semien) mountains. Partly marine and partly juvenile salt deposits occur in the Danakil depression.

2. Relief and Drainage.—The country can be divided into a number of clearly defined physical regions. The most important of these are the high plateaus of the western half of the country, which are bounded in the east by the line of the faults, which, from Lake Stefanie (Basso Naebor) in the south to approximately 9° N., have a northeast-southwest trend. On latitude 9° N., the faulting bounding the western edge of the Rift turns due north and eventually closely follows the line of the Red sea coast north of Massawa in Eritrea. At many places, notably at Ankober, Debra Sina and Alamata, this eastern wall of the plateau is precipitous and rises abruptly (about 7,000–8,000 ft.) from the Danakil desert to the east. It presents a formidable barrier to communications.



Village of Adi Arkai on the southwest road between Asmara and Gondar in the Amhara region. In the background is a section of the Simen massif which rises above the Ethiopian plateau to the west

Monk from the Ethiopian Orthodox convent of the Isle of the Sacred Lake (Lake Tana). Christianity was established in Ethiopia in the 4th century

VIEWS OF ETHIOPIA

View of the 18th-century ruins inside the "imperial enclosure," Gondar, former capital of Ethiopia. Buildings shown include (left background) the palace of Bakaffa, 1721–30, and (right background) the House of Song built by David III (Adbar Seged), 1716–21

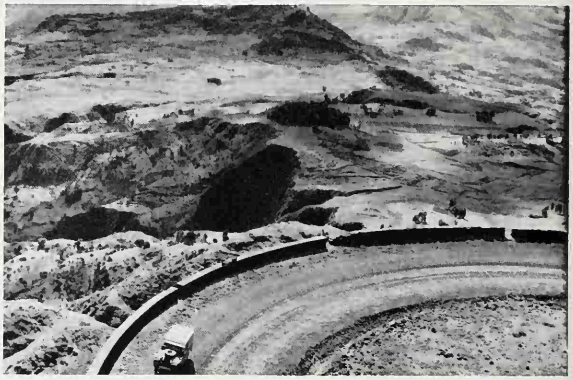
PHOTOGRAPHS, PAUL ALMASY

ETHIOPIA

Young Galla herdsman from Arusi. The Galla are a Hamitic (Cushitic) people who form the majority of the population in southern Ethiopia

Haile Selassie I avenue in Asmara, capital of Eritrea

Monument of the Lion of Judah in Addis Ababa, capital of Ethiopia. The lion symbolizes the nation and appears in the designation of the monarch as "King of the Kings of Ethiopia, Lion of Judah, the Elect of God"



The road northward from Addis Ababa to Asmara provides a means of transportation through the otherwise impassable mountain region of the Ethiopian escarpment

Monks reading an old manuscript written in Geez, the ecclesiastical language of Ethiopia from which developed the three major livng Ethiopic languages

Altar boys dressed in ceremonial robes and carrying umbrellas in a festival procession at Zagé. Ethiopian religious culture reflects a curious blending of Christian, Hebraic, archaic Semitic and pagan customs

SCENES IN ETHIOPIA AND ERITREA

PHOTOGRAPHS, (TOP LEFT, CENTRE RIGHT) CAMERA PRESS—PIX, (TOP RIGHT) WAAG—CAMERA PRESS—PIX FROM PUBLIX, (CENTRE LEFT, BOTTOM LEFT) PAUL ALMASY, (BOTTOM RIGHT) PIX FROM PUBLIX

The road from Addis Ababa to Asmara ascends and descends the escarpment at several points with considerable difficulty, surmounted only by spectacular road building. At only two points, between Massawa and Asmara and between Nazareth (Hadama) and Addis Ababa (where there is a break in the line of the escarpment), has it been possible to construct railways which lead to the plateaus.

The western plateaus slope generally toward the west with some of the highest points (Abuya Myeda 13,123 ft., Amba Sel 11,000 ft., Sarenga 12,000 ft. and Amba Alagi 11,500 ft.) situated along the line of the eastern escarpment. The general slope to the west is, however, interrupted by high mountain groups, which stand above the level of the plateaus and probably represent centres of earlier volcanic activity. The most notable of these is the Simen massif, whose highest point is Ras Dashan (Mt. Bejeda), 15,158 ft. high. Other mountain groups are: the Choke mountains in Gojam province, between Debra Markos and Lake Tana; the Lasta massif in Wallo province which is encircled by the upper course of the Takkaze river and whose highest point is Abuna Josef (13,881 ft.); the Amara-Saint massif south of the upper course of the Takkaze (Takazze), in Wallo province, whose highest point is Amba Farit (13,042 ft.); and in the south, in the latitude of Lake Abaya (Margherita), the Gamu massif with its highest point in Mt. Gughe (13,776 ft.). No peak in Ethiopia is permanently snow covered. The largest area of continuous plateau is to be found in Shoa province to the north and northeast of Addis Ababa and south of the Abbai river.

Since the general slope of the highlands is to the west nearly all the large rivers are tributaries to the Nile system. The most important is the Blue Nile, known in Ethiopia as the Abbai (Abbay), which has its source in Lake Tana, although popular belief locates the source in the Little Abbai river near Mt. Amedamit in the mountains of Gojam province to the south of the lake. This was the source identified in 1770 by the British explorer James Bruce. Lake Tana (*q.v.*) lies in a depression in the plateau and its waters have been dammed, near the town of Bahrdar, by a flow of lava which is crossed by the present course of the Abbai. It is at this point that the spectacular Tis Esat falls on the Abbai occur. After a drop of about 150 ft. at the falls, the river flows in a deeply eroded gorge which bends in a great semicircle around the mountains of Gojam and then westward toward the plains of the Sudan. At the point where the road from Addis Ababa to Dekra Markos crosses the river by means of a concrete bridge, there is a difference in height of about 5,000 ft. between the plateau and the bed of the river. The tributaries of the Abbai drain an enormous area of the central plateaus.

The northern part of the plateau is drained by the Takkaze ("the terrible"—a reference to its formidable appearance during the rainy season) and its tributaries. The Takkaze rises at over 9,000 ft. in the mountains of Lasta, describes a semicircle round the high mountain group of Simen where it flows in a gorge about 3,000 ft. deep and forms a barrier between the northern provinces and the central plateau. It crosses the border near Umm Hajar, after which it has the Arab name Setit. South of the Takkaze-Setit are a number of tributaries of the upper Atbara.

South of the Abbai drainage basin, the highlands are drained by two main river systems. In the west the Akobo joins the Pibor and Baro rivers to form the Sobat. The eastern parts of the southern highlands are drained by the Omo system which is a closed basin terminating in Lake Rudolf (1,230 ft.) on the southern frontier. The trench of the Omo forms a formidable barrier on the eastern side of Kaffa province which assisted in the isolation of this region and the establishment of separate kingdoms before the conquests by Menelik II.

The rivers of the western highlands with their variations in volume and frequent rapids and falls are not suitable for navigation but have possibilities for hydroelectric power. The Baro river in its plain is navigable as far as Gambela during the rainy season when large river boats come from Khartoum. In the east the Awash (Hawash) river rises to the west of Addis Ababa and terminates in Lake Abbé at the end of a series of salt lakes in the desert region west of the head of the Gulf of Tadjoura in French Somaliland. Other rivers include the Webi Shebeli which fails to reach the Indian ocean but the Webi Gestro, the Ganale Dorya and the Dawa unite to form the Juba (*q.v.*) which reaches the sea on the coast of the Somali Republic.

Between the western highlands and the southeastern highlands, in the upland trough of the Rift valley are a number of closed drainage basins feeding lakes: Zwai, the most northerly, Hora Abyata, Abaya, Chamo (Ruspoli), Stefanie, Langana, Shala and Awusa. Lake Rudolf lies in a smaller rift parallel to the main rift of the Ethiopian lakes.

The plateau of Eritrea is structurally separate as it consists mainly of exposed Basement complex and includes the western end of Tigre province south to the Takkaze river. The northwest plateau of Ethiopia, high country much of which is over 8,000 ft., limited on the east by the great escarpment and draining entirely toward the Nile, has Lake Tana at its centre. South of the Abbai and the plateau of Shoa are the southwest highlands, on the average lower in altitude and more maturely dissected. They contain a large number of smaller mountain groups. The southeast highlands border the rift on the east and form a long chain of mountain groups stretching from Harar into Sidamo province. This chain broadens out at latitude 7° N. into the high mountain group of Bale. To the southeast of this chain is the Harar Borana plateau at an average height of 1,000 ft. above sea level. The Rift valley and the Danakil lowlands divide naturally into three regions. The Rift valley, containing the lakes, has the highest elevation (*c.* 5,000 ft.) and is clearly limited by the parallel escarpment to east and west. At approximately 8° N. latitude the valley opens out to the triangular shaped region of the Danakil lowlands or Afar plain. The southern half of the plain (roughly south of the Dase-Assab road) drains toward the Awash river while the northern Afar region contains the structural units of the Danakil horst and the Kobar sink.

3. Climate.—The climate in Ethiopia is controlled by the range of altitudes found in the country. The plateaus and mountain groups above 8,000 ft. have a temperate climate with a wide diurnal range of temperature (night temperatures are often below freezing) and abundant rainfall, an average of 50 in. with the wettest regions in the southwest receiving as much as 80 in. All parts below 1,600 ft. are arid or semiarid and parts of the Danakil region are absolute desert. The Eritrean port of Massawa has the highest recorded annual average temperature of 86° F. The transition regions between the plateaus and the lowlands and the deep valleys of the plateau are subtropical and tropical with restricted quantities of rain. The Ethiopians have a traditional division of the country into *Dega*, *Woina Dega* and *Quolla* which describe the cool, warm and hot zones respectively.

The rainy season from July to September is known as *kremt* and is regarded as the winter season since cloud cover and heavy rain reduce temperature. This *kremt* rain is monsoonal in character and derives from a moist southwesterly air stream. *Bega*, the Ethiopian "summer," is actually the northern winter, but this is the dry season and absence of cloud cover allows a stronger heating effect. There is also an uncertain period of "little rains," as opposed to the "big rains" of *kremt*, which occurs for about four weeks during March and April. The high plateau frequently experiences very heavy falls of hail during *kremt;* these have sometimes given rise to reports of snow-covered mountains in Ethiopia. The highest parts of the mountains have an alpine climate.

4. Vegetation.—The country naturally divides into two main regions; the plateaus and mountains which have abundant rain but cool temperatures and the lowlands with high temperatures but deficient or uncertain rainfall. The natural vegetation of much of the highland is forest, with small groups of trees at higher levels and thickly wooded hillsides at medium altitudes. The most prominent trees are of the juniper type *tid* or *ted* (*Juniperus procera*) and *zigba* or *zegba* (*Podocarpus gracilior*). The fruit of the pinnated-leaved *cusso* or *kosso* (*Hagenia abyssinica*) is used by Ethiopians as a cure against tapeworm. In drier regions, on the higher parts and in deforested areas are mountain grasslands in which species of *Pennisetum* predominate but with species of *Eleusine*, *Setaria*, *Andropogon*, *Digitaria*, *Eragrostis*, *Sporobulus*, *Trifolium* and

sedges and rushes. In the highest regions afro-alpine vegetation, including giant senecios (*Senecio giganteus*) and giant lobelias (*Lobelia hynchopetalum*), is found. On the southwest plateaus, with slightly higher temperatures and higher rainfall, are large areas of broadleafed rain forest. The most frequent species are *karraro* (*Pouteria ferruginea*), *sissa* (*Albizzia schimperiana*), *tukur inchet* (*Pygeum africanum*), *sombo* (*Ekebergia rueppeliana*) and *dogma* (*Syzygium guineense*). The coffee plant is indigenous to this region and occurs naturally in the forest particularly in Kaffa province, whence it takes its name. Patches of bamboo forest (*Arundinaria*) occur particularly in the southwest. The foothills and lowlands are the habitat of acacia, *Euphorbia candelabra*, varieties of *Ficus* and *Olea*, and differing environments produce xerophilous woodland or open woodlands, savannas and subdesert scrub.

Climatic and soil conditions produce a variety of trees and fruits. Palms, mimosas, giant sycamores, laurels, myrrh and other gum trees, oranges, limes, pomegranates, peaches, bananas, grapes and other fruits, cotton, sisal, indigo and sugar cane are all found. In certain areas (the Gurage mountains and Sidamo), the false banana (*Musa ensete*) is grown for food. A fairly recent feature of the landscape is the imported eucalyptus (*Eucalyptus globulus*), widespread on the plateaus and round all Amhara settlements. Its local name is *bahr zaf* (tree from over the sea).

An Ethiopian national park is conserved at Managasha forest, 50 mi. W. of Addis Ababa.

5. Animal Life.—The fauna has links with those of the Palearctic and with those of the east African regions. Elephant and rhinoceros are found in low-lying districts, but rhinoceros is mainly restricted to the Akobo plain in western Ethiopia. The hippopotamus and crocodile inhabit many of the rivers and lakes, in some of which otters of large size are plentiful. Nile monitors are found in many lakes and rivers in western Ethiopia. Lions occur mostly in the south and southwest and in the Ogaden. Leopards, spotted and black, are numerous and often large; hyenas are found everywhere and are hardy and fierce; lynx, wolf, wild dog and jackal, wart hogs and badgers are also seen. The giraffe is found in the west; the zebra frequent the lower plains but the wild ass is rarer. Antelopes and gazelles of many varieties are numerous in most parts and include greater and lesser kudu, oryx, duiker, gemsbok, hartebeest, gerenuk and klipspringer; the dik-dik, the smallest of the antelopes, rarely weighing more than 10 lb., is common in the low countries and the foothills. The rarest of all these is the nyala, found in remote high mountain areas (*e.g.*, Simen and Arussi mountains). The civet is found in many parts but particularly in Kaffa province. Squirrels and hares and monkeys, notably the guereza and dog-faced baboon, range abundantly from the warm lowlands to heights of 10,000 ft.

Eagles, vultures, hawks, bustards and other birds of prey, partridges, duck, guinea fowl, sand grouse, curlews, woodcock, snipe, pigeons, thrushes and swallows are plentiful. A variety of ostrich is common. The Abyssinian ground hornbill is protected. Among birds prized for plumage are the marabou, crane, heron, blackbird, parrot, jay and many sunbirds of extraordinary brilliance.

Python, cobra and puff adder are found with a variety of smaller snakes. Malarial mosquitoes breed in low lying areas. Seasonal locust plagues occur in some eastern districts. There are thousands of varieties of butterflies and other insects, a large proportion of which remain unclassified. The most important game reserves are the Simen mountain area and Chillalo mountains (where nyala occur) and the Awash valley. (G. C. L.)

II. THE PEOPLE

1. Ethnology.—The inhabitants of Ethiopia consist of two main groups, Hamitic Ethiopian (sometimes called Cushitic) and Semitic, known collectively as Amhara. Both are intrusive, and entered the country from Arabia. Before their arrival the land was occupied by hunting peoples whose scattered descendants remain in southern Ethiopia as Watta (Wayto), Manjo, Fuga, and Doko; and by a Negro element which possibly appeared as early as the 8th millennium B.C. The Negroes may not have penetrated north of the Abbai and Jamma rivers, but they undoubtedly mingled with the Hamites, through whom some Negroid traces have passed to the Semitic stock. The Hamites seem to have entered northeast Africa from southwest Arabia long before the Semites in three principal waves: (1) Proto- or A-Hamites, possibly represented by the Barya (Barea) and Kunama in the north. (2) Early or B-Hamites, consisting of (a) the Agaw, whose original home was between the Takkaze and Abbai rivers and whose surviving groups, from south to north, are the Awiya, Quara, Kemant, Falasha (*see* FALASHAS), Khamta, Khamir, and Bilin; and (b) the Sidama peoples of southwest Ethiopia, of whom the most important are the Sidamo, Hadya, Kambatta, Wolamo (Wallamo), Kafa and the Gimira group. The most northerly Hamites, the Bedawiye or Beja (*q.v.;* the Blemmyes of the Romans), are perhaps to be included among the B-Hamites. (3) The later or C-Hamites, comprising the eastern peoples Galla (*q.v.*), Somali, Afar (Danakil) and Saho. The Sidama peoples, possibly the first to enter the Negro area, acquired some Negroid characteristics and then gradually pushed the Negroes westward; surviving Negro groups remain in southwest Ethiopia (Konso) and in Darfur (Koma, Berta, Ingassana and others, known collectively in Amharic as Shanqella or Shangalla). The C-Hamites coming last escaped the Negroid characteristics of the Sidama. In the 7th century B.C. or earlier, Semitic peoples from the Yemen began to colonize the northern Ethiopian highlands, where they established a kingdom with its capital at Aksum, out of which grew the kingdom of Ethiopia. These Semites did not remain ethnically pure, for they mingled with the Hamitic inhabitants of the country, though they preserved a characteristic "Ethiopian" physical type. At a much later date, further intrusive elements from the Muslim states in the eastern lowlands reached the highlands, and by the 10th century A.D. established a Muslim sultanate in Shoa. Though the Muslims, mainly of Hamitic stock, entered various parts of the kingdom of Ethiopia, their main influence was in the south and southwest to which trade routes led from the coast at Zeila in Somalia (*see* also CUSHITIC PEOPLES; AFRICA: *Ethnography*).

2. Languages.—Before the arrival of the Semites the main language stock in northeast Africa was the Hamitic (Cushitic), comprising: (1) Bedawiye; (2) the Agaw group; (3) the Sidama group, consisting of (a) Sidamo, Darasa, Hadya, Kambatta, Alaba; (b) Wolamo, Gofa, Kullo, Konta, Zala, Kuera or Koyra (Baditu), Kucha, Gamo; (c) Chara; (d) Basketo, Zaysse, Doko; (e) the Gongo group including Kafa, Shinasha, Mao; (f) the Gimira group including Gimira, Maji (Mazhi), Nao; (g) Yamma or Janjero; (4) the separate but related languages Galla, Somali and Saho-Afar.

The Semitic peoples brought with them a south Arabian Semitic speech which in Africa developed into Geez (Ge'ez), often called Ethiopic, which ceased to be spoken about the 10th or 11th century but remained the ecclesiastical and literary language of Ethiopia (*see* ETHIOPIAN LITERATURE). From Geez developed three major living languages: Tigrinya (Tigriña) and Tigrē, spoken in the north, and Amharic (*q.v.;* now the official language) spoken by about 40% of the people south of the Tigrinya area. Of these, Tigrinya is closest to Geez, Amharic farthest, with Tigrē in between; Amharic has acquired substantial Hamitic accretions. In the south, four minor Semitic languages of Ethiopic type were spoken: one, Harari (Adare) in the city of Harar, the second in Gurage south of Addis Ababa, the third (extinct) in Gafat in the Abbai river area and the fourth (fast disappearing) in Ankober northeast of Addis Ababa. English, the second language of education, is widely spoken by the intelligentsia and in official circles, while Italian and Arabic are understood in many places. (*See* AMHARIC; AFRICAN LANGUAGES; SEMITIC LANGUAGES.)

3. Religion.—The established religion is that of the Ethiopian church, the head of which is the abuna, with the emperor as titular leader. (*See* ETHIOPIA, ORTHODOX CHURCH OF.) The provinces of Tigré, Shoa, Gojam, Begemdir and Sidama are mainly Christian, and in those areas the religion plays an important part in the life of the people. A peculiar form of Judaism is practised by the Falashas (Agaw stock) in Dambya, Begemdir and Semen. Islam (Shafi'ite) has many followers in the north, east and south, with an important centre at Harar. Pagan elements survive in parts,

especially in Wallaga, Sidama, and among the Arusi Galla. Under the constitution of 1955 the rites of all religions are given full liberty unless used for political purposes.

4. Customs.—The national dress is a togalike garment called *shamma;* women wear a gown tied with a girdle, and a shirt and drawers, with a *shamma* over it when out of doors. Men wear tight-fitting trousers like jodhpurs. The staple diet is *enjara,* a flat unleavened bread made from the grain called tef (*Eragrostis tef*) and a spicy curry (*wat*). There are many religious festivals which draw large crowds, among them Temqat (Epiphany), Tensa'e or Fasika (Easter) and Masqal (held in September to commemorate the finding of the cross by St. Helena, though the proper date is in May). The Ethiopian year is seven years behind the Gregorian. It begins on Maskaram 1, which in 1958 was September 11; and has 12 months of 30 days, the remaining five (six in leap year), called Paguemen, coming at the end of the last month, Nahase (August–September). (G. W. B. H.)

III. HISTORY

In classical times Aethiopia was the name applied to a district of northeastern Africa (*see* ETHIOPIA, ANCIENT).

The plateau of Ethiopia and the lowlands which surround it on the borders of the Red sea and the Gulf of Aden were inhabited from the most ancient times by a variety of peoples. Because of the lack of written evidence and the scarcity of archaeological remains their history must be reconstructed through their oral traditions and on the basis of their geographical distribution. The greater part of the territory comprised in modern Ethiopia (including Eritrea) seems originally to have been occupied by the Cushites, a group of peoples speaking Hamito-Semitic languages. In the lowlands they occupied a large zone extending from the Beja region in the Sudan to the mouth of the Juba river on the Indian ocean; on the plateau they were divided into two principal groups: the Agau in the north and the Sidama in the south. The problem of the original home of the Cushites is still unsolved: were they of African origin (this was the opinion of the Austrian orientalist Leo Reinisch) or did they represent an early immigration from Arabia, as now seems more probable?

In the southern and western zones of the Ethiopic plateau the Cushites drove out or brought into subjection various tribes speaking Nilotic or Sudanese languages. These tribes may have been the aboriginal inhabitants of the area rather than immigrants from the lowlands of the Nile basin.

1. Kingdom of Aksum.—This was the distribution of ethnic groups in Ethiopia when new immigrants from south Arabia crossed the Red sea and founded the first Ethiopic state, the kingdom of Aksum (*q.v.*). This immigration did not follow military conquest but is assumed to have resulted from Arab trading settlements on the Red sea coast of Ethiopia, whence the immigrants, successively reinforced, gradually penetrated the interior and were able finally to assert their rule in the northern zone of the plateau. It is impossible to state with any certainty when the south Arabian immigration began. The most ancient south Arabic inscriptions in Ethiopia are the fragments found at Yeha which have been attributed tentatively to the 5th century B.C. The composite origin of the south Arabian immigrants is proved by the fact that, whereas one group, the Habashat (Arabic Habashah, whence Abyssinia), gave the country one of its names, another group, the Geez, left its name to the Semitic language spoken in Ethiopia by the immigrants and adopted as its official language by the kingdom of Aksum.

The kingdom of Aksum itself was probably the result of a gradual federation of south Arabian immigrants. The title of *nagash* (connected etymologically with the Semitic root *ngsh,* "to exact tribute") was perhaps originally borne by the princes of the various immigrant groups who were still paying tribute to their original rulers in the Arabian peninsula; later on it came to signify "monarch" or "prince," and when the king of Aksum affirmed his sovereignty over the whole territory and the first united Ethiopic state was founded, he was styled *negusa nagasht,* "king of kings," whence arises the commonly used title of negus.

The importance in the ancient world of the Ethiopic state of Aksum stemmed from its geographical situation on the sea route from the Mediterranean to India. The foundation of Alexandria in Egypt and the exploration of the Persian coast by the fleet of Alexander the Great led to the establishment of Greco-Egyptian colonies on the Red sea which served as market places and harbours for the trading ships along the route to India. The principal settlement of this kind on the coast of Aksum was Adulis (modern Zula, 30 mi. S.S.E. of Massawa).

When Egypt became a province of the Roman empire in 30 B.C. the safeguarding of the route through the Red sea was considered an important imperial interest, especially because during the long struggle between the Roman (afterward, Byzantine) empire and Persia, the sea route was the Roman way of communication with India, while Persian trade followed the land route through Mesopotamia and Syria to the Mediterranean. This compelled both empires to pay attention to the events on the Ethiopic seashore. In the anonymous *Periplus Maris Erythraei* written in the 1st century A.D., Adulis is described as an "open harbour" where a settlement of Greco-Roman merchants were dwelling for the purposes of trade; and Pliny recorded in his *Natural History* the mission of a freeman of Annius Plocamus, who had been sent to collect taxes from the communities on the Red sea and was finally driven by a tremendous hurricane in the Indian ocean toward Ceylon.

The commercial rivalry between Rome and Persia influenced several of the enterprises undertaken by the kings of Aksum; *e.g.,* their expeditions into southern Arabia in the 3rd and the 6th centuries A.D. for the purpose of harassing the Persians. Two episodes in the reign of the emperor Justinian illustrate the effect of this international rivalry on Ethiopia. The historian Procopius in his *De bello Persico* states that the emperor sent an embassy to the king of Aksum asking that Ethiopian merchants should be ordered to buy in India the greatest possible quantity of silk arriving there from China and to sell it later to the traders of the Roman empire. The purpose of this was to prevent Chinese silk being bought by the Persians or their allies who would thus increase their stocks of this raw material which was so highly valued in the luxury trade of the day. Julian, the ambassador of Justinian, concluded an agreement to that effect with the Ethiopian king, but this had no practical consequence because of the countermeasures taken by the Persians to corner the Indian market in silk. The other episode was related by the Greek monk Cosmas Indicopleustes in his *Topographia Christiana* (written *c.* A.D. 547). He told the story of Sopatros, a merchant of Adulis, who in a discussion before the king of Ceylon, extolled the value of the *nomisma* (the golden coin of the empire) in comparison with the *miliaresion* (a silver Persian coin).

It was through the Greco-Roman merchant communities in the Red sea ports that Christianity, which has been historically the strongest link between Ethiopia and the western world and gave to Ethiopic culture its typical characteristics, first reached the country. It was spread among the Ethiopian ruling classes by St. Frumentius who, during the reign of King Ezana in the first half of the 4th century, arrived in Ethiopia from his birthplace in Syria as a prisoner captured after the wreckage of a trading ship going to India. Early Ethiopic literature (*q.v.*) was, for the most part, based on translations, direct from the Greek, of works necessary for religious teaching and Greek was the second language used in the official inscriptions of the Aksumite kings. On the other hand, the so-called obelisks (steles) of Aksum are carved with figures typical of southern Arabia and there are traces of Indian influence, perhaps in the architecture of the Ethiopic monolithic churches and, more certainly, in a few words adopted by the language (especially words used for articles of trade). Finally there are some features of possibly Indian origin in the material culture of the country such as the suspended wooden bridges and the leather gourds employed as water wings for the crossing of the rivers. The development of the religious, cultural and artistic history of ancient Ethiopia was thus determined by a number of different factors.

In the second half of the 6th century and in the early 7th century A.D. the fortunes of the Aksumite kingdom declined. During the war waged by the Roman emperor Justin I against the Persians,

Kaleb, the king of Aksum, had successfully invaded Yemen and established an Arab prince there as his tributary. The purpose of this enterprise was the protection of the Christian community of Nagran (in Yemen) which was being persecuted by the Yemenite king. After a time, however, the puppet Arab ruler was overthrown and the Ethiopian commander in Yemen, Abreha, declared himself independent of Aksum. After many victories (attested by the long inscription on the Marib dam), Abreha led an unsuccessful expedition against Mecca, which is recorded in the Koran. Then the Persians, who could not afford to leave both banks of the Red sea in the hands of the allies of the Byzantine empire, intervened directly and put an end (probably *c.* 572) to Ethiopian rule in Arabia. Some years later the rise of Islam and of the Arab empire cut Ethiopia off from the Mediterranean. Thus isolated Ethiopia was forced to turn toward the interior of the country and the Ethiopic middle ages began, a long period for which there is little historical evidence.

2. Zague (Zagwe) and Solomonid Dynasties.—The Ethiopic kingdom had, up to that time, been oriented toward the Red sea and the valley of the Atbara river. When the road to the sea was closed by the rise of Islam the obvious route for the expansion of the Aksumite kingdom lay southward. This increased the influence of the purely African element in the kingdom and, after a dark period of decline, a local dynasty, the Zague of Agau origin, gained power and transferred the capital of Ethiopia from Aksum to a town in the region of Lasta (on the central plateau) which was then called Roha (the Syriac name for Edessa). The chronology of the Zague dynasty is uncertain. Three of their kings, Yemrehanna Krestos, Na'akueto la-Ab and Lalibela, and one of their queens, Masqal Kebra, are venerated as saints by the Ethiopian church. The legend of the power of the Ethiopian king to divert the course of the Nile by a dam toward the Red sea was first attributed to the Zague kings and, as it was brought by Ethiopian pilgrims to Jerusalem, it was widely diffused in medieval Europe until finally it became the subject of an episode of the *Orlando Furioso* by Lodovico Ariosto, the Italian poet of the Renaissance. Lalibela, of the Zague dynasty, reigned at the beginning of the 13th century and local tradition regards him as the principal builder of the monolithic churches, particularly in the capital, Roha, which later assumed his name, Lalibela. In 1270, however, the Zague dynasty was overthrown and power passed to the Solomonid dynasty whose traditional genealogy begins with Menelik I, son of Solomon and the queen of Sheba (Saba).

The first Solomonid king, Yekuno Amlak, was helped in his victory over the Zague by his countrymen of the province of Amhara in central Ethiopia and by the monasteries which had already acquired great political influence. The movement of expansion toward the south was thus given a further impulse under the new dynasty and the kings were also obliged to show particular favour to the monastic clergy. Both these tendencies were enhanced by the impending struggle against the Muslim states on Ethiopia's southern border.

Islamic missionary preaching and a fresh wave of immigration from Arabia led to the conversion to the Muslim faith of many pagans living in northeastern and southern Ethiopia. This and the closer contact with Arabia had been followed by the formation, in southern Ethiopia, of various sultanates, more or less independent of each other, but recognizing the leadership of the sultanate of Shoa, which was governed in the 13th century by a dynasty of princes tracing their genealogy to the Arab tribe of the Makhzumi, rivals of the Omayyads in Mecca. The last Makhzumite sultan of Shoa was defeated and deposed in 1285 by Ali ibn Walasma', sultan of Ifat, the country immediately east of Shoa. The hegemony of Muslim Ethiopia then passed to the sultans of Ifat, who claimed to be descendants of 'Aqil ibn Abi Talib, brother of the caliph Ali. The victory of Ifat was an incentive to a Muslim advance against Christian Ethiopia and the change of dynasty from the Zague to the Solomonids who were granting asylum to refugees from the Muslim states, seemed to present another favourable opportunity.

The wars between Christian Ethiopia and its Islamic neighbours lasted about two and a half centuries. One of the first events recorded in the royal chronicles is the attack made by Sabr al-din, sultan of Ifat, in 1332 on the Ethiopian king 'Amda Seyon (or Sion) I (reigned 1314–44). Sabr al-din was defeated and obliged to surrender to the Christian king, but his imprisonment did not put an end to the struggle and a league of Muslim princes, under the leadership of Qadi Salih and Jamal al-din, brother of Sabr al-din, again attacked the Ethiopian army. A hard-fought battle, near Das in the lowlands of Adal, ended in the complete victory of 'Amda Seyon. This war against the Muslims was celebrated by the court historian of the negus in a chronicle which is one of the best works in Ethiopic literature. Later, Kings Dawit I (1382–1411) and Yeshaq (1414–1429) took the offensive against Ifat; Dawit I advanced to the southeast and defeated and killed in battle Sultan Haqq al-din (1386–87). Yeshaq who conquered and slew Sultan Sa'ad al-din in 1415 penetrated as far as Zeila on the Gulf of Aden. Yeshaq's victory was praised by his soldiers in songs remarkable for their harsh but effective poetical inspiration. On the Muslim side Sultan Sa'ad al-din was also acclaimed as a hero and local traditions still celebrate his glory.

Zara Yakub (Zar'a Ya'qob), the son of Dawit I, succeeded to the throne in 1434. He became one of the outstanding personalities in the history of Ethiopia and reigned till 1468. His first task was to tackle the delicate problem of the Church. Because of the part played by the monasteries in the accession of the Solomonid dynasty, the monks had greatly increased their influence and had been granted much landed property by the kings and local chieftains. The monasteries with their numerous vassals had sometimes intervened even in disputes on the succession to the throne and at other times had openly fought the reigning king. Jealousy had also grown up between the monasteries and the clergy of the royal court, the latter being organized as the "chaplains of the camp" to serve the chapels in the king's camp in peace and war. The difficulties of the Church were aggravated by the attitude of the native clergy to the metropolitan who was appointed to Ethiopia from Egypt; by heretical movements, particularly in the northern provinces; and by pagan beliefs and customs in the south. Zara Yakub initiated severe reforms in the Church, eliminating abuses by strong measures and promoting the preparation of works for the refutation and outlawing of heresies and pagan practices (*e.g.*, the *Book of Light;* the *Book of the Nativity;* the *Book of Custody of the Sacrament,* etc.). Both the principal groups of heretics, the Stephanites and the Mikaelites, were repressed and their chiefs condemned to death. The leadership of the monasteries in Ethiopia, a privilege accorded to the abbot of St. Stephen on Lake Hayq as a result of the help given by that establishment to the founder of the dynasty, was limited by the favour shown by Zara Yakub to the monastery of Debra Libanos (in Shoa near the capital of the kingdom). Finally, during the reign of Zara Yakub's successor, the abbot of Debra Libanos assumed the leadership from his colleague of St. Stephen.

During the reign of Zara Yakub the sultan of Ifat, Badlay ibn Sa'ad al-din, attacked Ethiopia. He was defeated and killed in a great battle near Egubba, in eastern Shoa, on Dec. 26, 1445. This crushing defeat caused the decline of the sultanate of Ifat and the hegemony of the Muslim states of Ethiopia passed to the sultan of Adal.

For many centuries the Ethiopians had sent pilgrimages to the Holy Land and an Ethiopian community, established in Jerusalem, had acquired properties there and certain rights in the holy places. In 1441 some Ethiopian monks were sent by Niqodamos, abbot of the community in Jerusalem, to the ecumenical council summoned by Pope Eugenius IV in Florence to discuss the question of union between the Eastern and the Catholic Church. Their arrival in Florence and their subsequent journey to Rome caused great interest in learned Renaissance circles and Antonio di Pietro Averulino (nicknamed "the Filarete") represented the scenes in two basreliefs on the bronze door of St. Peter's in Rome. This is one of the first examples of the realistic trend in Renaissance portraits (the wooden sculptures by Antonio di Michele da Viterbo on the same subject seem to have been lost when Pope Paul V in 1606 ordered their removal from St. Peter's). The Ethiopic sovereign also figured in the legend of Prester John (*q.v.*). The fact that

Ethiopia was a Christian country helped establish it as the source and scene of that legend.

An Ethiopic narrative has represented the period of prosperity in the early 16th century by the tale of Negus Lebna Dengel (who reigned from 1508 to 1540). So proud was he and so bored by his undisputed sovereignty that he commanded his soldiers to whip the earth that it might give birth to some enemies for him to fight. The earth then brought forth the left-handed (the *Grañ*) imam of the Muslims and a terrible war began. The capital of the Muslim sultanate had been transferred from Adal to the town of Harar in 1520 and a military chief born there, Mohammed ibn Ibrahim, surnamed by the Christian Ethiopians *Grañ* and by the Muslims "the Conqueror" (Sahib al-fath), made himself master of the sultanate and assumed the title of imam, although he recognized as nominal local sovereign the sultan Abu Bakr ibn Mohammed of the ancient Walasma' dynasty. In the year 1530–31 the imam invaded the Christian kingdom of Ethiopia with the object of subjecting the country to Muslim rule and converting the people to Islam. The invasion was successful and after the victory of Shimbira Kure the Muslims conquered a great part of the Ethiopic plateau. Debra Libanos and Aksum, the most venerated places in Ethiopia, were burned and when the negus, driven from one mountain fortress to another in the north, died in 1540 and was succeeded by his son Claudius (Galawdewos), Christian Ethiopia was in the greatest danger.

3. The Portuguese Interlude.—Meanwhile Portuguese interests in India had made it imperative for Portugal to control the rival activities of Venice and Egypt on the Red sea route to the east and in the late 15th and early 16th centuries a number of embassies were sent from Portugal to Ethiopia. Accompanying an embassy sent in 1520 was the priest and writer Francisco Alvarez who later described the country and its peoples in a book *Verdadeira informaçao das terras do Presto yoao*, the earliest detailed description of Ethiopia. The Red sea route was still further endangered when the Turkish sultan Selim I conquered Egypt in 1517 and with the victories in Ethiopia of Grañ, who had received help from the Turks in Yemen, Muslim control of the Red sea seemed imminent. In face of this peril the Portuguese viceroy of India Estevam da Gama, son of Vasco da Gama, sent an expeditionary force which was disembarked at Massawa in 1541. The Portuguese army, under the command of Christovam da Gama, the viceroy's brother, joined forces with Claudius' Ethiopian army in their resistance to the Muslims. Christovam da Gama was taken prisoner and executed in 1542; but a few months later a combined Portuguese and Ethiopian army defeated the Muslims near Lake Tana and Grañ was killed in the course of the battle. This victory temporarily weakened the strength of the sultanate of Harar, but in 1551 a new leader, Nur ibn Mujahid, came to power there. Christian Ethiopia was again menaced on both sides, from the north by the Turks who had occupied Massawa in 1557 under their pasha Özdemir, governor of Yemen, and from the south by Nur ibn Mujahid who had married the widow of the late imam Grañ promising her a full revenge. The Turks were driven back in 1558 but Nur defeated and killed Negus Claudius on March 23, 1559, and the head of the unfortunate king was taken as a trophy to Harar. Nur died in 1567 and his tomb in Harar became a place of pilgrimage for the Muslims of Ethiopia.

4. The Galla Invasion.—Christians and Muslims were both exhausted after this long struggle and easily fell a prey to a new invader from the south, the Galla. Before their invasion of Ethiopia the Galla tribes inhabited the basin of the Juba river and a territory corresponding to the northern districts of Kenya. They had a tribal organization based on a system of "classes of age," which they had probably borrowed from the Bantu. Hard pressed by Somalis in the north and by the Bantu Wa-Nyika tribe in the south the Galla sought an outlet toward the Ethiopic plateau. Crossing the Ganale Dora (the upper reaches of the Juba river), which marked their frontier with the Muslim state of Bali, they rapidly crushed the forces of the Sidama states in southwest Ethiopia and attacked and defeated the amir Nur ibn Mujahid on his return from his victory over Negus Claudius. Finally they invaded the territory of Christian Ethiopia and in a few years reached Gojam and the region of Lake Tana. A consequence of the Galla invasion was the removal of the centre of the Ethiopic kingdom northward to the region of Gondar, which was more easily defensible. For the same reason in 1577 the capital of the Muslim sultanate in the south had been transferred from Harar to Aussa in the lowland.

5. Roman Catholic Missions.—During this period closer relations between Ethiopia and the western world, especially Portugal and the Holy See, and the need to obtain European aid against the Muslim danger gave the Roman Catholic Church an opportunity for an attempt to arrive at a full union with the Ethiopian Church. A mission with instructions from St. Ignatius de Loyola arrived in Ethiopia in 1554 under the leadership of Andrea de Oviedo, a Spaniard. Years of patient work and hardship followed during which theological differences between the two churches were discussed and although the Ethiopians held fast to their ancient monophysite doctrine no feelings of personal animosity were aroused. The ideas and methods introduced by the missionaries had a deep influence on Ethiopic literature and led the Ethiopians to look more closely into religious and ethical problems affecting their own lives. Some years later, in 1603, a great and learned Spanish missionary, Pedro Páez (*q.v.*), arrived in Ethiopia. His personal authority and eminent qualities were such that in 1614 Negus Susenyos was persuaded to accept the doctrine of the two natures of Christ and to notify his acceptance to Pope Urban VIII and the king of Spain (in that period Portugal was united with Spain under the same king). But Páez died in 1622 and Susenyos' enemies brought up against him his conversion from the old doctrine. Furthermore the military aid sought by the negus in his fight against the Muslims and the Galla had so far been refused by the European powers. Finally in June 1632, Susenyos proclaimed his return to the monophysite creed and abdicated in favour of his son Fasilidas.

The departure of the missionaries and the severe repression that followed the advent of Fasilidas severed Ethiopia's last link with the Mediterranean world. The decline in importance for European trade of the Red sea zone and the decisive strengthening of the Turkish empire in Egypt and Arabia contributed further to its isolation. A long period of decay followed and lasted about two centuries. It is known to the Ethiopian chroniclers as "the age of the Kings of Gondar." The kingdom, menaced continually by the Galla and by internecine struggles between pretenders to the throne, was reduced to desperate straits. The recovery of the Ethiopic state in modern times was, above all, the work of Kassa, who adopted the throne-name of Theodore II, and whose reign began in 1855.

6. Theodore II (1855–68).—Theodore won his crown through a long succession of battles against powerful feudal rulers whose interest was to maintain a state of anarchy in Ethiopia. His first major achievement was in 1853 when he overthrew Ras Ali of Begemder, the last of the Galla chiefs who as "maires du palais" had reduced the nominal sovereign to the status of a puppet king. He next attacked Wube (Ubye), the prince of Tigré, whom he captured in 1854 after a victorious campaign. Then, having been crowned as negus in Aksum in 1855, he turned south and, after a cruel massacre of the Muslims of Wallo, marched against Haile Malakot, king of Shoa, a scion of the Solomonid dynasty. Haile Malakot died during his campaign (1856) and Shoa then surrendered to Theodore who took with him as hostage the young prince Menelik, Haile Malakot's son. These victories, however, did not put an end to the turbulence of the local chiefs, who rose one after another against the newly asserted central power. So exasperated was Theodore by this situation that he became more and more cruel in his repressions. The country was desolated by massacres, torture, mutilation of captured rebels and their families and destruction of their villages.

Meanwhile Europe was again showing interest in the Red sea region; France had sent a mission to the chief of Tigré and Britain accredited W. C. Plowden as envoy to Theodore's court. He was succeeded by C. D. Cameron, who was falsely suspected by Theodore, then in a violent crisis of mistrust and passionate anger partly due to a sickness which had brought him to the verge of madness.

Cameron was accused of participation in a supposed Egyptian plot against Theodore and was imprisoned with other British subjects and some European missionaries. As the Ethiopian king stubbornly refused to liberate the prisoners, a British expeditionary force under the command of Gen. Sir Robert (later Lord) Napier disembarked at Annesley bay (Gulf of Zula) in Jan. 1868 and advanced into the interior. Theodore, isolated by the hostility of the feudal leaders and the hatred of his people and attacked by Napier in his stronghold at Magdala, committed suicide on April 13, 1868.

7. John IV (1872–89).—Theodore's only son Alamayahu was sent to England and after the departure of the British expeditionary force three claimants fought each other for supreme power in Ethiopia. In 1872 Kassa, the chief of Tigré, was victorious and, after his coronation in Aksum, assumed the name of John (Johannes) IV. His first purpose as emperor was to compel the feudal leaders to recognize their full dependence on the crown. This brought him into conflict particularly with Menelik, who in 1865 had escaped from the fortress to which Theodore had imprisoned him and had proclaimed himself king of Shoa, and with Takla Haymanot king of Gojam. John was helped by the dissension between his two adversaries. But, submission was, as usual, followed by fresh attempts to gain independence. Menelik first recognized the sovereignty of John after a military expedition against Shoa in Feb. 1879, but his submission was only temporary and, after a short interval, he again began to scheme not only to regain his kingdom of Shoa, but, as a descendant of the Solomonid dynasty, to win the imperial throne.

The external situation of Ethiopia was also full of danger. An Egyptian army, with Massawa as its base, aimed to occupy at least the northern plateau of Ethiopia but in 1876 John defeated the Egyptians decisively at Gura. Another threatening situation arose when the Mahdi came to power in the Sudan and mahdist chiefs launched predatory incursions into western Ethiopia as far as Gondar. Furthermore, the opening of the Suez canal in 1869 restored to the Red sea its value as the most direct sea route to India and rival European powers were competing for positions of advantage there. Britain and France established themselves in Somaliland. Italy, after acquiring the Bay of Assab in 1869 through an agreement between an Italian private company and the local chiefs, occupied Massawa in Feb. 1885. This was followed by an Italian advance toward the interior. John himself took command of an Ethiopian army in the lowlands west of Massawa but decided not to attack the Italian fortified position at Saati, although one of his commanders, Ras Alula, had successfully surprised an Italian detachment near Dogali, in Jan. 1887. Meanwhile Menelik had regained his kingdom in Shoa and the Italians, who had originally entered into negotiations in 1872, renewed their contacts with him. Menelik, in expectation of a further attack on Shoa, asked for arms from the Italians. John was in fact preparing to advance against Menelik when a mahdist army attacked western Ethiopia. Marching to check the Muslim advance John was killed in a battle at Metamma in March 1889.

8. Menelik II (1889–1913).—Menelik was crowned emperor of Ethiopia on Nov. 6, 1889, and the Solomonid dynasty was restored to the throne. During the crisis occasioned by the succession the Italians occupied Keren and Asmara. On Jan. 1, 1890, the colony of Eritrea (*q.v.*), consolidating Italy's acquisitions on the Red sea, was proclaimed. Prior to this a treaty of friendship and co-operation between Menelik and Italy had been signed at Uccialli on May 2, 1889. Under article 17 of that treaty, Menelik as emperor of Ethiopia agreed to avail himself of the services of the Italian government for all negotiations and business he might have with other powers or governments. In the Amharic text of the treaty, however, the wording was such that although it was open to the Ethiopian emperor to consult the Italian government on his foreign relations if he so wished, there was no compulsion on him to do so. When, therefore, the Italian government gave notice that article 17 of the treaty established an Italian protectorate over his country, Menelik refused to accept the interpretation and asserted Ethiopia's independence. After difficult and abortive negotiations war broke out between the two countries in Nov. 1895. Menelik advanced toward Tigré province which was occupied by the Italians and, after hard but victorious fighting against Italian outposts at Amba Alagi and Makale (Maqale), camped with his army in the neighbourhood of Aduwa. At the end of Feb. 1896 Gen. Oreste Baratieri, the Italian commander in chief, decided to take the offensive and in the ensuing battle of Aduwa the Italian army was defeated by overwhelming Ethiopian forces. This marked the end of the war. The treaty of Addis Ababa, signed in Oct. 1896, abrogated the Uccialli treaty, and re-established peace and Italy recognized the independence of Ethiopia.

During the reign of John IV, as rival feudal chiefs Menelik of Shoa and Takla Haymanot of Gojam had both tried to conquer the regions of southern and western Ethiopia which were inhabited by Galla and Sidama peoples. In 1882 an expedition of Takla Haymanot into the country south of the Blue Nile was opposed by Menelik and he defeated his rival at Embabo. After this battle Menelik pursued his advantage and, with the help of capable and experienced lieutenants, took steps, both military and political, to obtain the total submission of the small states founded in this area during the decline of the Christian Ethiopic kingdom as a result of the dismemberment of the southern Muslim states after the Galla invasion in the 16th century.

In pursuit of his aims Menelik played skilfully on the rivalries of the European powers concerned. After the massacre of an Italian geographical mission near Harar he obtained Italian aid to avenge the murdered explorers and in 1887 conquered Harar taking prisoner the last amir, Abdullahi. Similarly in 1897 he permitted a group of the French mission to the Nile led by Jean Baptiste Marchand to pass through Ethiopic territory and authorized one of his best officers, Ras Tasamma, to follow them. Ras Tasamma conquered the territory in the lowlands of the Nile basin as far as Gambela and the Sobat river. The conquest of the kingdom of Kaffa in 1897 concluded these successful enterprises which had most important consequences for the Ethiopic empire. Various treaties concluded with Great Britain, France and Italy between 1894 and 1908 fixed the new frontiers of Ethiopia with the neighbouring territories ruled by the European powers. (*See* also MENELIK II.)

9. Lij Yasu (1913–16) and Zauditu.—Menelik died in 1913 and was succeeded by Lij Yasu (Lidj Iyasu). Before he was crowned, however, World War I had broken out and Lij Yasu gave such clear evidence of his sympathy with the Muslims and the Central powers then allied with Turkey, that in 1916 a movement strongly supported by the leading chiefs and by the church dethroned him and proclaimed Zauditu (Zawditu), the daughter of Menelik, as empress. Ras Tafari (whose Christian name is Haile Selassie) became heir apparent. After a struggle in which Lij Yasu and his supporters were defeated a period of internal peace ensued and in 1924 the heir apparent tried to strengthen the country's relations with foreign powers by a tour in Europe, Egypt and Palestine. In Dec. 1925, however, an Anglo-Italian agreement accorded Italy a sphere of economic influence in western Ethiopia and Italy undertook in return to give diplomatic support to a British demand that Ethiopia should build a dam at Lake Tana to regulate the water supply for irrigation in Egypt and the Sudan. Ethiopia appealed against the agreement to the League of Nations. This episode caused strained relations between the parties concerned but Britain did not press the Lake Tana project and Italy tried to reduce the tension by concluding (in Aug. 1928) a treaty of friendship and arbitration with Ethiopia. To this was annexed an agreement concerning a free zone in the port of Assab (in Eritrea) which was to be connected by a motor road with Dessie (in central Ethiopia).

10. Accession of Haile Selassie.—Some months later, after the failure of an attempt at a *coup d'état* by the leading feudal chiefs, the heir apparent Ras Tafari was appointed negus with full powers by the empress Zauditu. In April 1930 the old empress died and on Nov. 2 of that year the negus was crowned emperor under the name of Haile Selassie I. Meanwhile relations with Italy had deteriorated owing to disagreement over the Assab-Dessie road convention and a number of disputes and incidents on the Eritrean and Somali boundaries. In Dec. 1934 an Anglo-

Ethiopian commission was inquiring into the grazing rights of British Somali tribes at a point where the boundaries of Ethiopia and British and Italian Somaliland met. At the Italian armed post of Walwal a dispute arose between the Ethiopian delegates and the Italian officer in charge of the post about the possession of the wells of Walwal. The British delegation thereupon returned to its base, but the Ethiopian escort of the commission and the Italian Somali troops fought fiercely until the Ethiopians retired northward. Italy demanded heavy reparations from Ethiopia for this incident. The Ethiopian government appealed again to the League of Nations and a commission of arbitration was appointed in accordance with the provisions of the Italo-Ethiopic treaty of 1928.

11. Italian Occupation.—The commission of arbitration concluded its meetings on Sept. 3, 1935, deciding that neither party was to blame in the Walwal incident, but Italian military operations against Ethiopia began on Oct. 3, from the frontiers of both Eritrea and Somaliland. The League of Nations voted economic sanctions against Italy. In northern Ethiopia, Aduwa and Makale were soon occupied by the Italians; then, after several months of preparation by both sides, the Ethiopian army attacked the Italian expeditionary force, commanded by Gen. Pietro Badoglio, at May Chaw (Mai Chio), near Lake Ashangi, on March 31, 1936, and was defeated. On May 5 Badoglio entered Addis Ababa. The emperor Haile Selassie escaped through Djibouti first to Jerusalem and later to Geneva where he appeared in person to plead the cause of his people before the assembly of the League of Nations. His claim against Italy was not supported; on the contrary, the assembly decided that the economic sanctions should be ended. Thus the Ethiopian empire was annexed and was joined with Eritrea and Italian Somaliland to form Italian East Africa, and the emperor sought refuge in England. Italy decided to carry out a big program of public works concentrating especially on motor roads and agricultural and industrial development. Resistance movements against the occupying forces, however, continued, first in western Ethiopia under Ras Imru and later in Gojam and Shoa. An attempt in Feb. 1937 to kill the Italian viceroy, Marshal Rodolfo Graziani, in Addis Ababa, was followed by a long period of cruel repression. When the duke of Aosta succeeded Graziani as viceroy in Nov. 1937 he adopted milder methods, but resistance continued in Shoa under the leadership of Ras Abbaba Aragawi. (*See* ITALO-ETHIOPIAN WAR.)

12. Restoration of Haile Selassie.—After Italy had declared war on the western Allies in World War II the emperor Haile Selassie moved from England to Khartoum in the Sudan where he established contact with anti-Italian movements. The British offensive in east Africa began early in 1941; the Italian forces were defeated at Keren in Eritrea, at Amba Alagi, where after a gallant fight the duke of Aosta was taken prisoner, and later at Gondar. Addis Ababa was occupied and Haile Selassie returned on May 5, 1941, and resumed his reign. By the peace treaty, signed in Paris in Feb. 1947, Italy surrendered any claim to Ethiopia. In 1950 the United Nations general assembly decided that Eritrea should constitute "an autonomous unit federated with Ethiopia under the sovereignty of the Ethiopian Crown." This was carried into effect in Sept. 1952.

Since his restoration Haile Selassie has carried out a program of modernization and of political and social reform. A parliamentary regime was established and the feudal system abolished. Commercial agreements have been concluded with many foreign countries and Ethiopia has taken an increasingly active part in the work of the United Nations, including military co-operation in Korea in 1951 and the dispatch in 1960 of an Ethiopian contingent to the Congo.

By an agreement signed on Nov. 29, 1954, the Haud area of Ogaden region, which had been administered since 1941 by British Somaliland, was transferred on March 1, 1955, to Ethiopian administration. Grazing rights in that area, secured to Somali tribesmen by an Anglo-Egyptian agreement of 1897, and the delimitation of the frontier between Ethiopia and the former Italian Somaliland, were the subjects of negotiation between the governments of Ethiopia and the newly established Somali Republic.

In Dec. 1960, during the absence of the emperor in Brazil, a *coup d'état* was attempted by senior officers of the imperial guard and a group of young administrators. Addis Ababa was seized, the crown prince was proclaimed emperor and Ras Imru was to be head of a new government. The rebellion was crushed by the regular army, who released the crown prince. Haile Selassie returned from Brazil and order was quickly restored.

See HAILE SELASSIE. (EN. C.)

IV. POPULATION

No comprehensive census has ever been made. The United Nations estimate for 1961 was 22,000,000, including Eritrea. There was a comparatively large resident foreign population of Italians, Greeks and Armenians in addition to foreign experts employed by the various ministries.

Towns.—Addis Ababa (*q.v.*), the national capital since 1889 and capital of Shoa province, is the largest town, with a population (1961) of 449,021. Dase (*q.v.*) is the second largest town with (1956 est.) 53,500 and Harar and Dire Dawa (*qq.v.*) have populations of more than 30,000. There are 28 towns with populations exceeding 5,000. The most important are Gondar (*q.v.*), capital of Begemdir province, Aksum (*q.v.*), the ancient capital in Tigré province; Aisha (Aiscia) and Miesso on the railway between Addis Ababa and Djibouti; Gambela, a river port on the Baro river, Jimma, a coffee centre and capital of Kaffa province; and Debra Markos, capital of Gojam province. The total urban population is about 750,000. Large regular markets attended by thousands of people are held at Batie, Sokota and Waldia in Wollo province; Goba, Jijiga, Asba Tafari and Wardere in Harar province; Wanbera in western Gojam; Gore, capital of Illubabor province; Agaro and Bonga in Kaffa province; Fiche, Debra Birhan, Nazareth (Hadama) and Waliso in Shoa province; Dilla, Wondo and Negelli in Sidamo province; Makale and Adigrat in Tigré province; Gimbi, Dembidollo (Saio) and Nakamti in Wallega province; Debra Tabor and Metamma in Begemdir province; and Asselle and Shashamana in Arussi province.

There are many important historical centres in Ethiopia and Eritrea, besides Gondar, Aksum and Harar; notably Debra Libanos (monastic centre in Shoa), Ankober (former capital in Shoa), Lalibala (*q.v.*) (site of famous rock churches in Wollo province), Adulis (the ancient port of Aksum south of Massawa), Coloe (ancient Aksumite centre in Eritrea) and Debra Damo (ancient monastery in Tigré).

V. ADMINISTRATION AND SOCIAL CONDITIONS

1. Constitution.—Ethiopia is a constitutional monarchy. The head of state, who has wide powers and prerogatives, is the emperor Haile Selassie I (born 1892), the 225th ruler in the Solomonic line since Menelik I. The constitution granted by the emperor in 1931 and revised in 1955 established a prime minister, a council of ministers and a parliament consisting of an appointed senate and a chamber of deputies elected for four years, with a system of supreme, high, provincial and local courts. Sixteen ministries deal with the usual categories of state business, ministers being appointed by the emperor. There are no political parties and the emperor is the primary political force. Local and tribal institutions function, however, within the constitutional framework.

2. Local Government.—The provinces are administered under the ministry of the interior; the governors general of the provinces and governors of districts are appointed by the emperor. Excluding Eritrea, there are 13 provinces, divided into subprovinces (*awrajas*), again divided into districts (*weredas*) and finally 1,104 subdistricts (*mutikil weredas*). The provinces are: Kaffa, Shoa, Harar, Tigré, Gojam (*qq.v.*), Arusi, Begemdir, Gamu-Gofa, Bale, Illubabor, Sidamo, Wallega and Wollo. Of the larger centres, 124 have a municipality with a mayor responsible to the ministry of the interior. The governors exercise authority through the *balabats* (local chiefs).

Under the federal act of 1952 uniting Ethiopia and Eritrea, the latter maintains a separate administration headed by the emperor's representative in Eritrea but retains considerable local autonomy in domestic matters.

3. Living Conditions.—Little information is available about average incomes and standards of living. Cash incomes in rural districts are extremely low as subsistence farming is supplemented by exchange and barter. Labourers' incomes in the larger centres range from Eth. $20–24 per month while graduates of the University college have an established salary of Eth. $450 per month. The ministry of commerce family budget survey of 1952 in Addis Ababa showed that about 50% and 11% of incomes between Eth. $20 and $200 per month were spent on food and rent respectively. Except in Addis Ababa and other large towns, dwellings (tukuls) are built mostly of mud and mortar with conical thatched or tin roofs.

4. Health.—The ministry of public health, established in 1947, administers the nutritional, preventive, educational, social and sanitary measures of health protection, and in co-operation with other ministries and private and semiofficial bodies (Ethiopian Red Cross society, Imperial Welfare trust, mission clinics, medical missions from other countries, research teams from the World Health organization and the United Nations children's fund) has a full program for raising health standards. There are 20 government hospitals and clinics in Addis Ababa and 36 government hospitals in the provinces; in addition there are about 80 private and mission clinics. Special institutions include the Institut Pasteur d'Ethiopie which is the national central laboratory and research institute; and the Haile Selassie I public health college and training centre, Gondar, organized in co-operation with WHO, the U.S. technical assistance mission and UNICEF to train health officers for community centres.

Malaria, tuberculosis, dysentery, pneumonia, smallpox and typhus are believed to cause most deaths. In the 1950s pilot projects for malaria control were initiated, with international aid, in several regions. Campaigns against tuberculosis and eye diseases (especially trachoma) are carried on through the schools, and the aim is to develop the rural schools into community centres which will train people in sanitation, child care and pest control. A leprosy control service is centred in the Princess Zenebework memorial hospital in Addis Ababa. There are training schools for nurses at four major hospitals in Ethiopia.

5. Education.—Modern education in Ethiopia dates from the establishment of schools in the capital by Menelik II. Progress was seriously interrupted during the Italian occupation 1936–41, when schools were closed and many of the graduates of the education system lost their lives. After 1941, under the personal leadership of Haile Selassie I, an ambitious 10-year program of government elementary, secondary, post-secondary and college education was instituted to combat the widespread illiteracy. Elementary education in the provinces is financed from a special education land tax; all other education is financed from the central treasury. Eritrean education is financed separately by the Eritrean government.

The government system has expanded from 240 schools, with 32,000 children and 1,000 teachers, in 1946 to more than 630 schools with about 160,000 children and 4,200 teachers in the early 1960s. In addition, there are foreign mission schools, private schools and Ethiopian Orthodox Church schools. A further 15,000 children are accommodated in mission schools, and the church schools, with an unspecified number of pupils, contribute greatly to the spread of literacy and the preservation of the national culture. At the apex of the education system are the University college (founded 1950), the College of Engineering, the Ethio-Swedish building institute and the Theological college (all in Addis Ababa), the Gondar health college, the Military academy (Harar), the Naval academy (Massawa) and the Alemaya College of Agriculture and Mechanical Arts (Harar). There is a four-year teacher training institution of secondary level in Harar and a special education course for secondary teachers at the university college. Many students are sent abroad on postgraduate scholarships and students from other African countries are offered scholarships at colleges in Ethiopia under a plan initiated by the emperor in 1958.

6. Justice.—The legal system is said to be based on the Justinian code. A penal code based on new concepts came into force in 1958, and in 1960 civil, commercial and maritime codes were promulgated by the emperor. Under the judicial system established in 1942, the hierarchy of courts consists of a supreme imperial court, an imperial high court, general governorate courts, provincial courts, regional and subregional courts. Upon federation with Eritrea the two imperial courts were vested with federal judiciary powers to sit as federal courts. Jurists, including some Europeans, are appointed by the emperor. There are special tribunals for Muslims.

7. Defense.—A ministry of defense was created in 1953 to reorganize the armed forces following the federation with Eritrea, with jurisdiction over the army, navy and air force. A British military mission assisted with the training of the army from 1942 to 1950; thereafter the Ethiopian authorities established training schools for officers and other ranks. In 1958 an Indian army mission assisted with the establishment of a military college for officers in Harar. The imperial bodyguard has a separate organization. The air force maintains a training centre at Debra Zeit near Addis Ababa where Swedish and U.S. instructors train Ethiopian pilots in the handling of jet aircraft. Under the department of marine, training schools with Norwegian instructors for petty officers and officer cadets were set up at Massawa. The emperor is commander in chief of the armed forces.

VI. THE ECONOMY

The Ethiopian economy is primarily agricultural and pastoral, with a maximum of only about 200,000 people engaged in non-agricultural occupations. Exports are almost entirely agricultural products and derivatives. In the period 1942–56 separate plans for long-term development of various branches of the economy were drawn up, such as a five-year program for agriculture, the Assab harbour development project, the Addis Ababa city plan and other programs. In 1956 the emperor established a central planning board to co-ordinate the separate projects. The first general five-year plan (1957–61) established priorities for investment and geared the education system to the training of the various personnel required. The low level of power production received particular attention and investigations were begun into the hydroelectric potentialities of the Blue Nile river system. The general trends were toward improvement and extension of agricultural output with particular attention to industrial crops to provide a further basis for local industries. A land reform program to introduce an equitable system of land ownership was proclaimed in 1959.

1. Agriculture and Livestock.—Agriculture and stock raising are the main sources of private and public income. The variety of climatic and physical conditions enables a great range of temperate, subtropical and tropical plants to be cultivated. Many regions are exceptionally fertile and with the abundant rainfall can yield several crops a year. In the domestic economy, however, livestock products are the most important. Much of the meat and milk produced is sold locally and may fetch Eth. $300,000,000 annually in addition to exported hides and skins valued at Eth. $20,000,000. The total livestock (excluding poultry but including cattle, sheep, goats, horses, mules, donkeys and camels) was estimated about 70,000,000. Much of the land is admirable for stock raising. Most of the cattle (a type of Zebu) are kept in the dry plains, but sheep, horses and dairy cattle are raised on the pastures of the wetter uplands. Mules and donkeys (estimated together at nearly 5,000,000) are fundamental to the internal commerce of the country. By the early 1960s the incidence of anthrax, rinderpest and other animal diseases was reduced.

The most important cash crop is coffee which provides an annual export value of about Eth. $100,000,000. Several varieties are marketed (Harar, Jimma, Lekemti, Sidamo) and both plantation coffee and wild coffee from the forests of the southwest are important. The long rainy season and the general elevation of the land (5,000–6,500 ft.) is ideally suited to the coffee tree. In the Harar, Arusi, and Sidamo highlands, rainfall is less certain and irrigation schemes have been developed. A national coffee board was established in 1957 to maintain standards in production and processing, and to set up fixed and mobile cleaning plants.

In the highlands, grains of many kinds, including teff (the indigenous love grass *Eragrostis abyssinica*), millet, maize (corn), barley, oats and wheat, are grown. Oilseeds of about 12 different species (niger [noug], castor bean, cabbage seed, flax, peanut, rapeseed and sunflower seed) are grown at slightly lower altitudes and represent a valuable export. The sugar plantation at Wonji in the Awash valley supplies Ethiopia's needs and cotton cultivation is widely encouraged. Catha or chat (a stimulant) is an important local product near Harar.

2. Minerals.—A variety of mineral deposits exist but there has been no extensive exploitation. In the southwest, ultrabasic rocks of Pre-Cambrian age have a promising mineral content. There are platinum deposits at Yubdo in Wallaga province and gold, both alluvial and in quartz veins, occurs in several districts. The Italians during their occupation claimed to have located copper, lead, magnesium and iron. A disadvantage is that most of the Pre-Cambrian outcrops lie far from the main communication network. The sedimentary series includes economic deposits of lignite (especially at Chilga in Begemdir), limestone, rock salt and building stone. There is a possibility of oil-bearing strata in the Mesozoic rocks of the Ogaden region in the southeast and a drilling concession is held by a West German company. A school for mining engineers and geologists is conducted under the ministry of mines.

Salt is exploited chiefly in the Danakil depression, Tigré, Wollo, Harar and Sidamo provinces; there are also extensive solar plants on the Eritrean coast.

3. Industry and Power.—There are about 140 major and minor industrial establishments. Textiles (especially cotton), leather goods (especially shoes), food products (flour, bread, macaroni, canned tomatoes, edible oils), drink (beer, wine and mineral water) and tobacco and cigarettes are most important. There are also sawmills and furniture factories and a cement factory at Dire Dawa uses local limestone deposits. Further industrial development may be based upon hydroelectric installations completed at Koka on the Awash river south of Addis Ababa. These supply additional electricity to Addis Ababa, Dire Dawa and Harar, and settlements in between.

4. Trade and Finance.—Ethiopia's trade was about Eth. $411,900,000 in 1960 (imports Eth. $219,300,000; exports and re-exports Eth. $192,600,000). The balance of trade depends to a great extent upon coffee, the export of which fluctuates between 30,000 and 55,000 metric tons annually. Efforts are being made by the national coffee board to improve the production and marketing of this crop. At the same time the government seeks to extend the range of agricultural exports including oilseeds, hides and skins, cereals and pulses. The principal imports are textiles, petroleum products, machinery and tools, and motor vehicles.

About 70% of imports come from 6 countries (Italy 13%, India 13%, U.S.A. 12%, Japan 12%, U.K. 10%, Federal Republic of Germany 10%). Indian and Japanese trade is almost entirely one way. About 40% of exports (mainly coffee) go eventually to the United States market (although about one-third of this is exported in the first instance to Aden and Djibouti), about 30% (coffee, oilseeds, hides and skins) go to Europe and the remainder (largely foodstuffs) to the middle east. There is a small export to other African countries.

Banking and Currency.—The State Bank of Ethiopia (created in 1942 and extended to Eritrea in 1952) is the central bank of the empire and the only one permitted to deal in foreign exchange and to issue currency. The Maria Theresa silver thaler or dollar (a very fine coin bearing the head of the empress Maria Theresa and minted in Vienna with the date 1780) was, as in many adjacent areas, the ruling currency in Ethiopia toward the end of the 19th century and probably 60,000,000 were in circulation on the eve of the Italian invasion in 1935. Several attempts were made to replace it with national currency but by the liberation in 1941 it was still generally current, this time with the addition of the east African shilling. A currency reform in 1945, when coinage and paper currency were issued (5 cent, 10 cent and 25 cent coins and 1, 5, 10, 50, 100 and 500 dollar notes) was carried out in accordance with the Bretton Woods agreement, with the value of the currency based on gold. The Maria Theresa dollar was officially withdrawn in 1946 but is still found in remote areas. It is from this period, when Eth. $10 equaled £1 sterling, that the terminology developed whereby the 50-cent piece became known as "one shilling" and Eth. $10 became known as "one pound." In remote areas barter continues; bars of salt are sometimes used as currency and the copper coinage is often more acceptable than paper money as individual transactions are generally of low value. By the early 1960s the rate of exchange was Eth. $1 = 40 cents U.S.A., or Eth. $7 = £1.

The Development bank, a savings and investment bank which replaced the old Agricultural bank in 1951, assists in channeling savings into productive development in agriculture and industry. One feature is the granting of small agricultural loans to individual farmers for agricultural improvement.

National Finance.—The financial year in Ethiopia begins on Sept. 11 and revenue is derived from all the normal sources. There are two budgets: (1) an ordinary budget covering normal revenues (mainly from taxation—direct, indirect, excise and federal) and expenditures (mainly entered under ministry and departmental headings); (2) an extraordinary budget covering realization of state assets, loans, reparations, capital outlays and special items. In the late 1950s total actual revenue exceeded Eth. $168,000,000 and total budgetary expenditure exceeded Eth. $175,000,000. Tax revenue amounts to about 88% of total revenue excluding loans. Normal public debt is not large and the bulk of loan revenue (from the Export-Import bank and the International Bank for Reconstruction and Development) is attached to specific projects such as the construction of Assab harbour, highways and airports.

5. Transport and Communications.—The road system comprises about 2,850 mi. of all-weather roads (of which about 950 mi. are bituminized) maintained by the imperial highway authority and about 10,500 mi. of unimproved roads; finance for development is supplemented by a loan from the World bank. These roads radiate from Addis Ababa to Asmara and Massawa through Dase and Adigrat, to Debra Markos over the Abbai bridge, to Nakamti (due west of Addis Ababa), to Jimma in the southwest, south to Negelli and Moyale on the Kenya border and east to Dire Dawa through Nazareth and Awash. In addition there are important road connections between Asmara and Gondar through Aksum; Asmara and Tessenei through Keren; Asmara and Massawa; across the desert from Dessie to Assab; and from Dire Dawa through Harar and Jijiga to Hargeisa in the Somali republic. The roads built during the Italian occupation were primarily strategic and their alignment with the Italian outlets of Massawa and Mogadishu does not always fit Ethiopian requirements. The spectacular but rapid mountain road building at Debra Sina and Alamata on the edge of the escarpment presents large maintenance problems and the special climatic problems of the wet and dry seasons inflate maintenance costs. Road transport is a vital feature of internal commerce and there is a large construction program for feeder roads to the main highways.

Ethiopia has two railways. The 487-mile single-track metre-gauge line from Addis Ababa to Djibouti in French Somaliland, completed in 1917, is operated by a Franco-Ethiopian company. The 191-mile railway from Massawa to Asmara and Agordat, in Eritrea, is operated by the Eritrean government and has a much steeper gradient as it climbs the 7,000-ft. escarpment in about 50 mi. Since the development of the port of Assab there is a three-way split in external trade between Massawa, Assab and Djibouti. In the early 1960s there was an increase in tonnage handled at all three ports but the trend is to develop the Eritrean ports at the expense of the French-controlled port of Djibouti. The rivers are for the most part unnavigable. The Baro river is navigable (May–October) from Gambela to Khartoum for boats of 3 ft. draft or less.

The mountainous nature of the country gives a special importance to internal air traffic. Ethiopian Airlines, government-owned and established in 1945, operates about 30 airports most of which are open throughout the year. There is a considerable passenger and freight traffic between the provincial landing grounds and Addis Ababa, and freight rates often undercut road transport.

Postal, telegraph and telephone facilities are operated by the Imperial Board of Telecommunications, a semiautonomous agency. *See* also references under "Ethiopia" in the Index volume.

(G. C. L.)

BIBLIOGRAPHY.—James Bruce, *Travels to Discover the Source of the Nile,* 5 vol. (1790); C. F. Rey, *Unconquered Abyssinia as It Is Today* (1923), *In the Country of the Blue Nile* (1927), *The Romance of the Portuguese in Africa* (1929), *The Real Abyssinia* (1935); Evelyn Waugh, *Remote People,* published in U.S. as *They Were Still Dancing* (1932); L. de Castro, *Etiopia, Terra, Uomini e Cose,* 2 vol., 2nd ed. (1936); D. R. Buxton, *Travels in Ethiopia* (1949); Addis Ababa Chamber of Commerce, *Guide Book of Ethiopia* (1954); J. Doresse, *Ethiopia, Ancient Cities and Temples* (1959); G. Stefanini, *Saggio di una Carta geologica dell' Eritrea, della Somalia e dell' Etiopia* (1936); Y. A. Haggag, *A Contribution to the Physiography of Northern Ethiopia* (1961); Thomas Pakenham, *The Mountains of Rasselas* (1959); G. K. N. Trevaskis, *Eritrea, A Colony in Transition: 1941–52* (1960); Bjorn von Rosen, *Game Animals of Ethiopia* (1953); Jurgen Freiherr von Wolff, *Mammals of Ethiopia* (1955); R. E. G. Picci Sermolli, *Una Carta geobotanica dell' Africa Orientale* (1957); C. Conti Rossini, *Etiopia e Genti d' Etiopia* (1937); C. F. Beckingham and G. W. B. Huntingford, *Some Records of Ethiopia* (1954); E. Ullendorf, *The Semitic Languages of Ethiopia* (1955), *The Ethiopians* (1960); International African Institute, *Ethnographic Survey of Africa: North-eastern Africa* (1955–56); A. N. Tucker and M. A. Bryan, *The Non-Bantu Languages of North-Eastern Africa* (1956); A. H. M. Jones and E. Monroe, *History of Abyssinia* (1935); Sir E. Wallis Budge, *History of Ethiopia, Nubia and Abyssinia,* 2 vol. (1928); *Atti del convegno internazionale di studi etiopici,* published in Rome (1960) and setting forth the condition of Ethiopic studies in various fields; C. Conti Rossini, *Storia d'Etiopia,* vol. i (1928), *Historia regis Sarsa Dengel,* 2 vol. (1907) and *Italia ed Etiopia dal trattato di Uccialli alla battaglia di Adua* (1935); O. G. S. Crawford (ed.), *Ethiopian Itineraries c. 1400–1524* (1958); E. Cerulli, *Etiopi in Palestina,* 2 vol. (1943–47), *Documenti arabi per la storia dell' Etiopia* (1931), *Scritti teologici etiopici dei secoli XVI e XVII,* 2 vol. (1958–60); J. Perruchon, *Histoire des guerres de Amda Syon, roi d'Ethiopie* (1890), *Les Chroniques de Zar'a Ya'eqob et de Ba'eda Maryam, rois d'Ethiopie* (1893); E. Littmann *et al., Deutsche Aksum Expedition,* 4 vol. (1913); W. E. Conzelman, *Chronique de Galawdewos roi d'Ethiopie* (1895); F. M. Esteves Pereira, *Chronica de Susenyos, rei de Ethiopia,* 2 vol. (1892–1900); Guebre Sellassie, *Chronique du regne de Menelik II,* French trans. by Tesfa Sellassie, 2 vol. (1930–32); C. Sandford, *Ethiopia under Haile Selassie* (1946); C. Beccari (ed.), *Rerum Aethiopicarum scriptores occidentales inediti,* 15 vol. (1903–17); Ministry of Information, *The Abyssinian Campaign* (1942); M. F. Perham, *The Government of Ethiopia* (1948); E. S. Pankhurst, *Ethiopia: A Cultural History* (1955); Board of Trade, *Report of U.K. Trade Mission to Egypt, Sudan and Ethiopia* (1955); Ministry of Commerce, Addis Ababa, *Economic Progress of Ethiopia* (1955); *Economic Handbook* (1958); H. P. Huffnagel (ed.), *Agriculture in Ethiopia* (1961); W. E. H. Howard, *Public Administration in Ethiopia* (1956); R. Pankhurst, *An Introduction to the Economic History of Ethiopia* (1961).

Current history and statistics are summarized annually in the *Britannica Book of the Year.*

ETHIOPIA, ANCIENT (AETHIOPIA) is not an exact geographical term. Ethiopia was a general Greek name for the country of the dark-coloured peoples living south of Egypt. There seems to be little reason for supposing that the derivation of the name is anything other than its literal meaning: "burned-faces." *Aithiopes* are first mentioned in Homer, who describes them as the furthest of mankind; the gods go to their banquets and probably the sun sets in their country. They are divided by Herodotus into two main groups, a straight-haired race and a woolly-haired race, dwelling respectively to the east and west (vii, 70). In later classical times they are the inhabitants of the kingdom of Cush, with its capital first at Napata (*q.v.*) and then at Meroe (*q.v.*). For the archaeology and history of this area, *see* NUBIA; SUDAN, REPUBLIC OF THE. The name Ethiopia (*q.v.*) has in modern times been adopted by Abyssinia.

(A. J. AL.)

ETHIOPIA, ORTHODOX CHURCH OF, the national church of Ethiopia, which has been a Christian country since the 4th century. In the year 341 a certain Frumentius on a journey from Tyre with his brother was captured at an Ethiopian port and sent to King Ezana at Aksum. The two young men won the confidence of the king and began to evangelize the country. Later Frumentius was consecrated by Athanasius of Alexandria to be its first bishop.

Toward the end of the 5th century A.D. monasticism reached Ethiopia, traditionally being brought by nine monks from Syria, under whose influence the Scriptures were translated into Ge'ez and the Ethiopian Church was strengthened in the monophysite doctrine which it owed to its connection with Alexandria (*see* MONOPHYSITES). In the 7th century the conquests of the Muslim Arabs cut off the Ethiopian Church from contact with its Christian neighbours save for the minority Coptic Church in Egypt. At the same time the Aksumite kingdom was engaged in a struggle with the pagan peoples in the south of the country, whom they gradually converted. Under the Solomonid kings (late 13th century onward) this process was completed, but the church became infected with various syncretistic beliefs. Its monastery in Jerusalem kept a certain contact with the outside Christian world, and Ethiopian delegates attended the Council of Ferrara-Florence (1438–45). King Zara Yakub (15th century) made determined attempts to root out heresy and paganism by his enlightened reforms. In the 16th century Portuguese priests tried to convert the Ethiopians, and one of them, Pedro Paez, persuaded the emperor Susenyos (1607–32) to become a Roman Catholic. But the latter's attempt to force his new church on the whole country aroused such opposition that he abdicated, and his son ejected the Jesuit missionaries. Franciscans and other western missionaries who later came to Ethiopia had little success; the Ethiopian Church meanwhile was continually struggling with its Muslim neighbours. In the 20th century the Ethiopian Church began to press for complete independence from the Coptic Church, from which it had always taken its archbishop. This was finally achieved in 1959, when the Ethiopian archbishopric was raised to a patriarchate; although consecrated by the patriarch of Alexandria, who retains a primacy of honour in Ethiopia, the patriarch of Ethiopia is fully independent of the Coptic Church.

The Ethiopian Church shares the general characteristics of the Eastern Churches. Lay officials called *dabtaras* play a leading part in church affairs and are expert in theology. The priesthood on the whole is not learned. Secular priests may be married; regular priests, belonging to monasteries, may not. Monasticism is widespread, individual monasteries often concentrating on the teaching of special subjects in theology or church music, and pupils therefore traveling from one to the other. There are theological seminaries in Addis Ababa and Harar, which sometimes send students on to Cairo, Athens or Halki. Churches may be round or rectangular, more frequently the former. Round churches, which have four doorways facing the four cardinal points, consist of three concentric circles, the innermost containing the sanctuary where the altar stands, and where the ark, a small wooden coffer representing the ark of the covenant, is kept. Here the priests and the emperor receive the Eucharist. Outside the sanctuary is the "holy place" where the congregation may receive, and the outermost circle is called the choir. Each church has its own church school, generally within the outer wall of the church green. Until about 1900, church schools were the sole source of Ethiopian education.

The church retains the liturgy and the Scriptures in Ge'ez, though a translation of the liturgy into Amharic (the language of the country since the 10th century) has been made. In 1960 the church authorized a new revised version of the Scriptures in Amharic. The Ethiopian canon of Scripture contains books not included in that of other churches, among them the Shepherd of Hermas, the Ascension of Isaiah and the Book of Enoch.

The Ethiopian Church holds that Christ has one nature, but insists that he is perfectly human as well as perfectly divine. Several customs of Jewish origin are still observed: circumcision is universally practised, the Sabbath is kept, the ark is an essential item in every church and a focal point of outdoor festivals on feast days, and a redemption ceremony including the sacrifice of a goat or a lamb is sometimes performed for sick people.

BIBLIOGRAPHY.—J. B. Coulbeau, *Historie politique et religieuse d'Abyssinie,* 3 vol. (1929); E. Cerulli, *Studi Etiopici,* 4 vol. (1936–51); S. Trimingham, *Islam in Ethiopia* (1952); D. O'Hanlon, *Features of the Abyssinian Church* (1946); S. Pankhurst, *Ethiopia* (1955). *See* also bibliography to ETHIOPIA.

(W. D. O'H.)

ETHIOPIAN LITERATURE. The literature of Ethiopia (Abyssinia) consists of (1) writings in the classical Ge'ez (Ethiopic), constituting very largely a literature of translations from Greek and Arabic; and (2) works in Amharic, the modern official language.

Beginnings of Ethiopic Literature.—The ancient Ge'ez language, usually called Ethiopic in the western world, was used by Semitic immigrants who had at an uncertain date crossed from southern Arabia to settle on the African mainland. By the 1st century B.C. they had consolidated into a kingdom centred upon Aksum in the north of present-day Ethiopia. From the early years of the Christian era there survive a few inscriptions and some legends on coins; of the former, some are in a Sabaean script, written right to left, others in unvocalized Ge'ez letters: the fully vocalized alphabet, written from left to right, was probably established about the time of the introduction of Christianity (340). Greek also was employed officially in early times. Moreover, Christianity was introduced by Greek-speaking Syrians, and the earliest extant literary works in Ge'ez are translations from Greek, which may have influenced their style and syntax.

No manuscripts of any Ethiopic text are known from before the 14th century, but the oldest compositions are of a much earlier period. Thus the translation of the Bible was probably begun by the "nine saints"—Syrian monks who arrived in Ethiopia toward the end of the 5th century. The original used for the Old Testament appears to have been the Septuagint in a recension of Antioch, though later revisers (in the 14th century and again after 1500) seem to have used also the Alexandrian version, and possibly the Hebrew masoretic text. The Ethiopic Bible contains, besides the canonical books of the Old and New Testaments and the usual Apocrypha, certain other books, also translated from the Greek. The most notable are *Jubilees* and *Enoch*, the latter being important since no complete version in any language was known until manuscripts brought back from Ethiopia by the traveler James Bruce were translated by R. Laurence (1821).

The Golden Age of Ge'ez.—From the 7th century to the 13th, a period marked by political disturbances, there was no new literary activity, but with the restoration of the "Solomonian" dynasty, *c.* 1270, there began the most productive era of Ge'ez literature, again characterized by translation, not from Greek but from Arabic, though the originals were frequently Coptic, Syriac or Greek. The subject matter is mostly theological or strongly flavoured by religious considerations. The most interesting work occurring at the beginning of this period is *Kebra nagast* ("Glory of Kings"), a combination of mythical history, allegory and apocalypse the central theme of which is the visit of the queen of Sheba to Solomon and the birth of a son, Menelik, who became the founder of the Ethiopian dynasty; it was doubtless written to justify and glorify the newly restored royal house though it probably enshrines previously existing traditions and legends (*see* ETHIOPIA: *History*). Although apparently an original work by an Ethiopian, an ecclesiastic of Aksum claims in a colophon that he translated it from an Arabic version of a Coptic original. This may be a literary fiction but it gains some support from Arabisms in the Ge'ez text. The work is, however, not extant in any other language.

It is not known when the canon of the Mass, with its numerous anaphoras, and the other liturgical books of the Ethiopian Church (*q.v.*) were produced. Some at least were probably translated from Greek during the early years of Ethiopian Christianity but in their extant form they appear to date from the 13th or 14th century. Of the same period are two other ecclesiastical books highly valued in Ethiopia though of little literary interest: *Sinodos* and the *Didascalia*. They are collections of canons, constitutions and statutes translated from Arabic versions of the Greek originals. *Sinodos* is reckoned as a canonical book of the New Testament, some of its ordinances being attributed to Christ himself. Two semiliturgical works are ascribed to Abba Giyorgis of Gasechcha (fl. 1314–44), namely *Sa'atat* ("Book of Hours") and *Sword of the Trinity*, a book of daily devotions. A more important name of this century is Abba Salama, an Egyptian Copt who became metropolitan of Ethiopia in 1350. Not only was he responsible for a revision of the text of the Bible but he translated, or induced others to translate, several books popular among the Ethiopian faithful. The rhapsodical *Weddase Maryam* ("Praise of Mary"), probably a Syriac work of St. Ephraem Syrus, is appended to the Psalter and thus has almost canonical status. Another Syrian author, Philoxenus of Mabbog, is represented by *Filkesyus* (a corruption of his name), questions and answers on monastic life. *Gebra Hemamat* is a lectionary for Holy Week; *Gadla Sama'etat* and *Gadla Hawaryat* ("Acts of the Martyrs" and apocryphal "Acts of the Apostles") treat a favourite subject of Ethiopian devotional literature—a holy life culminating in a painful martyrdom. To a slightly later period, about the beginning of the 15th century, belong various separate lives of saints and martyrs, including St. George (the patron saint of Ethiopia). Moreover, the translation of the Arabic *Synaxarium* was undertaken at this time. This contains lives of saints—one or more—for every day in the year. Although the Ge'ez version dates from this time, most of the manuscripts of the *Senkesar* include two original features: five-line rhymed verses at the end of all the more important lives, each beginning *Salam* (salutation) and summarizing the virtues of the saint; and lives of several native holy persons (including some 15th- and 16th-century kings), which shows that at least some of these accretions belong to a later date.

The early 15th century also saw the translation of several apocalyptic books, notably the *Apocalypse of Peter* and the *Testament of Our Lord Jesus Christ*, which inspired two original compositions. *Fekkare Iyasus* ("Elucidation of Jesus") was written during the reign of Theodore I (1411–14); its object was partly political, as it foretells the advent of a ruler named Theodore who would bring peace and prosperity to his people. *Mystery of Heaven and Earth* was written by Ba-Hayla-Mika'el probably somewhat later; its contents were dismissed by the 17th-century scholar Job Ludolf as *futiles et absurdissimas narrationes*, but in 1926 J. M. Harden described its vigorous account of the struggle between Michael and Satan as "Miltonic." This book must not be confused with another original work of the same period, the *Book of Mystery* by Giyorgis of Sagla, a refutation of heresies composed as a result of theological disputations with an unidentified European. Heresy was also attacked by King Zar'a-Ya'qob (1434–68) in his *Mas'hafa Berhan* ("Book of Light"), which is concerned with improper practices then corrupting the Ethiopian Church. Magico-religious prayers and incantations catered for a depraved popular taste; many are still current, the most considerable being *Lefafa Sedeq* ("Bandlet of Righteousness"), little more than a catalogue of magical names. To counteract such nonsense, Zar'a-Ya'qob encouraged the writing of devotional manuals and hymns; he himself is credited with *Arganona Maryam* ("Organ of Mary") and *Egzi' abher nagsa* ("The Lord Reigneth"), a collection of hymns for each day of the year.

The large hymnals and antiphonaries called *Deggua*, *Mawase'et* and *Me'raf* also probably date from this time, though some of the anthems may be older; the attribution of the *Deggua* to Yared (6th century) is probably due to a confusion: the word also means "ecclesiastical music" in general, and Yared seems to have been an inventor of musical modes, not a writer. Another type of religious poetry first composed during the 15th century was the *malk'e* ("likeness"), consisting generally of about 50 five-line rhyming stanzas, each addressed to a different physical or moral attribute of the saint apostrophized. A *Malk'a Maryam* is attributed to King Na'od (1494–1508), who also wrote some hymns in another verse form called *sellase*. As a last example of the religious literature of the "golden age" may be mentioned the *Miracles of Mary*, translated from Arabic in 1441–42; enormously popular, it is found in several recensions.

Secular literature of the period is limited to history. Most important are the indigenous royal chronicles, the first of which, probably contemporary, relates to the wars of Amda-S'eyon I. There is then a gap until Zar'a-Ya'qob, but thenceforward the series is continuous, though no author's name is recorded in this period.

The Silver Age and Decline of Ge'ez.—In 1527 began the Muslim incursions directed by Ahmad (Mohammed) Gran, and for 15 years Christian Ethiopia was at the mercy of the marauders. Literary activity ceased, and many manuscripts perished. In 1543 the invaders were repulsed but the country never fully recovered. Many Christians had embraced Islam and the first efforts of the church were directed toward their reconversion. A useful agent

was a Muslim merchant who had been converted to Christianity and eventually, as Enbaqom (Habakkuk), became prior of the monastery of Dabra Libanos. He wrote *Anqas'a amin* ("Gate of Faith"), a justification of his conversion, to persuade apostates to recant. Other similar works were *Mas'hafa qeder* ("Book of Impurity") and *Mas'hafa nesseha* ("Book of Penitence"), the latter being translated from Arabic by an Ethiopian monk in Egypt. A greater danger to the Ethiopian Church was the advent of Roman Catholic missionaries. Several works were written defending the Monophysite faith including a brief *Confession* by the king himself (Galawdewos or Claudius, 1540–59) and another original work called *S'awana nafs* ("Refuge of the Soul"); while from Arabic were translated *Fekkare malakot* ("Elucidation of Godhead") and *Haymanota abaw* ("Faith of the Fathers"). The latter, a collection of extracts and treatises largely on Monophysite doctrine, was translated by Maba S'eyon, son of a nobleman at the court of Galawdewos.

Ge'ez had by now lost its vigour; it had long ceased to be a vernacular (exactly when is not known) and, while it remained the liturgical language, few lay people were thoroughly conversant with it. During the 16th century Amharic, the principal spoken language, was already beginning to be used for literary purposes; even the royal chronicles gradually became interlarded with Amharic words and expressions, the jargon being called *lesana tarik*, "language of history." However, about 1600 a few substantial works in Ge'ez appeared: *Hawi*, an enormous theological encyclopaedia, translated by Salik of Dabra Libanos; the *History* of Johannes Madabbar, bishop of Nikiu, which contains a valuable account of the Arab conquest of Egypt and is important since the Arabic original is lost; the *Book of the Wise Philosophers*, an anthology of aphorisms from such diverse sources as David and Solomon, Plato and Aristotle, Cicero and Chosroes of Persia. Of practical value was *Fetha nagast* ("Legislation of the Kings"), a compilation of canon and civil law made in Egypt in the 13th century and for more than 300 years the only legal code in Ethiopia. Mention should also be made of *Hatata Zar'a-Ya'qob* ("Inquiries of Zar'a-Ya'qob"), a philosophical work purporting to date from the 17th century but now known to have been composed by Giusto da Urbino, an Italian Roman Catholic missionary, about 1850. Ge'ez poetry (*qene*) in various formal styles flourished particularly at Gondar in the 18th century, and continues to be practised at many monasteries, though comparatively little is preserved. Some poems of Alaqa Tayya were printed at Asmara in 1921, and an important anthology was compiled by Heruy Walda Sellāsē and published at Addis Ababa in 1926.

Falasha Literature.—The Jewish population, known as Falashas (*q.v.*), who live mostly in regions north of Lake Tana, use Ge'ez as their sacred language. Besides the Old Testament (including *Jubilees*) they have a few books peculiar to themselves, notably *Te'ezaza Sanbat* ("Ordinance of the Sabbath") and a homily called *Abba Elijah*. The source and date of these and of other Falasha books are uncertain; perhaps most are translations from Arabic of the 14th century. A *Falasha Anthology* was published by Wolf Leslau in 1951.

Amharic Literature.—The only early known Amharic compositions are songs celebrating the victories of Amda-S'eyon I (1314–44). From the 16th century onward theological works, mostly commentaries on Ge'ez texts, were produced. A translation of the Bible was made in Cairo early in the 19th century by "Abu Rumi Habessinus"; neither his name nor the quality of his Amharic suggests a true Ethiopian. This version formed the basis of editions issued by missionary societies, revisions being made by foreigners with an inadequate knowledge of Amharic. A more scholarly version of the New Testament was printed in Addis Ababa in 1955, to be followed by the Old Testament. Of the early 19th century also is a medical encyclopaedia. The first official chronicles wholly in Amharic were those of Theodore II (1855–68); there are two versions, by Alaqa Walda Maryam and Alaqa Zanab. Zanab also wrote a short didactic work, *Talks on Body and Soul* (1932, reissued 1952).

A translation of Bunyan's *Pilgrim's Progress* by Gabra Giyorgis Terfe (1892) pointed the way to a popular form of original composition—the allegorical novel, often partly in verse, with a religious bias. The first was *Libb wallad tarik* ("Imaginative Story") by Afawarq Gabra-Iyasus (1908). During the regency of Ras Tafari (1916–30; afterward Emperor Haile Selassie) Blatengeta Heruy Walda-Sellase (d. 1938) became the leading Amharic writer, especially notable for allegorical compositions such as *Wadaje lebbe* ("My Heart as My Friend") and *Addis alam* ("A New World"), though he also wrote biographical and other works.

With the restoration of Ethiopian independence after the Italian occupation of 1936–41, a great impetus was given to Amharic literature, the emperor encouraging authors to produce many types of books, especially on moral and patriotic themes. Among writers of merit may be cited Bitwadded Makonnen Endalkachaw (allegorical novels and plays), Kabbada Mika'el (verse dramas, some history and biography) and Takla-S'adeq Makurya (histories). Notable translations include Samuel Johnson's *Rasselas* by Sirak Heruy and *Herodotus* by Mars'-Hazan Walda Quirqos.

BIBLIOGRAPHY.—The standard work, brief but comprehensive, is I. Guidi, *Storia della letteratura etiopica* (1932), which gives details of editions and translations of the principal Ge'ez texts; some later publications are listed in E. Cerulli, *Storia della letteratura etiopica* (1956). *See* also J. M. Harden, *Introduction to Ethiopic Christian Literature* (1926); E. Littmann, "Die äthiopische Literatur" in *Handbuch der Orientalistik* (1954). English translations of many important texts were issued by E. A. Wallis Budge, largely for popular consumption; they often show marks of somewhat hasty work. Of the Ethiopic liturgy a new translation of the current use was made by Marcos Daoud with Ethiopian assistance (1954); *see* also J. M. Harden, *Anaphoras of the Ethiopic Liturgy* (1929). For Amharic literature *see* L. Fusella, "Recenti pubblicazioni amariche in Abissinia," *Rassegna di studi etiopici*, vol. 5 (1947); *Guide-book of Ethiopia*, pp. 407–16 (1954); S. G. Wright in D. A. Talbot, *Haile Selassie I: Silver Jubilee* (1955).

(S. G. WT.)

ETHIOPIC LANGUAGE: *see* SEMITIC LANGUAGES.

ETHNOLOGY, SOCIETIES OF: *see* ANTHROPOLOGY AND ETHNOLOGY, SOCIETIES OF.

ETHNOLOGY AND ETHNOGRAPHY, originally sciences concerned with the history and distribution of human races (*see* RACES OF MANKIND), in modern usage (in English-speaking countries) are generally restricted to the study of culture. Ethnography refers to the descriptive studies of customs of particular tribes or peoples of particular regions, whereas ethnology has a more theoretical orientation. A further differentiation, developed in England, restricts ethnology to a concern with historical problems of cultural development and utilizes social anthropology for comparative and generalizing studies of society and culture, but U.S. anthropologists prefer to use ethnology as a general term for all these activities.

As branches of anthropology, ethnology and ethnography have close relations with archaeology, linguistics and physical anthropology, and with geography and history as well.

See ANTHROPOLOGY: *Ethnology;* SOCIAL ANTHROPOLOGY; *see* also references under "Ethnology and Ethnography" in the Index volume.

(F. R. E.)

ETHOLOGY, a term applied by European zoologists to that aspect of the study of animal behaviour concerned with as complete a description as possible of the objective behaviour of an organism in its natural environment. Ethologists emphasize the unlearned aspects of behaviour and have been responsible for renewed interest in the concept of instinct. The study has centred mostly upon the experimental analysis of behaviour in lower animal forms. *See* ANIMAL BEHAVIOUR; BEHAVIOURISM; INSTINCT.

(ED. H. H.)

ETHYL, in chemistry, the alkyl (aliphatic hydrocarbon) radical C_2H_5. (*See* CHEMISTRY: *Organic Chemistry.*) The hydride of ethyl is better known as ethane; the alcohol C_2H_5OH is the ethyl alcohol of commerce (spirit of wine) (*see* ALCOHOL); and the oxide $(C_2H_5)_2O$ is ordinary ether (*q.v.*). Ethyl is also a trademark denoting a fluid (the active ingredient is lead tetraethyl) mixed with motor fuel to eliminate knocking in internal-combustion engines (*q.v.*).

See ORGANOMETALLIC COMPOUNDS.

ETHYL ALCOHOL (ETHANOL) is the most important member of the class of organic compounds which are given the general name of alcohols. Comparison of its molecular formula

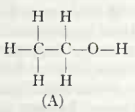
(C_2H_5OH) with that of the saturated hydrocarbon, ethane (C_2H_6), suggests that ethyl alcohol may be regarded as a hydroxyl derivative of ethane in which one of the six hydrogen atoms of the ethane molecule has been replaced by a hydroxyl (—OH) function. It is this relationship between ethane and ethyl alcohol which makes the systematic name, ethanol, widely acceptable.

History.—Ethanol must have been one of the earliest organic compounds to have been investigated. It is the intoxicating ingredient of many beverages whose production involves a fermentation process, and the word alcohol is often used to describe the ethyl alcohol content of such beverages. It is often stated that ethyl alcohol was known first by the Arabs, but this is believed to be incorrect and it is more likely that its discovery occurred in Italy.

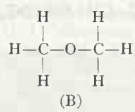
Despite the Arabic origin of the word (*see* ALCOHOL), the separation of ethyl alcohol from wine was apparently not known to the Arabian alchemists. According to Edmund O. von Lippmann, their distillation apparatus did not cause the condensation of low-boiling-point liquids (the boiling point of pure ethanol is 78.32° C.). This point may well be questioned, but P. E. M. Berthelot and von Lippmann both note that the earliest known description of the concentration of ethyl alcohol by distillation of wine occurs in a Latin manuscript of the 12th century, Mappae Clavicula. Since a 10th-century version of this manuscript contains no reference to this process, it is assumed that ethyl alcohol was first separated by distillation around the 11th century, probably in the wine districts of Italy.

Although ethyl alcohol had been known for a considerable time in a crude state in wines, mead and beer, it was not until 1796 that Johann Tobias Lowitz first prepared it free from water. Nicolas Théodore de Saussure determined its composition in 1808, although A. L. Lavoisier had earlier established qualitatively that it contained carbon, hydrogen and oxygen.

The molecular formula of ethanol was thus shown to be C_2H_6O and the major credit for the establishment of its structural formula belongs to Sir Edward Frankland, although its correct structural formula was first proposed by Auguste Laurent in 1846. Frankland's investigations form one of the cornerstones of modern structural theory, and the interpretation of his arguments in terms of current valency language is as follows. Accepting the valency of carbon as four, oxygen as two and hydrogen as one, there are two possible isomeric formulas which may be written:



(A)



(B)

The choice between (A) and (B) for ethanol rests upon the following evidence. Reaction of ethanol and sodium metal gives sodium ethoxide, $2C_2H_6O + 2Na \rightarrow 2C_2H_5NaO + H_2$. Ethanol and hydrogen bromide give ethyl bromide, $C_2H_6O + HBr \rightarrow C_2H_5Br + H_2O$. These experiments demonstrate that one of the hydrogen atoms of the ethanol molecule is differently placed from the other five, and the unique character of one of the hydrogen atoms is compatible only with formula (A). Formula (B) is that of an entirely different compound, dimethyl ether. The chemical reactions of ethyl alcohol discussed in detail below are all easily interpretable in terms of (A).

Manufacture.—There are three main processes used for the manufacture of ethanol: (1) the fermentation of carbohydrates; (2) the hydration of ethylene; and (3) the reduction of acetaldehyde prepared by the hydration of acetylene. Prior to 1930 all the ethanol used for industrial purposes was prepared by fermentation processes, and this is still the most important method in most countries. However, in the early 1960s in the United States 70%–75% of the total output of ethanol was produced by hydration of ethylene. The importance of ethanol in the chemical industry is shown by production statistics at mid-20th century which showed an annual output of between approximately 400,000,000 and 500,000,000 proof gallons in the United States, and of between 50,000,000 and 60,000,000 proof gallons in the United Kingdom.

Fermentation of Carbohydrates.—The fermentation process involves the transformation of carbohydrates into ethanol by growing yeast cells (*see* FERMENTATION). The fermentation of D-glucose may be represented simply by the over-all equation

$$C_6H_{12}O_6 \longrightarrow 2C_2H_5OH + 2CO_2 + 50{,}000 \text{ cal.}$$

But it is now recognized that this equation is a gross oversimplification and that 11 steps are involved in going from glucose to ethanol. They are: D-glucose → glucose-6-phosphate → fructose-6-phosphate → fructose-1:6-diphosphate → glyceraldehyde-3-phosphate → 1:3-diphosphoglyceric acid → 3-phosphoglyceric acid → 2-phosphoglyceric acid → phosphoenolpyruvic acid → pyruvic acid → acetaldehyde + CO_2 → ethyl alcohol. Most of the steps in this scheme are enzymatically controlled and the appropriate coenzymes for some of the steps such as adenosine triphosphate or diphosphopyridine nucleotide are known.

Fermentation of a few carbohydrates other than D-glucose also yields ethanol. The fermentable sugars are D-mannose, D-galactose, D-fructose and the disaccharide maltose; other disaccharides or polysaccharides must be converted either by acidic or enzymatic hydrolysis into fermentable sugars. The chief raw material for the manufacture of ethanol by the fermentation process is molasses, which is the noncrystallizable residue from the refining of sugar by crystallization. The molasses, which contains about 50% sucrose, is diluted with water so that the sugar concentration is 14%–18%.

The pH is then adjusted to five by the addition of sulfuric acid, which prevents the growth of harmful microorganisms. Yeast is added and the fermentation proceeds until the ethanol concentration rises to 14%, when fermentation ceases. The yeast provides the enzyme invertase capable of converting sucrose into glucose and fructose:

$$\underset{\text{Sucrose}}{C_{12}H_{22}O_{11}} + H_2O \xrightarrow{\text{invertase}} \underset{\text{Glucose}}{C_6H_{12}O_6} + \underset{\text{Fructose}}{C_6H_{12}O_6}$$

The glucose and fructose yield ethanol by the process detailed above.

The magnitude of the operations associated with the fermentation of molasses on an industrial scale is indicated by the size of the fermenters, which are of 500,000 gal. capacity, and the molasses storage tanks, which are of 6,000,000 gal. capacity. The fermenters are totally enclosed both to prevent loss of ethanol and to collect the carbon dioxide for use as a refrigerant.

Industrial ethanol is also prepared from potatoes. These are first heated with superheated steam, mashed with malt and water and then fermented by the addition of yeast. Crude spirit results from the distillation of the fermented mash.

Hydration of Ethylene.—This is now the most important manufacturing method, although it was not developed as an industrial process until 1929. It is interesting to note that the hydration of ethylene to ethyl alcohol was first reported by Henry Hennel a century earlier in 1828.

The hydration of ethylene, $C_2H_4 + H_2O \rightarrow C_2H_5OH$, is achieved by passing a mixture of ethylene and a large excess of steam, at a temperature of 300° C. and a pressure of several thousand pounds per square inch, over a solid catalyst. The catalyst normally used is phosphoric acid supported on an inert material, but in the 1950s a silicophosphoric acid was found to give higher yields. The ethylene used in this process is obtained during the cracking of petroleum oil for the production of gasoline (*see* PETROLEUM).

In an older process for the hydration of ethylene, it was absorbed into concentrated sulfuric acid forming ethyl hydrogen sulfate, and this was hydrolyzed by adding water and heating.

$$C_2H_4 + H_2SO_4 \longrightarrow C_2H_5O.SO_3H$$
$$C_2H_5O.SO_3H + H_2O \longrightarrow C_2H_5OH + H_2SO_4$$

Reduction of Acetaldehyde.—Some ethanol has been prepared by the reduction of acetaldehyde, which is itself prepared by the hydration of acetylene (*see* ACETIC ACID).

$$C_2H_2 + H_2O \longrightarrow CH_3.CHO$$
$$CH_3CHO + H_2O \longrightarrow C_2H_5OH$$

However, economically this process is not as satisfactory as the direct hydration of ethylene described above.

Manufacture of Pure Absolute Ethanol.—Ethanol produced either by fermentation or by synthesis is obtained as a dilute aqueous solution and must be concentrated by fractional distillation. Direct distillation can yield at best only the constant-boiling-point mixture containing 95.6% by weight of ethanol, since the boiling point of the constant-boiling-point mixture is slightly below that of pure ethanol.

Dehydration of the constant-boiling-point mixture yields anhydrous or absolute ethanol. In the laboratory absolute ethanol is obtained from rectified spirit by drying over calcium oxide, calcium or magnesium. On an industrial scale, azeotropic distillation is used to remove the water by codistillation with added benzene or trichloroethylene. These substances form ternary azeotropes of lower boiling point than that of the ethanol-water constant-boiling-point azeotrope, so that the water is removed preferentially leaving anhydrous ethanol.

Properties.—Pure ethanol is a colourless liquid (boiling point, 78.32° C. at a pressure of 760 mm.) with an agreeable ethereal odour and a burning taste. It solidifies at −114.5° C. and may therefore be used as a low-temperature thermometric liquid. It is very hygroscopic and is miscible with water in all proportions. When ethanol is mixed with water, heat is evolved and a contraction in volume takes place. Thus one volume of water plus one volume of ethanol gives 1.93 vol. of solution at the same temperature. Ethanol forms with water a constant-boiling-point (azeotropic) mixture which boils at 78.2° C. Benzene, ethanol and water form a ternary azeotrope of boiling point 64.85° C.

Ethanol burns in air with a nonluminous flame producing carbon dioxide and water. Its heat of combustion is 7.09 kg.cal. per gram as compared with 11.5 kg.cal. per gram for gasoline. Mixtures of ethanol and air are explosive.

The chemical properties of ethanol are typical of primary alcohols. Thus oxidation with potassium dichromate and sulfuric acid gives acetaldehyde, which is then transformed into acetic acid, $CH_3.CH_2OH \rightarrow CH_3.CHO \rightarrow CH_3.CO_2H$. Ethanol reacts with metals such as sodium, potassium, magnesium or aluminum giving ethoxides. With concentrated sulfuric acid it gives ethyl hydrogen sulfate, $C_2H_5.O.SO_3H$, and diethyl sulfate, $(C_2H_5)_2SO_4$. When ethanol is heated with excess concentrated sulfuric acid at 170° C., ethylene, $CH_2=CH_2$, is formed, but when concentrated sulfuric acid is treated with excess ethanol at 140° C., the product is diethyl ether, $(C_2H_5)_2O$. The reaction of ethanol with hydrogen chloride and zinc chloride gives ethyl chloride; hydrogen bromide and hydrogen iodide give ethyl bromide and ethyl iodide respectively. When ethyl alcohol is heated with carboxylic acids in the presence of hydrogen chloride or a few drops of concentrated sulfuric acid, ethyl esters are formed, $R.CO_2H + C_2H_5OH \rightarrow R.CO_2C_2H_5 + H_2O$ (R representing a hydrocarbon radical). Ethanol reacts with bleaching powder giving chloroform, $CHCl_3$, and with iodine and potassium hydroxide giving iodoform, CHI_3. Mercury fulminate, $Hg(CNO)_2$, was prepared by Baron Justus von Liebig (1823) by treating ethanol with nitric acid and mercury, and essentially the same process is still used industrially.

Ethanol is toxic to the animal organism and although considerable differences of opinion have been expressed, it would appear that the rapidly lethal dose is of the order of 10 ml. per kilogram of body weight. During 90 min. after taking it, the concentration of ethanol in the blood rises and then gradually decreases until it is normal after 12–15 hours. Ethanol is converted in the body into carbon dioxide and water via acetaldehyde.

Detection and Estimation.—Ethyl alcohol may be detected by oxidation to acetaldehyde by means of chromic acid, and testing for acetaldehyde with Schiff's reagent. It may also be detected by warming the sample with iodine and potassium hydroxide, so producing iodoform, which is recognizable by its characteristically rosette-shaped crystals and its melting point, 119° C. The two tests are not specific, but ethanol may be positively identified as a crystalline derivative such as the 3:5-dinitrobenzoate.

The percentage of ethanol present in a mixture of pure water and ethanol is readily established by determining its specific gravity and then referring to tables specially prepared for the purpose. This process is called determination by alcoholometry. If other substances are present, as in the case of wine or beer, it is necessary to separate the ethanol by distillation. For example, in order to determine the percentage of ethanol in beer, a 200-ml. sample of the beer can be distilled until 100 ml. are collected and then the density of the distillate determined by means of a pyknometer or hydrometer. The percentage of ethanol found in the distillate divided by two is the alcoholic strength of the beer.

The standard tables relating the specific gravity of aqueous alcoholic solutions to the percentage of ethyl alcohol present are expressed in various terms depending upon the requirements of the government or other authority by whom the tables were prepared. The usual tables give the percentage of absolute ethyl alcohol by weight (degrees according to Richter) or by volume (degrees according to Tralles). Arbitrary standards are in general usage in Great Britain, the United States and in other countries. U.S. proof spirit contains 50% ethanol by volume; British proof spirit contains 49.3% ethanol by weight and 57% by volume.

Anhydrous ethanol suspected of containing water may be tested by the addition of four volumes of gasoline or kerosene. The mixture becomes turbid if the concentration of water exceeds 1%. Absolute ethanol should give no turbidity on the addition of anhydrous benzene, and no pink coloration should be produced when it is shaken with potassium permanganate crystals. A sensitive test for water in ethanol is due to C. Claus (1877). Anhydrous ethanol, a trace of anthroquinone and sodium amalgam give a green colour, but if a trace of water is present, a red colour is produced. A precise method for determining even minute traces of water is the Karl Fischer method, which involves the direct titration of the ethyl alcohol with a methanolic solution of iodine, sulfur dioxide and pyridine; the reagent is self-indicating.

See also references under "Ethyl Alcohol" in the Index volume.

BIBLIOGRAPHY.—J. F. Thorpe and M. A. Whitely, *Thorpe's Dictionary of Applied Chemistry*, 4th ed., vol. 1, p. 175 (1937); I. Mellan, *Industrial Solvents*, 2nd ed. (1950); L. A. Underkofler and R. J. Hickey (eds.), *Industrial Fermentations* (1954); A. K. Doolittle, *The Technology of Solvents and Plasticizers* (1954); R. S. Aries, "Alcohol, Industrial" in R. E. Kirk and D. F. Othmer (eds.), *Encyclopaedia of Chemical Technology*, vol. 1 (1948–57); E. H. Rodd (ed.), *Chemistry of Carbon Compounds*, vol. 1a, p. 285 (1951). (W. D. Os.)

ETHYL CHLORIDE, a compound that finds use in medicine as an anesthetic. It is prepared from hydrogen chloride and ethyl alcohol. Although it is a gas at room temperature, ethyl chloride is stored and sold under pressure in glass bottles fitted with trigger-controlled spray nozzles, as a colourless, mobile liquid with a characteristic ethereal smell and sweet, burning taste.

Uses.—It has been used (1) as a local anesthetic for small incisions, tooth extractions and needle punctures, by spraying it on the surface of the skin or mucous membrane in a fine stream with particular care that the end of the spray stream is focused upon the point to be frozen. Its rapid evaporation causes local freezing (−35° C. can be obtained).

(2) As a general anesthetic, it is usually given in doses of three to five cubic centimetres in a closed inhaler. Short administrations produce an anesthesia similar in type to that of nitrous oxide with an equally rapid loss of consciousness, a quick recovery and slight aftereffects. It is suitable only for short operations such as extraction of teeth or opening abscesses. Ethyl chloride like chloroform as a general anesthetic has its popularity chiefly in Europe. It does not produce the muscular relaxation effected by chloroform, which is desirable especially in protracted operations. Moreover, it is inflammable. Ethyl chloride has been recommended for anesthesia of short duration in the belief that it was less dangerous than and as satisfactory as nitrous oxide. Its effect on the circulatory and respiratory systems is slightly stimulating, causing flushing of the face, acceleration of the pulse rate by 10–20 beats per minute, and a rise in systolic blood pressure of from 10 to 20 mm., together with deeper and more rapid breathing. (*See* ANESTHESIA AND ANESTHETICS.)

(3) It also is used in organic chemical synthesis as an ethylating

agent; *i.e.*, as a reagent for the introduction of the ethyl group, C_2H_5, into compounds of various types.

Properties and Commercial Preparations.—Ethyl chloride boils at 12.5° C. and has the formula C_2H_5Cl. The vapour burns with a green flame. Chloryl and Kelene are trade names. Anestile is a mixture of ethyl and methyl chlorides. Somnoform is composed of the two chlorides with ethyl bromide. (A. B. L.; X.)

ETHYLENE (Ethene), one of the most important raw materials in the organic chemical industry, is a colourless gas of sweetish taste and odour; it occurs in both petroleum and natural gas. Ethylene undergoes polymerization to polyethylene, which is a plastic material of great commercial use, particularly in the form of packaging films, wire coverings and "squeeze" bottles. The polymerization is carried out at elevated pressure (1,000 atm.) and temperature in the presence of minor amounts of oxygen or peroxides, or at low pressures and temperatures in the presence of catalysts.

Ethylene gas is used in anesthesia. (*See* ANESTHESIA AND ANESTHETICS.)

Ethylene is capable of undergoing reactions characteristic of an olefin (*see* OLEFIN: *Monoolefins*). It reacts with water in the presence of sulfuric acid to form ethanol (ethyl alcohol, *q.v.*); it combines with chlorine and bromine to yield the corresponding ethylene dihalides. It reacts with oxygen in the presence of a silver catalyst to form ethylene oxide, which is used in the manufacture of antifreeze, detergents, paints and plastics. Ethylene reacts with isobutane in the presence of aluminum chloride to form 2,3-dimethylbutane (component of gasoline). It combines with benzene in the presence of catalysts to form ethylbenzene, which is dehydrogenated to styrene for use in the manufacture of plastics and synthetic rubber.

Pure ethylene is best prepared by passing vapours of ethyl alcohol over alumina pellets heated at 420° C. Commercially ethylene is produced by the pyrolysis of ethane and/or propane.

Ethylene, whose formula is C_2H_4, or $CH_2{=}CH_2$, boils at −103.9° C. and freezes at −169.4° C. Its critical temperature is 9.6° and its critical pressure 50.7 atm. *See* also references under "Ethylene" in the Index volume. (H. Ps.; X.)

ETIQUETTE may be defined as a system of conventional rules controlling social or professional behaviour.

The taboos of savage tribes provided both the etiquette and the law by which the tribe lived. In more complex societies, constitutionally accepted rules of social behaviour are upheld and enforced by legal codes, while within particular classes there will also be found unwritten codes, binding upon their own members. To enforce conformity to these codes there are no laws; an offender faces no formal trial or sentence, the penalty lying in the disapproval of other members of his group. Etiquette in its true sense is only possible in a highly stratified society, where every man knows the behaviour expected from him toward others and from others toward himself.

Courts are the natural homes of etiquette, because at the centre there is a monarch around whom niceties of behaviour spread in expanding circles. The author of *Beowulf*, writing of Anglo-Saxon society, describes Wealtheow the queen, "mindful of etiquette," carrying the goblet first to the king, then to the courtiers, in a clearly defined order of precedence.

The middle ages were a golden period for etiquette; the feudal system was strictly stratified; *e.g.*, Jean Froissart in his *Chronicle* speaks of the Black Prince waiting at table upon the captive King John of France, after the battle of Poitiers. The late 18th and early 19th centuries showed another great flowering of etiquette in Britain when exquisites like Beau Nash and Beau Brummell imposed their whims as rules upon polite society; even the prince regent would not leave his waistcoat unbuttoned to a greater degree than Brummell prescribed. In the late 19th and early 20th centuries those in the upper strata of society regarded the observance of the most trivial demands of etiquette as at once a diversion and, for the women, an occupation. More and more elaborate rituals were designed to create a delightful sense of exclusiveness for the initiates and to keep the unworthy, ignorant of them, at a distance.

World wars and increasing social equality have resulted in a simpler code, appropriate to the faster tempo and less pampered conditions of life in society generally. Yet etiquette remains active on royal or ceremonial occasions and in the more formal aspects of professional or communal life. No rule of law or principle of morality decrees that a soup plate should be tilted away from, never toward, the diner, or that (in Great Britain) a surgeon shall be known as "Mr." while a physician is addressed as "Dr.," but etiquette ordains it and it is obeyed. Etiquette is not necessary to the state but it is so to society. Since the framework and content of the communities of which society is formed are constantly changing, the habits of etiquette must change with them. But etiquette is the formula of good manners, without which there would be no agreeable social relationships. Wherever there is polite society, therefore, some form of etiquette will be found. *See* also ADDRESS, FORMS OF.

BIBLIOGRAPHY.—Emily Post, *Etiquette*, 9th ed. (1955); Millicent Fenwick (ed.), *Vogue Book of Etiquette* (1948); F. B. Fawcett, *Court Ceremonial and the Book of the Court of King George VI* (1937); Ann Page, *Complete Guide to Wedding Etiquette* (1950). (O. K. C.)

ETNA, MOUNT, an active volcano in Catania province on the east coast of Sicily (Gr. AITNE, from *aitho*, "I burn"; Lat. AETNA; Sicilian MONGIBELLO). Its height and crater are subject to constant change. The highest point varies above 10,500 ft. and tends to decrease (10,870 ft. in 1861; 10,758 ft. in 1900; 10,705 ft. in 1950. It is the highest volcano in Europe, and the highest peak in Italy south of the Alps. It covers approximately 500 sq.mi. and has a base circumference of about 90 mi.

Geologically, Etna is a large volcanic cone which stands in a great Pliocene subsidence bay of the Ionian sea. It is in contact with Tertiary sedimentary rocks in the northwest. There occur

STURE JOHANNESSON

MOUNT ETNA, THE CENTRAL CONE

gray Eocene and yellow and red Miocene sandstones and limestones with Pliocene marine marl and plastic clay. These latter entirely compose the floor of the platform upon which the volcanic pile has been built. Near the coast ancient lava flows overlie Pleistocene conglomerates and gravels, thus proving that the subaerial activity did not begin until middle Quaternary times. The volcanic rocks consist of crystalline and vitreous lavas and tuffs, essentially basaltic in composition, largely porphyritic and wholly nonleuctic.

Other types are augite and phonolitic andesites and labradorite rock. The andesites were the earliest erupted rocks. Basic plagioclase feldspar and augite are the most common minerals, olivine being subordinate or missing altogether. The constancy of composition indicates derivation from a remarkably uniform magma. At Acireale, a town at the southeast foot of Etna, the lava has assumed the prismatic or columnar form in a striking manner; at the rock of Aci it is in parts spheroidal. The Grotte des Chèvres has been regarded as an enormous gas bubble in the lava. The remarkable stability of the mountain appears to be due to the innumerable dikes which penetrate the lava flows and tuff beds in all directions and thus bind the whole mass together.

In shape Etna is a truncated cone with the great rift of the Valle del Bove, a sterile abyss 3 mi. wide and 2,000–4,000 ft. deep, on its east side. This latter is the original crater, the volcanic axis having shifted to the present summit crater. There are more than 200 subsidiary cones, some of them more than 3,000 ft. high, which have risen on lateral fissures that converge on the central crater. The mountain is built up of three superimposed parts which correspond to three well defined zones of vegetation—the cultivated, the wooded and the desert. The lowest, *regione coltivata,* extends up to 3,000 ft., has a gentle gradient and is a great belt of volcanic products. It is densely populated and of luxuriant fertility (vine, olive, fruit, vegetables, corn, etc.) and often yields five harvests a year. On it are the towns Catania (*q.v.*), Nicolosi and Acireale. The middle zone, *regione boscosa,* has a gradient of one in five and grows abundant pine, genista broom, chestnut, beech and fern. The highest and central zone, *regione deserta,* is a desolate region of lava flows and volcanic ash, almost destitute of vegetation.

There is a narrow zone of subalpine shrubs, but no true alpine flora. In the last 2,000 ft. only five phanerogamous species are to be found, the first three of which are peculiar to the mountain: *Senecio etnensis, Anthemis etnensis, Robertsia taraxacoides, Tanacetum vulgare* and *Astragalus siculus.* No trace of animal life is to be found in that zone; for the greater part of the year it is covered with snow, but by the end of summer this has almost all melted, except for that preserved in the covered pits for use in Catania and elsewhere.

The ascent of Etna is best undertaken in summer or autumn, seven to eight hours being required from Nicolosi. The easiest route is by the Etna road from Catania through Nicolosi to its terminal at 6,171 ft., a distance of 34 km. (21 mi.). From the end of the road a cableway reaches the building called Torre del Filosofo (9,564 ft.), a Roman ruin believed to commemorate the ascent of Etna by the emperor Hadrian. Above, at 9,666 ft., is the observatory, originally a hut built by British armed forces in 1811, at the base of the terminal cone.

The Greeks believed Etna to be the workshop of Hephaestus and the Cyclops or, as in some texts, a column supporting the sky. Underneath Etna the giant Typhon rested, but whenever he twisted or turned the earth trembled. Hesiod (*Theogony,* 820–68) speaks of eruptions of Etna; Pindar and Aeschylus refer to the famous eruption of 475 B.C. Aetna was, according to Diodorus Siculus, the name given in 476 B.C. to the town of Catana (mod. Catania) by Hieron I, tyrant of Syracuse. Inessa (about 10 mi. N.W. of Catana) was also renamed Aetna by Hieron's colonists who settled there after they were driven out of Catana by the Sicel leader Ducetius (*q.v.*).

There were more than 135 recorded eruptions up to the early 1960s, all but 21 of which took place in the Christian era. It is believed that only the strongest of eruptions that occurred in classical times were recorded, such as that of 396 B.C., which kept the Carthaginian army, commanded by Himilco, from reaching Catana. Among later eruptions that of 1381 reached the sea with its lava flow; but the most violent on record was that of March 11, 1669, when the twin peaks, Monti Rossi, were formed. A crack then appeared in the mountainside, reaching from the crater down to the village of Nicolosi, and the lava flow reached the city of Catania and the sea. There were 16 eruptions during the 18th century and 19 during the 19th. The eruption of 1852 threatened the village of Zafferana Etnea and took a heavy toll of human lives. That of 1928 cut the railway that runs round the base of the mountain and buried the village of Mascali; the series of eruptions between Nov. 1950 and Oct. 1951 threatened the villages of Rinazzo and Fornazzo and are believed to have been the most violent in over 200 years.

BIBLIOGRAPHY.—Sartorius von Waltershausen, *Atlas des Atna* (1880); E. Chaix, *Carta Volcanologica e topographica dell' Etna,* showing lava streams up to 1892; G. de Lorenzo, *L'Etna* (1907); C. S. du Riche Preller, *Italian Mountain Geology,* part iii (1923); Touring Club Italiano, *Guida di Sicilia* (1953). (G. KH.)

ETON, a town and urban district of Buckinghamshire, Eng., on the Thames opposite Windsor, 22 mi. W. of London by road. Pop. (1961) 3,901. It is famous for its college, the largest of the great public boarding schools of England. The "King's College of Our Lady of Eton Beside Windsor" was founded by Henry VI in 1440–41 and endowed mainly from the revenues of the alien priories suppressed in 1414. King Henry followed the model of William of Wykeham's foundations of Winchester and New college, Oxford, and simultaneously with Eton he founded King's college, Cambridge, to which scholars from Eton were to proceed. King Henry's statutes in their final form provided for the establishment of a college consisting of a provost, ten fellows who were to be secular priests, 70 scholars, a schoolmaster and an usher, ten clerks, 16 choristers and 13 poor infirm men. The founder also provided for the education of boys other than scholars, some of whom, among them the sons of noble or powerful persons, were to live in the college, while for the rest inns and houses in the town were reserved. These, under the name of "oppidans," now form the principal body of the boys.

The present constitution of the college was established by new statutes made under the Public Schools act of 1868. Under them the college consists of a provost, appointed by the crown, a vice-provost and ten nonresident fellows, who together form the governing body, a headmaster and lower master and 70 scholars. The body of fellows comprises the vice-provost, the provost of King's college and representatives nominated respectively by the universities of Oxford and Cambridge, the Royal society, the lord chief justice and the masters of the school, with four co-opted fellows. The scholarships are awarded on the results of an annual competitive examination open to all boys between 12 and 14 years of age who are natural-born British subjects. The scholars, now known as collegers, live in the main college buildings and the remainder of the boys, the oppidans, numbering over 1,100, live in boarding-houses in the town. In the past the whole cost of the maintenance and education of the scholars was met from the college endowments, but after 1923 they were charged a fee which is remitted in whole or in part in cases of need. The fee payable by oppidans covers the whole cost of maintenance and education, none of which is met from endowments, except the cost of chapel services. There is a teaching staff of over 100 masters, and the curriculum embraces classical and modern subjects. The connection with King's college is continued by the reservation of 24 scholarships there for Etonians who reach the requisite standard. The college also maintains a choir school for the free education of about 26 choristers and an almshouse in the town.

The construction of the original college buildings was begun in 1441, but they were not completed till 50 or more years later. They consist of two quadrangles. The outer or School yard is enclosed by the chapel, the Upper and Lower schools and the original scholars' dormitory (Long chamber), later transformed. It has in its centre a bronze statue of the royal founder. Part of the Upper school range was destroyed by a German bomb in 1940 and was rebuilt. The buildings enclosing the lesser and inner quadrangle (the cloisters) contain the residences of the officials,

the library, dining hall and various offices. The chapel represents only the choir of the immense church which the founder had intended to build, but as this was not completed when he was driven from the throne, William of Waynflete, the bishop of Winchester, who had been the second provost of Eton and possibly the first headmaster, added an antechapel (*c.* 1480) in place of the nave that Henry had planned. The chapel is built in the Perpendicular style, and on the walls are an important series of paintings of about 1479–87. A chantry contains the tomb of Roger Lupton, provost 1504–35, whose most notable monument is the fine tower between School yard and the cloisters. Most of the glass, which was inserted in the mid-19th century, was destroyed in 1940 and the new east window, of a striking design, was made by Evie Hone. The library in the cloisters contains many manuscripts, including an oriental collection, and rare books.

The college has far outgrown its ancient buildings, and newer buildings include a chapel for the junior boys, built from the designs of Sir Arthur Blomfield, 25 boardinghouses in which each boy has a separate bed-sitting room, several blocks of schoolrooms, laboratories, music, art and handicraft schools, museums, a gymnasium and a school hall and library block built as a memorial to Etonians who fell in the South African War, 1899–1902. An extensive program of modernization is being carried out.

The main athletic occupations of the boys are cricket and rowing, between which each boy has to choose (those who choose cricket being termed dry bobs and those who choose rowing wet bobs), the peculiar Eton forms of football known as the field game and the wall game, association and Rugby football, which are played in the Lent schooltime, and the Eton form of fives which originates from the space between two buttresses of the chapel. The principal annual celebration is held on June 4, the birthday of King George III, when there are speeches in the Upper school, a procession of boats, cricket matches and fireworks. Another annual celebration takes place on St. Andrew's day, Nov. 30, when there is a match at the wall game between collegers and oppidans, and other football matches are also held. The chief cricket matches are those between Eton and Winchester, played alternately at the two schools, and between Eton and Harrow at Lord's ground, London. The rowing eight competes at Henley Royal regatta.

The school dress is a black tail coat and waistcoat, dark trousers, white bow tie and top hat, except for boys under a certain height who at first wear an Eton jacket and black tie. It became customary to wear this dress in the middle of Queen Victoria's reign. During World War II the top hat was discarded and subsequently was only worn by a few boys on special occasions. There is no reliable evidence for the story that the duke of Wellington remarked that the battle of Waterloo was won on the playing fields of Eton. He is, however, recorded as saying that he believed that he learned his spirit of enterprise at Eton, and this may have given rise to the story.

BIBLIOGRAPHY.—Sir H. C. Maxwell Lyte, *History of Eton College From 1440–1910,* 4th ed. (1911); R. Willis and J. W. Clark, *Architectural History of Cambridge . . . ,* vol. i (1886); A. C. Benson, *Fasti Etonenses . . .* (1899); Sir Lionel Henry Cust, *A History of Eton College* (1899); R. A. Leigh, *An Illustrated Guide to the Buildings of Eton College* (1930); Sir Wasey Sterry, *Annals of Eton College* (1898), *Eton College Register 1441–1698* (1943). (C. A. E.)

ETRURIA, the Roman name for the part of Italy occupied by the Etruscans (*q.v.*). Though it might originally have designated the area of Etruscan penetration on the mainland south of the Alps at its great extent, it came eventually to mean the country bounded, approximately, by the Tyrrhenian and Ligurian seas on the west, by the Magra river and the Apennines on the north, by the Tiber valley on the east and by Latium, with Rome, on the south.

The alternative Roman name, Tuscia, is reflected in that of Tuscany (*q.v.*) or Toscana, though the modern Italian *regione* is smaller than the older Etruria, having lost much territory to Lazio in the south and to the modern Umbria in the east.

The subjection of Etruria by the Romans began with the conquest of Veii in 396 B.C. and was completed after great vicissitudes in 281 B.C. Organized as the seventh region of Italy under the Augustan settlement, Tuscia was joined with Umbria to form a single consular province under Diocletian. After the collapse of the Roman empire in the west and the fall of the Ostrogothic kingdom in Italy, most of Tuscia was overrun by the Lombards in the 6th century and organized as a Lombard duchy. On the destruction of the Lombard kingdom by the Franks in the 8th century the country was incorporated into Charlemagne's dominion. For later history *see* TUSCANY.

The name Etruria, long politically extinct, was revived by Napoleon Bonaparte in 1801, when he converted the grand duchy of Tuscany into the kingdom of Etruria for Louis of Bourbon-Parma (1773–1803), nephew by marriage and son-in-law of Charles IV of Spain. Louis' son Charles Louis (1799–1883; finally Charles II of Parma from 1847 to 1849) succeeded him as king, under the regency of his mother, the infanta Luisa of Spain, from 1803 to 1807. Then in Jan. 1808 the dynasty was dispossessed by Napoleon, who in 1809 revived the grand duchy of Tuscany for Élisa Bonaparte. The refusal of the congress of Vienna to restore the kingdom was a setback for Spanish diplomacy, though the infanta Luisa received the duchy of Lucca in 1815 as compensation for the dynasty's exclusion from Parma during the lifetime of Napoleon's consort Marie Louise.

See also references under "Etruria" in the Index volume.

ETRUSCAN LANGUAGE was the language of the Etruscans (*q.v.*), in north central Italy, from about the 8th century B.C. to the 4th century A.D. From the days of Dionysius of Halicarnassus, who declared the Etruscan was like no other language, to the mid-20th century, many attempts of various kinds, none completely successful, have been made to interpret the extant monuments of this difficult language. Modern scholars base their labours on the internal study of the language and employ the principle of combination and chronology; definite results have been obtained by using established details to elucidate less known details.

Material.—There survive: (1) A certain number of words preserved by Latin and Greek authors such as *aisar* "gods" (Etr. *ais* "deity"), *ister* "actor," whence the Latin *histrio.* (2) Over 10,000 inscriptions from the 7th century B.C. to the time of Augustus, nearly all of which have been collected in the *Corpus;* 90% of them are only proper names. The longest inscriptions are the tablet of S. Maria di Capua (300 words) and the Cippus Perusianus (120). (3) A liturgical text of about 1,500 words written in the wrappings of the famous mummy of the museum of Zagreb. External evidence dates this in the 2nd century B.C. It should be remembered that Etruscan was used after the date of this material and that according to Ammianus the soothsayers used it as a liturgical language up to the 4th century A.D.

Alphabet.—The alphabet closely resembled the Central and Greek alphabets of the Western type especially and shows that Etruscan only possessed surd occlusives which tended to become spirants, as is shown by the variations in the records both of pure Etruscan words as *sec* and *sech, lart* and *larth, acsi* and *achsi,* and in words borrowed from the Greek such as *spurida* (acc.) and *sporta, thriambos* and *triumpus,* so that to the influence of an Etruscan substratum scholars have attributed the change from *c* to *h* in the speech of Florence; *e.g., hasa* for *casa, poho* for *poco.* In Etruscan *f* was weakened to *h* as in *cafatia* beside *cahatia.* Then the short interior vowels were later reduced and often disappeared, perhaps with an initial accent. *Menelaos* is written *Menle; Heracles, hercle; Achilleus, achle; Clutaimestra, Clutmsta.* The Latin loan word *pronepos* becomes *prumths.*

Structure.—Etruscan-Latin bilingual texts are rare, very short, do not always agree and are of very little help. The few definite features are as follows. In names borrowed from Greek, Etruscan distinguished gender by the terminating *e* for the masculine (*hercle, achle,* etc.) and *ai, ei* for the feminine (*Elinai, -ei,* "Heeleene," *phersipnai, -ei,* "Persephone"). In certain true Etruscan names the feminine is sometimes in *i, larth,* fem. *larthi,* or in *ia, arnth,* fem. *arnthia,* or in *tha, lautni* ("freedman"), fem. *lautnitha.* These variations may depend on the form of the root as in the case endings. The nominative is the normal form of the root. The genitive is in *s* or *ś* with roots ending in a vowel (*sethre* and *sethres*), in *uś,* with roots ending in a consonant as *velthur,*

velthurus; thanchvil, thanchviluś. The most characteristic genitive is in *l* and indicates filiation, thus *arnth, arnthal,* "son of Arruns." It may even be added to another genitive, to form a genitive or filiative of the second degree as *velthur* gives *velthuruś,* "son of Velthur," and *velthuruśl,* "son of the son of Velthur." The dative is sometimes in *ś* identical with the genitive as *Mi* ("this") *thupltha*ś ("to Thupeltha") *alpan* ("as a gift") *turce* ("he gave"). The more usual form is in *śi* as *auleśi* ("to Aulus") *meteliś* ("Metellus") *ve(lus)* ("of Vel") *Vesial* ("of Vesia") *clenśi* ("to the son"). "To Aulus Metellus son of Vel and Vesia." The word for son is *clan* in the nominative, *clenś,* genitive, and *clenśi,* dative. It is not known whether an accusative existed. The plural is in *ar: clan* "son," *clenar* "sons"; *ais* "deity," *aisar* "deities."

Verb.—As to the verb, all that is definite is that the third person singular of the perfect tense was in *ce: amce* "has been"; *turce* "has given"; *tece* "has raised"; *lupuce* "has lived." Sometimes this ending is missing, as *avils xxxvi lupu* "lived 36 years."

Numerals.—The order of the numbers on the dice found at Toscanella is open to question, but it is regarded as probable that *mach* is "one"; *zal* "two"; *thu* "three"; *huth* "four"; *ci* "five"; and *śa* "six." It is known that seven, eight, nine are *cezp, semph, muv,* but their exact identification is not settled. Tens were formed by adding *-alch;* thus, from *ci* "five" we get *cealch, celch* "fifty," and from *cezp* we get *cezpalch.*

Vocabulary.—The number of words definitely known by bilingual inscriptions or as contained in concordances is so limited that it is still impossible to decipher accurately any part of a lengthy text such as that on the Zagreb rolls. The following words are known: *usil* "sun"; *tiv* "moon"; *alpan* "offering"; *fleres* "statue"; *zamathiman* "golden brooch"; *hinthial* "soul"; *lautn* "family"; *ril, avil* "age" or "year"; *ati* "mother"; *puia* "wife"; *scc, sech* "daughter"; *thura* "brother"; *nefts* "grandson"—borrowed from Latin *nepos; mi* "this"; *-m, -c* "and." The syntax is known only through very simple phrases such as: *pumpui larthi puia larthal clevsinas aveleśla sech sentinal thanchvilus,* "Larthi Pompeia, wife of Larth Clevsinas son of Aulus, daughter of Tanaquilla Sentinei"; or *trepi thanchvil vipenas arnthal arnthialiśla puia,* "Tanaquil Trebia, wife of Arruns Vipena, son of Arruns."

Linguistic Affinities.—Opinion now strongly favours the Lydian origin of the Etruscans as recounted by Herodotus. A few regard them as indigenous to Italy on the ground that, apart from archaeological evidence, it is impossible, on the linguistic material available, to identify the affinities of Etruscan with any other form of speech. Others point to the use in Lydian and that in the Caucasian languages of the genitive in *l,* to the plural in *ar* found in Caucasian speech, to the copula *-c* found in Lydian, to the presence in Lycian and Etruscan of a number of nominal suffixes in common and to the discovery at Lemnos of an inscription in a language closely allied to Etruscan. (*See* ANATOLIAN LANGUAGES.)

These views are not necessarily contradictory. It may be surmised that the first migration of Etruscans (*Tu[r]sci*) moved from Asia Minor at an early date to Greece (*Pelasgians*) and were soon separated from the other branch (*Turseni*) of the same stock who remained in Asia Minor whence they spread much later. This hypothesis explains the fact and also the paucity of contacts between Etruscan and Anatolian languages. Until a lengthy bilingual text has been discovered, the language must be studied internally, a slow, difficult but relatively sure method.

See J. Whatmough, *Foundations of Roman Italy,* ch. ix (1937); M. Pallottino, *Testimoniae linguae Etruscae* (1954). (J. WH.)

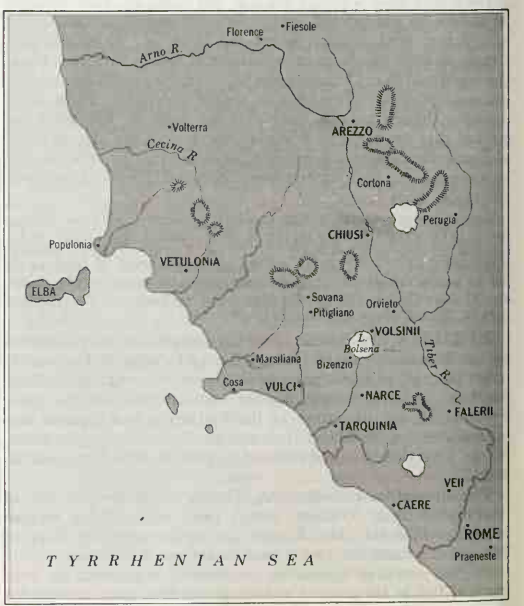
ETRUSCANS. The Etruscans were a people living in north central Italy from the 8th century B.C. onward, called by themselves *rasna* and by the Greeks *Tursenoi.* Before the middle of the 7th century they occupied an area of west central Italy bounded on the east by the valley of the Tiber and on the north by the Cecina river; these bounded classical Etruria in which Etruscan civilization reached its height in the 6th century B.C.

ORIGINS

No definite answer to the question of the origins of these people has been found, and scholars divide in support of three main opinions: that of Herodotus in the 5th century B.C., who considered the Etruscans to be immigrants from Lydia or from the Greek parts of Asia Minor; that of Dionysius of Halicarnassus in the 1st century B.C., who looked upon them as an Italic people native to the area they occupied; and a third theory, little supported, which sees them as migrants from north of the Alps. Modern scholarship has stressed the importance of the study of the conditions under which the Etruscans grew to a dominant cultural status. These conditions were certainly supplied in Italy, for no community of the same ethnic and social characteristics occurred elsewhere in Europe or Asia. The Etruscans spoke a non-Indo-European language of unknown origin. Little more than names and titles can be read, but many words of religious significance survived in Latin. (*See* ETRUSCAN LANGUAGE.)

The "sea peoples" who shortly after 1200 B.C. overran Asia Minor and the coastal lands of the east Mediterranean basin included tribes with names which seem philologically connected with peoples later found settled in the coastlands of the west Mediterranean. Such were the Tshkl and Shrdn, whose names are possibly connected with the Sikels of Sicily and the Sardes of Sardinia. Another group bore the name Turshw, which again finds an echo in the Greek name *Tursenoi* for the Etruscans, but was also applied by the Greeks to certain coastal peoples in Asia Minor. From iconographic evidence and archaeological sources it is, however, impossible to say whether people ancestral to the Etruscans in the west were, in the 11th century B.C., part of a great folk movement from Asia Minor of peoples seeking a foothold in Mediterranean lands. A long time gap divides the sea raiders from the appearance in Italy of elements which can be distinctively recognized as Etruscan.

The only monument outside Italy that can be connected with Etruscan origins is a tombstone engraved with the figure of a helmeted warrior found at Kaminia on the island of Lemnos. Around the figure and on the lateral edges of the stone is a long inscription in an archaic Greek 7th-century script. The language, however, is not Greek but is judged by its inflections and word formations to be similar to Etruscan. This inscription has been taken to point to an Ionian origin of the Etruscans, in agreement with Herodotus' theory of an origin in Asia Minor. It is, however, not known how this stone should be interpreted, since both Etruscan and the similarly alphabetized Lydian language of Asia Minor



IMPORTANT ETRUSCAN CITIES

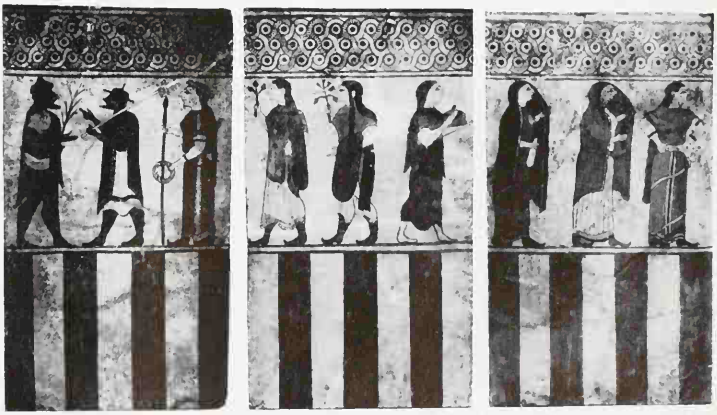
Sarcophagus, dating from the end of the 6th or the beginning of the 5th century B.C. From Cerveteri (Caere). Museum of the Villa Giulia, Rome

Funerary group in soft limestone, 5th century B.C. From Chianciano. Archaeological museum, Florence

Bronze Chimaera from Arezzo (Arretium), 5th century B.C. Found in 1553, it is of pure Etruscan origin. Archaeological museum, Florence

Orientalizing silver bowl, 7th century B.C. Phoenician. The Walters art gallery, Baltimore

The *Mater Matuta*, a funerary statue in soft limestone, 5th century B.C. Found near Chianciano. Archaeological museum, Florence



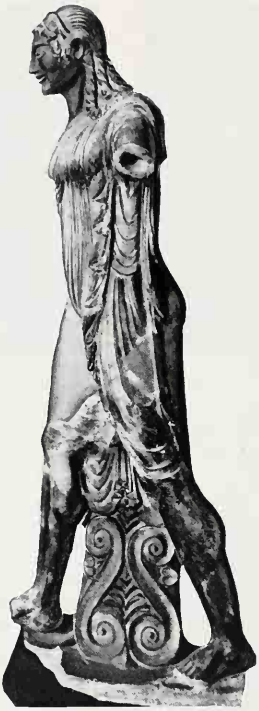
Panels of a wall painting from Cerveteri (Caere), depicting the "Judgment of Paris," probably 7th century. British museum

ETRUSCAN SCULPTURE, PAINTING AND APPLIED ARTS

BY COURTESY OF (CENTRE RIGHT) THE WALTERS ART GALLERY, BALTIMORE, (BOTTOM RIGHT) THE TRUSTEES OF THE BRITISH MUSEUM; PHOTOGRAPHS (TOP LEFT, TOP RIGHT, CENTRE LEFT, BOTTOM LEFT) ALINARI

ETRUSCANS

Terra-cotta statue of Diana found at Civita Castellana (Falerii) in Etruria; dating from the 5th century B.C. City Art museum of St. Louis, Missouri



The Veii Apollo, painted terra cotta, probably the work of Vulca; c. 500 B.C. Museum of the Villa Giulia, Rome

Bronze cauldron on tripod found in the Regolini-Galassi tomb, Caere (Cerveteri); c. 650 B.C. Etruscan museum of the Vatican, Rome

Molded pottery jug in the shape of a duck with painted figure of Lasa, a female Eros purely Etruscan in origin; dating from the 4th–3rd century B.C. The Louvre, Paris

Bucchero (black pottery) wine jug with engraved design, dating from the 7th century B.C. The Metropolitan Museum of Art

Bronze base, decorated with reliefs, of the Loeb tripod from San Valentino near Perugia; c. 540 B.C. Antikensammlungen, Munich

Figure of a warrior on an archaic Etruscan inscribed gravestone found at Volterra; probably 7th-century period. Archaeological museum, Florence

ETRUSCAN SCULPTURE AND APPLIED ARTS

BY COURTESY OF (TOP LEFT) THE CITY ART MUSEUM OF ST. LOUIS, (BOTTOM LEFT) THE METROPOLITAN MUSEUM OF ART, ROGERS FUND, 1921; PHOTOGRAPHS, (TOP CENTRE AND RIGHT, BOTTOM RIGHT) ALINARI, (CENTRE LEFT, BOTTOM CENTRE) FROM M. PALLOTTINO, "ART OF THE ETRUSCANS," THAMES & HUDSON, 1955, PHOTOS BY MARTIN NURLIMANN

are unknown. Its presence in Ionia points to an important factor for the understanding of the conditions in which Etruscan cultural status grew, namely, its close connections with Ionia and Greece.

Attempts to solve the problem of Etruscan origins by the systematic study of ancient craniology have been inconclusive and study of modern Tuscan blood groups, while showing an isolated group with possible eastern affinities in the central Etruscan area, which historically and ethnically has been little disturbed, are difficult to interpret.

The archaeological evidence, while full of difficulties, allows recognition in the Iron Age cultures of northern and central Italy of new elements which made their appearance in the second half of the 8th century B.C. At this period exotic elements appear in tombs of the Villanovan culture, so-called from its first recognition at Villanova, a few kilometres south of Bologna (*see* VILLANOVANS). The little-known Villanovan culture was widespread in north central Italy, reaching the Julian Alps in the north and Campania in the south. It was characterized by a cremation burial rite in which the cremated bones were buried in rough pots of biconical form, a shape which appears in Italian Bronze Age ceramics before 1000 B.C., accompanied by clasps, bronze and iron rings and bronze crescent-shaped razors, some of which objects bear great similarity to metal types in the east central European urn field cultures. No dwelling sites are known and it is uncertain how much the chronological and typological divisions of the culture worked out for the Bologna area are applicable to its southern and western aspects. The Villanovan graves of Etruria, and particularly of Latium, have more flamboyant features than those of the north.

It is in the northwestern coastal districts of Italy that new elements first appear. To the normal Villanovan pit burials are added, in the Vetulonia cemetery of Poggio alla Guardia, new forms of burial inside wells that are carefully lined with stones and surrounded in groups with circular curbs of dressed stones. In the Poggio delle Granate graveyard at Populonia, primitive built tholos tombs containing typically Villanovan objects were in use. In Latium, particularly in the graveyards of the Alban hills, the Villanovan dolium (funerary jar) was replaced by a pottery receptacle for holding the ashes, which was in the form of a circular bust with exterior details in impasto. Inside these funerary vessels were placed weapons of bronze tipped with iron, women's belt clasps, silver hair rings, bronze pins and buckles, the latter valuable for working out chronology and typology.

In the Latium and Etruscan areas the use of beaten bronze suddenly became predominant. Helmets with double crests were placed over the ossuaries (bone urns) of warriors, particularly at Tarquinia (ancient Tarquinii) and Veii (modern Veio). Cauldrons and situlae (buckets) of beaten bronze and lozenge-shaped belt plaques with embossed designs are particularly common, and the biconical funerary urns were imitated in bronze. A number of these objects are paralleled in the northern Villanovan area; their decoration of *repoussé* figures of birds and ducks links them with southern Hallstatt metalwork, particularly as found in Slovenia, but new shapes in beaten metal are entirely Italic. Among these must be listed the flat-sided "pilgrim flask" that occurs at Vetulonia, small tables with attached cups and mounted on wheels, firedogs, thrones—particularly at Clusium (*q.v.*; mod. Chiusi)—along with short swords with crescent-shaped handles of Asiatic design. Certain cast-metal types appear to be intrusive: prong-and-hole belt clasps appear in large numbers and are of Caucasian type, and a new arc-shaped type of clasp is paralleled in the near east.

Impasto is the most common form of pottery decoration in the southwest Villanovan area: animals, human figures, swastikas and other geometric shapes were applied to the surfaces of vessels. A new painted technique appears, however, in southern Villanovan centres at Veii, Vulci and Bizenzio, where large jars for storing wine were painted with hatched triangles and schematized human shapes. In the northern region, at Vetulonia in particular, the pottery tradition contained different elements such as fine vertical-spouted jugs which link this area with the cultures of prehistoric Sardinia.

The hut-urn burials of Latium spread north to Bizenzio, Vetulonia and Populonia. In general the exotic elements in the Villanovan culture spread initially inland in a northeasterly direction from such coastal settlements as Caere (mod. Cerveteri), Populonia, Tarquinia and Vetulonia in a manner consistent with the arrival of a new population on the seaboard. Although no single argument for an eastern origin of the Etruscans is conclusive, the cumulative comparisons which can be drawn in religion, art, tomb structure and building of ships between Etruria and the east strongly suggest that Etruscan civilization consisted of an intrusive east Mediterranean aristocracy ruling an Italic substratum. It is probable that the growing study of the Etruscan language will ultimately be of greater value than archaeology in the determination of the origins of this population.

EXPANSION

By the middle of the 7th century B.C. the chief Etruscan towns had been founded. Before they had reached the Arno river in the north and included all Tuscany in their dominion, the Etruscans embarked upon a series of conquests initially probably not coordinated but undertaken by individual cities. The pressing motive for expansion was that by the middle of this century the Greeks not only had obtained a grip on Corsica and expanded their hold on Sicily and southern Italy, but also had settled on the Ligurian coast (northwestern Italy) and in southern France. Etruscan expansion to the south and east was confined at the line of the Tiber river by the strong Italic Umbrian people settled beyond it on the south and the Picenes on the east. To the northeast no such united power opposed their expansion, since the Apennine mountains in Aemilia (mod. Emilia) and Tuscany were held merely by scattered Italic tribes. Through these the Etruscans were able, in the middle of the 6th century, to push into the Po valley. As capital of this northward region they established the old Villanovan centre at Bologna (the Etruscan city of Felsina) and on the banks of the Reno founded Marzabotto. On the Adriatic coast to the east Ravenna, Rimini (ancient Ariminum) and Spina traded with Istria (ancient Istra) and the Greek Dalmatian colonies. From the Po valley contacts were made with the central European La Tène cultures. Etruscan conquests in the northeast extended to include the modern cities of Piacenza, Modena, Parma and Mantua. To the south they were drawn into Latium and Campania from the end of the 7th century and in the 6th century they had a decisive impact on the history of Rome, where the Etruscan dynasty of the Tarquins is said to have ruled from 616 to 510/9 B.C. It is possible that the Roman Tarquins were connected with a family called Tarchu known in inscriptions.

Rome before the Etruscan advent was a small conglomeration of villages. It was under the new masters that, according to tradition, the first public works such as the walls of the Capitoline hill and the Cloaca Maxima were constructed. Considerable evidence of the Etruscan period in Rome's history has come to light in the region of the Capitol. That there were rich tombs in Rome itself cannot be doubted—tombs similar to those in the Latin town of Praeneste (modern Palestrina). Meanwhile, by the beginning of the 6th century the Etruscans had included Fiesole (ancient Faesulae) and Volterra (ancient Volaterrae) in their northern limits and at the same time began to push southward into Campania. Capua became the chief Etruscan foundation in this region and Nola a second; a necropolis has been found in the Salerno region and Etruscan objects in low levels at Herculaneum and Pompeii. The coastal region was still, however, in Greek hands. When the Etruscans attacked the Greek foundation of Cumae in 524 B.C., their advance was finally checked by their defeat at the hands of Aristodemus of that city. The rivalry between Greek trade in the western Mediterranean and that carried on between the Etruscans and Carthage had already come to a head at the battle of Alalia in 535 B.C., a battle which the Greeks claimed to have won but which so upset them that they determined to abandon Corsica to Etruscan influence.

In the last quarter of the 6th century when Etruscan power was at its height from the Po to Salerno, small settlements of Etruscans might have been planted outside Etruscan limits proper. At

Spoleto (ancient Spoletium) in the north and Fossombrone in Liguria their power was not, however, to last long: the check delivered at Cumae was the first of the sharp waves of resistance that came from the Greeks, Samnites, Romans and Gauls. About 509 B.C. the Etruscans were chased from Rome under circumstances reflected in the story of the expulsion of Tarquinius Superbus, the intervention of Lars Porsena of Clusium and the Latin victory over Aruns Porsena's son at Aricia. When Latium was lost, relations between Etruria and its Campanian possessions were broken with disastrous effect. A series of piecemeal feuds between Etruscan cities and Rome led to the incorporation of the former into the Roman sphere—first the nearby town of Veii in 396 B.C., after which Capena, Sutri and Nepet (mod. Nepi) fell in turn, thus beginning the end of that first Etruscan attempt at the unification of Italy. (*See* also ROMAN HISTORY.)

ARCHAEOLOGY

Local and Imported Pottery.—From the first period when the complicated artistic and archaeological elements called Etruscan come first to be distinguished from the Villanovan cultural background, pottery with geometric designs painted on a light ground pose a problem. Almost all of these were made in Italy but copy (in design rather than in shape) the geometric pottery of Greece in the 9th–8th centuries B.C. While these show evidence of some intercourse between Greece and Italy as early as the second quarter of the 8th century, actual imports of Greek geometric vessels are restricted in the Etruscan area to a few *aryballoi* (narrow-topped globular vessels) of a style dating to the third quarter of the 8th century and immediately preceding the importation of proto-Corinthian vases into Italy, which are considerably more common than late geometric imports. In addition to the best pieces imported from Greece there are several local imitations of Corinthian vessels called Italo-Corinthian ware. This early link in the ceramic styles of Etruria and Greece was to continue for five centuries and demonstrates that, whatever the origins of the Etruscans, their Aegean orientation was strong. A further distinctive factor of the Etruscan facies was the development of a well-purified black glazed pottery called *bucchero* which, while it appears earlier in some of the primitive fabrics of Vetulonia, becomes predominant only in the middle 7th century. In shapes the forerunners of this pottery seem to be east central European Iron Age and Anatolian, while the fabric type is also known in Cyprus and Rhodes.

Besides imported Greek pottery, imported oriental objects are the only other source of outside dating. A vase of blue faïence bearing the cartouche of the Egyptian Pharaoh Bocchoris (Bakenrenef), who reigned in the last quarter of the 8th century B.C., is the only object for which a definite date is possible. It provides, however, only a *terminus post quem* (earliest possible date) for the tomb in which it was found, the Bocchoris tomb at Tarquinia, and not a firm one at that since the vase is possibly a Phoenician copy in circulation considerably later than the Pharaoh's death. Goldwork and other objects from the Bocchoris tomb show a mixture of Villanovan and proto-Etruscan styles and, even if the tomb should be dated slightly before 700, which is very doubtful, it is clear that the Etruscan elements in the population had not crystallized and that the Etruscans must have settled in large enough numbers to affect the Villanovan culture only a little earlier than the bulk of Greek colonization of Italy and Sicily.

Burial Customs.—Significant changes in Villanovan burial customs in coastal areas have already been noted, and in the course of the 7th century these changes took place throughout the entire area of Etruria. While at Vetulonia and Marsiliana d'Albegna the Villanovan "pit" and "well" tombs were continued, they were grouped in circles and covered with tumuli. In the inland settlement at Clusium the pit tombs with a clay representation of a human figure, often seated on a throne, retained conservative forms. The small burial vaults already noted at Populonia came into use at Vetulonia; chambered tombs with rectangular chambers and long corridors were cut in the rock at Caere and Tarquinia (*qq.v.*); at other sites chambered tombs were constructed of carefully dressed stones or partly cut in the rock and vaulted with false vaults. These were covered with tumuli and often retained by circular curbs of dressed stones. The largest of these is the so-called Regolini-Galassi tomb at Caere beneath a tumulus 48 m. in diameter and with five chambered tombs dug into its circumference. At Caere are also found tombs with numerous chambers, and there are partly built and partly rock-cut tombs at Vetulonia, Vulci, Veii, Volsinii (modern Orvieto), Chianchiano, Castellina (in the Chianti mountains) and Praeneste. Many of these are later than the 7th century, and vault tombs of circular plan known at Casal Marittimo, Casaglia (near Montecatini) and at Quinto (near Florence) are of the 6th or early 5th century.

These cemeteries of chambered tombs were most often situated on small hillocks, and rock-cut tombs with molded doors are cut in the faces of sheer cliffs or rocks at Norcia, Viterbo, Castel d'Asso and Sovana. Unusual are the 5th-century Populonia built tombs which imitate rectangular houses set on high plinths. Prehistoric Villanovan tombs are known to have preceded built or rock-cut Etruscan tombs at Cortona, Bologna, Clusium, Volterra and Vetulonia. A further type of tomb consisted of a freestanding coffin made of stone slabs and crowned with an ovoid column. This was a type particularly common at Marzabotto. Inscribed gravestones are not common, but there is an early one commemorating Aule Pheluske at Vetulonia, and other examples are from Monte Gualando and Perugia (ancient Perusia) and a series with warriors in low relief from Volterra.

Orientalizing Period.—It was the great wealth of objects contained in these chambered tombs which caused the rifling of Etruscan tombs during the Renaissance and gave birth to the scientific spirit of etruscology in the second half of the 19th century. Although there are a number of gold plaques with purely geometric ornament, the most sumptuous objects found in the tombs bear patterns derived from the orient and hence characterize the earliest well-defined period of Etruscan archaeology, known as the orientalizing period.

Egyptian scarabs and other small objects were imported in the 8th century and scarabs are found in purely Villanovan graves. The bulk of objects which were either imported from the east or were local copies of imports fit into a context of *c.* 650–625 B.C. From the rich Caeretan tombs of Regolini-Galassi, the Barberini and Bernardini, and the Castellani tombs at Praeneste comes a wealth of gold and silver objects with beaten and modeled designs of sphinxes, winged bulls and lions and lotus blossoms introduced from the east. From these tombs come also a set of silver jugs and bowls of Egyptianlike beaten design.

Both the jugs and the bowls find their closest analogies in Cyprus whence they were traded by Phoenicians. One of the silver bowls found in Etruria bears the inscribed name in Phoenician of its owner or maker. This bowl and another from Praeneste are certainly imports, while bowls from Caere, Salerno, Vetulonia and Praeneste, together with others now in Baltimore, Md., and Amsterdam, Neth., probably were made by Phoenician workmen or their descendants settled on Etruscan soil. The earliest of these silver vessels is a cup from the Tomb of the Prince at Vetulonia, Greek rather than Phoenician in shape, and from the Regolini-Galassi tombs silver vessels of purely Etruscan shape accompany the Phoenician vessels and show that silversmiths were working locally.

This imported metalwork had a profound effect upon the shapes of Etruscan *bucchero* vessels which, in the latter half of the 7th century, often bear impressive designs of winged animals or Phoenician lotus patterns.

Artistic impulses in the orientalizing period did not, however, all come from Cyprus or Phoenicia. The closest parallels to the long-legged animal style of the 7th century with its winged bulls, horses and stags walking quietly in procession are to be found in the orientalizing art of Crete, while the lion style initially owes something to the sculptured lions of the Syro-Hittite region in southeastern Anatolia. Another source of inspiration was the metalwork of the Vannic kingdom of Urartu (*q.v.*; mod. Armenia and eastern Turkey). Cauldrons with siren handle attachments reached Etruria from the same source as they reached the shrines of Archaic Greece (probably by way of Ionia and the Greek Black

sea colonies) and along with them tripods and metalwork stands. Cauldrons were imported to Etruria half a century before the orientalizing period proper, the earliest being that from the Cauldron Circle at Vetulonia.

The routes by which other orientalizing objects reached Etruria can only be conjectured. While Carthage seems to have played no part in this orientalizing trade, direct relations between Etruria and Cyprus and Phoenicia are indeed possible, but small blue faïence flasks of Rhodian manufacture and indeed the Urartian bronzes might have found their way westward by trade in the hands of Phocian Greeks, who were busy in the western Mediterranean in the second half of the 7th century. It is known that the Phocian Greeks had a settlement called Pyrgi (probably mod. Santa Severa) on the coast near Caere, but it is not known whether the neighbouring port of Punicum (site unidentified) indicates a settlement of Carthaginians or Phoenicians.

Tombs.—Ultimately the chronology of the rich Etruscan tombs depends upon that of the imported proto-Corinthian and Corinthian pottery, which is somewhat controversial. L. Pareti classified the tombs according to their contents and the extent of imports and made an acceptable scheme. It is fraught with difficulties, however, because these tombs were explored at a period when the archaeological significance of the "tomb group" was not appreciated, and the assemblages were often salted with other finds to improve their commercial value to museums. Also several burials, often spaced in date, took place in some of the larger tombs and were confused in excavation.

There can be little doubt that the earliest tombs with exotic products and imports are the Circle tombs of Vetulonia (*q.v.*), where native Villanovan tradition was unbroken. Fifteen *pozzeti* (well tombs) were placed inside circles on the east side of the hill on which Vetulonia stood, and silver and amber ornaments were found in these, although it is not certain that they should be regarded as Etruscan proper. Nearby on the east and south sides of the Poggio alla Guardia were similarly encircled graves in which inhumation as well as cremation is in evidence (Circle of Bracelets, Bes Circle, Tomb of the Cone, etc.). These tombs were rich in gold, silver and bronze objects. Gold clasps and pins bore animal designs in free granulation, and bronze vases for serving wine had curly handles terminating in the figures of griffins. Only slightly later is the Circle of the Trident which, in addition to the huge trident after which it was called, contained a chariot with innumerable horse trappings, all of Italiate work.

In the first truly Etruscan period may be placed the Giulimondi tomb at Caere, excavated by G. Pinza, the Warrior tomb and the Bocchoris tomb at Tarquinia, all of which contain only geometric pottery, although in the Giulimondi tomb are two early proto-Corinthian cups. Bronze work, including a "pilgrim flask" from the Warrior tomb, presents new types and gold plaques are of geometric decoration.

To a second phase, predominantly geometric but characterized by the first proto-Corinthian imports, must be placed two tombs at Cumae, a tomb at Falerii (modern Civita Castellana) and the material of the right alcove of the Regolini-Galassi tomb at Caere.

In a third phase the first orientalizing imports appear. Outstanding are the Tomb of the Prince at Vetulonia (burials II, V) and the burial in the main chamber of the Regolini-Galassi tomb. To these may be added tombs at Ponte Sodo, Vulci, the Tomb of La Pietrera, Vetulonia (fossa V), Tomb 10 of the Buccace at Bizenzio and material from the Poggio la Sala, Clusium.

After the second and third phases there is a wealth of Corinthian imports, early Corinthian in the Pontecapuana tomb, Salerno, the Villa Tarquinia tomb, Falerii, the Tripod tomb, Caere, and the contents of the forechamber and left alcove of the Regolini-Galassi tomb (late in the phase). These constitute a fourth phase which is increasingly rich in orientalizing material and in which the development of *bucchero* and Italo-Corinthian ware occurred.

The fifth phase is characterized by imports of middle-Corinthian pottery with much relief decoration on bronzes and metal, Phoenician silver vessels, pyriform *bucchero* jugs and exotic *bucchero* vessels with plastic caryatids and horse and griffin heads, etc. So provided are the Barberini and Bernardini tombs, Praeneste, the Periphery tomb III in the Regolini-Galassi tumulus, pit graves 1, 2 and 4 of the Pietrera tomb at Vetulonia, the Tomba Campana at Veii and Tumulus of the Razors at Populonia, and the Circle Tombs of the Ivories and of the Fibulae at Marsiliana d'Albegna.

To a final phase dating slightly before 600 B.C., in which Ionian motifs become recognizable, belong the Tomb of the Snare, Clusium, and the tumulus at Castellina in Chianti.

ART

Sculpture.—Etruria, Latium and the Faliscan districts fall into three main areas of artistic production: northern, central and southern, each centred upon cities with a distinctive artistic style. In the southern areas the chief centres were Caere and Veii, in which the Etruscan style most closely approached the Greek. In central Etruria Vulci was evidently the leading art centre, although Tarquinia was unsurpassed in the beauty of the wall paintings. Several potters were at work in Vulci, and the greater part of the central Etruscan bronzes, artistically the best, were produced there. The north was dominated by Clusium, although Perugia seems to have been important along with lesser centres at Volterra and Fiesole.

The very earliest pieces of Etruscan statuary are flat rectilinear figurines from Vetulonia and Capodimonte di Bolsena. These figures occur in later contexts in the Regolini-Galassi and Bernardini tombs, both of which contain pieces in a more advanced style which cannot have developed much later. These are statuettes of women with pigtails and long skirts depicted in a manner in which the impulses are perhaps north Syrian, although this female type, copied in ivory and amber, is certainly local.

The earliest evidence of Greek influence is the presence of centaurs, perhaps transmitted on Corinthian vases. Their style in Etruria is orientalizing with a slim body and elongated legs, perhaps reflecting Cretan influence. These and other mythical creatures found great favour with the Vulci stonemasons. To archaic works of early Etruscan sculpture certain Greek parallels can be found in the late 7th and early 6th centuries, and in general characteristics the works still followed the Greek Archaic Daedalic tradition. The next change in style took place *c.* 550 B.C., when art became distinctively Ionian. These new influences can be seen earliest in such pieces of bronze work as the Loeb tripod from San Valentino near Perugia and the Monteleone chariot platings (in the Metropolitan museum, New York city), but soon become apparent also in the relief designs on *bucchero pesante* (heavily embossed black pottery) and in architectural reliefs like those from Tarquinia. By the end of the 6th century B.C. Veii possessed an excellent school of terra-cotta sculptures in Ionian styles. The statues of Apollo and of a votaress suckling a child are elaborately stylized in features, draperies and muscles. Clay statuary, still retaining traces of former painting, was made in many Etruscan centres. Examples in the more mature classical style that began in the last quarter of the 5th century are the satyr-and-maiden groups from Satricum (modern Conca) in the Villa Giulia museum, Rome, which contains a rich collection of architectural terra cottas from Caere, Falerii, Veii, Satricum and other sites.

These pieces of statuary were designed to stand on temple roofs, and the socketed bases by which they were fixed have survived. Terra-cotta sculpture was also used for antefixes for these temples, but above all for funerary sculpture. Sarcophagi with the sculptured figures of the husband and his wife reclining on the lids begin late in the 6th century with an excellent piece from Caere. Bronze sculptures also span this period from the end of the 6th to the end of the 5th centuries, beginning with the famous Capitoline wolf, the symbol of modern Rome, the later Arezzo chimera and the Todi Mars of the late 5th century.

But in spite of considerable achievements in sculpture in the round, distinct preference for low relief was shown and a large series of 6th–5th-century relief sarcophagi, the finest from

Clusium, depicting dances, funeral games and banquets, or the journey of the dead to the underworld, are a major source of information on Etruscan everyday life. Superbly carved gravestones of the late mid-6th century are known from Clusium and Settimello, but the disk- and horseshoe-shaped gravestones of the Bologna, Fiesole and Populonia graves have crude reliefs.

Painting.—Painting is known almost entirely from tombs. Sixty painted tombs survive at Tarquinia, 20 at Clusium, and a number at Caere, Veii, Orvieto and Vulci. This painting was not merely decorative or commemorative but fulfilled a religious function in an attempt to recreate around the dead person his earthly environment. Subjects were oriental in theme; *e.g.*, processions of griffins and winged horses, etc., with occasional scenes from Greek mythology. But later, scenes of funeral obsequies and funeral games appear. The painting was sometimes done on free-standing whitened terra-cotta slabs largely in black and purplish-brown like the examples in the British and Villa Giulia museums. In similar style but advanced technique are the Campana slabs from Caere (in the Louvre), showing two seated men talking and a winged creature and an archer carrying off a woman. Dull red is here added to purple and brown, and the influence of Corinthian vase painting can be seen. These slabs date to the end of the 6th century B.C.

The Tarquinia school was more dramatic and added a blue and green and several intense browns. Most famous is the Tomb of the Augurs with its scenes of wrestlers, a dancing woman and Silenus, musicians and a referee (not an augur) hastening to watch the scene. The Tomb of the Lioness, with its dancers, musicians and banquet scene, is in a style of which there are many examples at Tarquinia among tombs dating *c.* 520 B.C. Only slightly later is the Fowling and Fishing tomb with boys hurling slingstones at multicoloured birds and diving into the sea from rocks.

Gradually at Tarquinia and Clusium Etruscan costume gives way to Ionic and the scenes become rather Bacchic, as on the Tomb of the Leopards and of the Triclinium, the Tomb of the Funerary Couch and the Justiniani tomb.

The rising influence of Attic art in the years 490–470 B.C. shows in the paintings of a new classical style in which Clusium rose to prominence; *e.g.*, the Tomb of the Monkey. From then until the end of the century, at both Tarquinia and Clusium, the influence of Attic red-figure vase painting, already apparent in the Triclinium and Justiniani tombs predominates and the reality of the scenes is enhanced by the use of shadowing and dimensional devices.

Between 430 and 400 B.C. there are few notable examples of tomb painting, but after the submission of Etruria to Roman artistic hegemony there was a renaissance of painting, now in Hellenistic style. The earliest examples are the Golini tombs at Orvieto and the Orcus tomb at Tarquinia. This new painting tradition continued until the middle of the 1st century B.C. in such tombs as the Campanari and the François tombs at Vulci and the Tomb of the Typhon at Tarquinia. This late phase is noted for its intense chiaroscuro, three-quarters stance, use of perspective, full-field grouping, as in the "Sacrifice of the Trojan Captives" in the Francois tomb, and the use of ethereal blues.

ARCHITECTURE

What little remains of Etruscan stonework has survived by virtue of its massiveness. Foundations of city walls survive at Volterra, Volsinii and Cortona, but those of Perugia are more complete and have surviving vaulted gates. The tombs, themselves replicas of house interiors, show something of moldings, arches and vaults. True vaulting and arching were known and these enabled larger buildings than those of Greece to be constructed. Wood was the chief building material for domestic purposes and also for temples, from which the terra-cotta decorations alone survive. Blocks of houses in carefully paved streets are known only from Marzobotto, although the layout of tombs with paved streets in the Banditaccia at Caere and the Cuccumella at Vetulonia makes them veritable cities of the dead. The Latian hut urns show that Villanovan houses were circular in plan with conical roofs, but a stone urn from Clusium is modeled in the form of a rectangular Etruscan house erected on a tall, stepped platform. The roof is a hipped gable and has a gabled gallery over it. In this connection it is interesting that the Romans attributed to the Etruscans the construction of the atrium house. Temples were rectangular in plan and divided into three cellae (chambers); ground plans and nothing more are known from Rome and Bolsena.

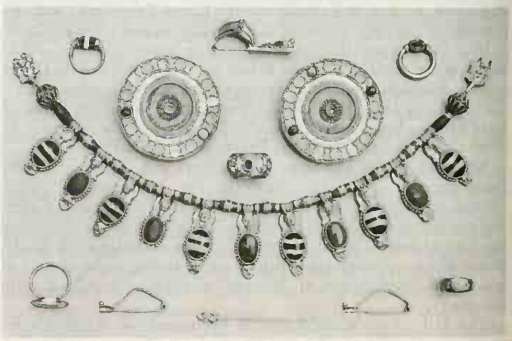
Etruscan temples were built of wood and brick upon high platforms of dressed stone and were consequently more perishable than their Greek equivalents. They were crowned and decorated by brightly painted statues and revetments of terra cotta, many of which have survived.

ARTS AND CRAFTS

Metallurgy.—Casting of bronze was expertly practised by the Etruscans for statuettes, chariot pieces and harness. The technique of beaten or laminated bronze inherited from the Villanovans was used by the Etruscans in 7th-century tombs for the manufacture of thrones, cauldron stands, censors, etc., all elaborately embossed, and also for fans, shields with gorgon and Archilous heads, all of which are characteristic of Etruscan bronze work of the early period. This does not survive the 5th century and is gradually replaced by cast metal. Beaten bronze work of Villanovan style is also found in cultures peripheral to the Etruscan in the early period, such as the beaten bowls of the Picenes and the situlae of Hallstatt influence from Este, Certosa di Bologna and Vace in Slovenia, with registers of embossed marching animals, banquets and funerary games.

In Etruria bronze engraving becomes common in the 5th century, particularly in engraving the backs of mirrors with scenes of classical mythology, and for the bronze cylindrical cistae (coffers) with lids in the form of wrestling boys common in the 4th century. Bronze was used in preference to iron for spears and swords, although a number of iron axes are known. Helmets were made of bronze and followed the Corinthian pattern. On the statuettes of warriors bronze laminated body armour is shown, and although there are solid breastplates from the northern Etruscan region, none of the coats of mail survives.

Jewelry.—The earliest jewelry is that of Vetulonia, on which granulation and filigree wire are applied and plaited gold wire is used on the interstices of gold straps. Elaborately granulated and embossed earrings and beads come from the orientalizing tombs and represent the products of jewelers' shops probably centred at Praeneste. The styles of ornaments in use in the earliest Etruscan period, although they have no direct parallels elsewhere, seem to stem from an area in touch with ancient Mesopotamian technique. Particularly attractive are the pronged shoulder clasps for cloaks modeled with gold granulated sphinxes and lions. Jewelry styles changed sharply in the late 6th century, and the types most common are disks for stitching on clothing and barrel-shaped earrings. Spina, on the Adriatic coast, seems to have been a centre of Ionic-style jewelry production.



BY COURTESY OF THE METROPOLITAN MUSEUM OF ART, DICK FUND, 1940

SPECIMENS OF ETRUSCAN JEWELRY FROM VULCI (*c.* 500 B.C.)

BY COURTESY OF MUSÉE DES BEAUX-ARTS, DIJON

GRANULATED GOLD ETRUSCAN FIBULA. IN THE MUSÉE DES BEAUX-ARTS, DIJON, FRANCE

Large objects of goldwork come from the orientalizing tombs, fretwork coverings of golden animals for wooden situlae from the Castellani and Regolini-Galassi tombs and a gold-covered coffer from Vetulonia. After the silverware produced in the 7th century by resident Phoenician workmen, silver was rarely used in Etruria although it was mined in Populonia (*q.v.*) and was used there and elsewhere for the earliest issues of coinage.

Ivory.—As part of the orientalizing influence new materials were brought into use, among them ivory imported from Phoenicia or Carthage. A few carved pieces from the Bernardini tomb are oriental imports but raw ivory was locally worked. The commonest pieces are small ivory caryatids holding their pigtails, of a type copied in amber and *bucchero*. More ambitious were the ivory arms and situlae, fine examples of which come from Clusium where a situla showing the escape of Ulysses from Circe is found. Ivory pieces including combs come from the circle of ivories at Marsiliana d'Albegna. They probably date to the third quarter of the 7th century and are strongly orientalizing, although they do not compare in beauty with the Etruscan-style plaques from Volsinii dating to the last quarter of the 6th century.

Amber.—Beads and parts of fibulae of amber are common in the Iron Age tombs of Italy, and in the early 7th century melon-shaped amber beads and small amber amulets in the shape of women and monkeys appear in Villanovan graves and seem to be the products of one amber shop, perhaps at Narce. Amber as an artistic medium became popular in the late 6th and early 5th centuries when a series of exquisite miniature amber carvings were produced in Greek style depicting, after the natural contours of the amber piece, intricate groups of reclining people. Some of these pieces were mounted on fibulae and worn as much for the amuletic and medicinal properties attributed to them as for their artistic merits. Some of the finest have been found in the Ancona region and it is possible that they were produced by the experienced Picenian amber workers of that region.

Pottery.—Although the Etruscans had their own black pottery, the light *bucchero sottile* and the heavily embossed *bucchero pesante*, they were quick to imitate imported Greek vases beginning with those of Corinth. Few Italo-Corinthian vases have any artistic quality and it is the black-figured vases made in the second half of the 6th century and the first years of the 5th, particularly a set of ewers supposedly manufactured at Caere, probably by a settlement of Ionian potters, which introduce seriously artistic vase painting. Plenty of red-figure vases were produced after Attic models, but a group of red-figure vases beginning shortly after 400 B.C. appears to have been produced in Faliscan territory under non-Attic inspiration. Other specialties practised chiefly in the 4th century were the overpainting of red-figure vases with details in white and yellow paint and the making of molded jugs in the forms of ducks, Negro heads and the heads of devils.

COMMERCE

Archaeological research in many lands has indicated the wide extent of Etruscan commercial relations. Although to the south Sicily, except for the west Phoenician colony of Motya, provided no markets, Carthage was a ready buyer of *bucchero* and wine ladles. A number of objects were exported to Spain in the 5th century, the period of Etruria's greatest trade expansion. *Bucchero* vases are commonly found in the Greek and Ligurian settlements of southern France, and the Rhône-Rhine areas were particularly important channels of Etruscan trade, by which bronze vessels reached the tables of La Tène France, Switzerland, Germany and ultimately Britain and played an important role in the development of metallurgy among Celtic peoples. The most common exports were the bronze Schnabelkannen (slope-spouted jugs), most probably made at Vulci. These have been found in Switzerland, Germany, France, Spain, Hungary and north Africa. Istria had Etruscan connections and imported Etruscan metalwork, and some of the elaborate vessels from a rich grave at Vix (France) are Etruscan.

While small settlements of Etruscan traders were probably established in many foreign cities, no permanent colonies were made overseas, with the possible exceptions of Elba and the east coast of Corsica. Trade was carried on with the Sardinians, who probably were in some areas related to the Etruscans, but there is no evidence of a permanent colony of Etruscans on Sardinian soil.

RELIGION AND SOCIAL LIFE

According to Livy, the Etruscans were more addicted to religious practices than any other nation. These practices consisted largely in various forms of divination to ascertain the will of the gods. Bird watching, observation of lightning and other weather phenomena, the ebbing and flowing streams were all part of that *Etrusca disciplina* passed on to the early Romans. Interesting is their practice of liver reading, an art largely Babylonian but also practised by the Canaanites. A bronze model liver from Piacenza is divided into 45 areas, each with the name of a presiding deity written in it, and in shape it is similar to the clay model livers from Babylonia. The cult of running waters is not a widespread usage, but is well attested in Anatolia. Places, trees and stones probably all had individual spirits, and a number of sacred meteoric stones standing on carved plinths have been found.

Over earth and heaven ruled a group of gods and goddesses whose roles were individualized largely under Greek influence. Supreme among them was Tinia (Tin, Tina), but Latin authors also speak of a god Voltumna (Veltha, Veltune) of whom figural representations survive. Tinia, along with Uni and Minrva (Minerva), comprised a divine triad which, with 9 other deities, made up a heavenly council of 12. It is possible that Voltumna was a version of Tinia peculiar to the Volsinii. Other local deities are Sethlans of Perugia, Fufluns (later equivalent to Bacchus) of Populonia and the eponymous Tarchon at Tarquinia. Celestial bodies were deified (Usil, the sun; Tiv or Tiuv, the moon), although there is no evidence of interest in a solar cult or astrology.

Many of the names of deities sound so similar to the Greek gods with which they were combined that it is possible that they were imported from the Hellenic world in the first place, hence Minrva, Minerva; Nethun, a sea-god, Neptune; Turm, Hermes; and Trithn, Triton, a water spirit. The Etruscan Velchan might, however, be original and have become the Roman Vulcan.

The nature of the deities ruling the underworld is not clearly understood, but the chief seems to have been a goddess equivalent of the Ceres-Demeter type. Also connected with the fortunes of death was the satanic figure of Charun (Gr. Charon). With these deities the Etruscans communicated by deep stone-lined wells (Lat. *mundus*, "world"), the stone covers of which were opened on sacred days. There is some evidence that these wells were occasionally included in the tripartite temples. The area of the sky set apart for an individual omen of lightning or bird flight was called a *templum* and became the Latin word for temple. So careful were the Etruscans to provide the dead with the furniture of everyday life, even including carts and chariots, beds and thrones, that belief in some form of life after death is implicit.

It was religion which held the Etruscans together politically. Throughout their history they seem to have remained a loose federation of 12 tribes, each drawing its name from a metropolis; *e.g.*, Tarquinienses, Volsinienses. Each year the tribes called a convention at the sanctuary of Voltumna near the lake of Bolsena. Although these conventions were essentially religious and were doubtless coupled with competitive games, political assemblies met at the same time and elected a king for the year, whose functions were probably nominal.

After a short monarchic phase, reflected in Etruscan influence in Rome's early history, Etruscan society became essentially oligarchic. Numerous titles survive of ruling aristocrats bearing such offices as that of *zilath* and *purthne*, but the exact functions of these are not known and the only insignia of authority known is the bronze fasces, the double ax surrounded by the rods of correction, from Vetulonia, an object later associated with the Roman lictors.

Almost nothing is known of the strata of Etruscan society, although it is likely that foreign artisans and descendants of other Italic peoples formed a lower stratum, and slavery was certainly practised. Nor are the internal politics of the Etruscan cities understood, or the reasons why certain marginal cities like Perugia grew to great influence while others like Bizenzio, a flourishing centre in the early period, became slowly extinct. In a civilization so essentially urban as the Etruscan, the movements of industry and the wealth that followed it must have played a decisive role.

See ETRUSCAN LANGUAGE; PRE-HELLENIC ARCHITECTURE; ROMAN HISTORY; POTTERY AND PORCELAIN: *History of Ceramics: Italy;* VILLANOVANS; *see* also references under "Etruscans" in the Index.

BIBLIOGRAPHY.—Raymond Block, *The Etruscans* (1958); M. Pallottino, *The Etruscans*, Eng. trans. by J. Cremona (1955); N. Åberg, *Bronzezeitliche und früheisenzeitliche Chronologie*, part i, *Italien* (1930); A. Åkerström, *Studien über die etruskischen Gräber* (1934); F. Altheim, *Die Ursprung der Etrusker* (1950); A. Andrén, *Architectural Terracottas From Etrusco-Italic Temples* (1940); J. D. Beazley, *Etruscan Vase-Painting* (1947); C. Clemen, *Die Religion der Etrusker* (1936); G. Dennis, *The Cities and Cemeteries of Etruria*, rev. ed. (1878); P. Ducati, *Storia dell' arte etrusca* (1927), *Etruria antica*, 2 vol. (1925); E. Gerhard, *Etruskische Spiegel* (1868); G. M. A. Hanfmann, *Altetruskische Plastik* (1936); O. Montelius, *La Civilisation primitive en Italie*, vol. ii (1921); H. Mühlestein, *Die Kunst der Etrusker* (1929); L. Pareti, *La Tomba Regolini-Galassi*, etc. (1927); D. Randall-MacIver, *Villanovans and Early Etruscans* (1924); G. M. A. Richter, *Handbook of the Etruscan Collection*, Metropolitan Museum of Art (1940); P. J. Riis, *An Introduction to Etruscan Art* (1953); *Tyrrhenika: an Archaeological Study of Etruscan Sculpture in the Archaic and Classical Periods* (1941); F. Schachermeyr, *Etruskische Frühgeschichte* (1929); J. Whatmough, *The Foundations of Roman Italy* (1937). (WM. C.)

ETTMAYER, KARL VON (1877–1938), Austrian Romance linguist, who advocated the psychological approach to linguistic problems, was born on July 22, 1877, at Jessenetz, Moravia. He studied at the universities of Innsbruck, Graz and Vienna, obtaining his doctorate in Vienna in 1899, and taught at the Universities of Freiburg (Switz.) from 1905 to 1911, Innsbruck until 1915, and the University of Vienna, where he remained as professor of Romance linguistics until his death. Ettmayer's most important work is the *Analytische Syntax der französischen Sprache* (1930–36), a book of over 1,000 pages. Dealing mainly with Old French, it constitutes an attempt at psychological (as opposed to logical, or grammatical) syntactic analysis of a language. The method found few adherents, and was severely criticized in the decades after Ettmayer's death. Other books are *Bergamaskische Alpenmundarten* (1903) and *Vademecum für Studierende der romanischen Philologie* (1919). He died on March 24, 1938, at Vienna. (E. PM.)

ETTRICK, a river and parish of Selkirkshire, Scot. Pop. of parish (1961) 260. The river rises in Capel fell in the extreme southwest of the shire and flows northeast for 32 mi. to the Tweed, its principal affluent being the Yarrow. The Ettrick is noted for its salmon and trout. Ettrickdale has literary and family connections with Sir Walter Scott. James Hogg (*q.v.*), the "Ettrick shepherd," was born in Ettrick parish, the site of the cottage being marked by a monument erected in 1898. About 2 mi. below Ettrick church is Thirlestane castle, and beside it is the ruined stronghold that belonged to John Scott of Thirlestane under James V. Two miles up Rankle burn, a right-hand tributary, lies the site of Buccleuch, also a stronghold of the Scotts, which gave them the titles of earl (1619) and duke (1663). Remains of other Scott strongholds are Tushielaw tower, of which there is a fragment on high ground opposite the confluence of the Rankle and the Ettrick, and Oakwood tower, near the site of a Roman camp.

ETTY, WILLIAM (1787–1849), English painter, was born at York on March 10, 1787, the son of a miller. He served as an apprentice printer on the *Hull Packet* until 1805, and was admitted to the Royal Academy schools in Jan. 1807; in July, Sir Thomas Lawrence accepted him as a private pupil for a year. He first exhibited at the Royal Academy in 1811. His sketchbooks of this period contain lively records of groups of children, etc., observed from nature and in a variety of mediums. John Flaxman and Lawrence were the strongest influences, but by 1818 Etty had developed considerable powers as a portraitist. His first journey abroad was to France in Jan. 1815, followed by another to France and Italy in 1816. He enhanced his growing reputation in 1821 by his "Cleopatra's Arrival in Cilicia" (Lady Lever art gallery, Port Sunlight) which, although an unwieldy composition, revealed his powers as a colorist. He only attained maturity, however, after a visit to Italy (1822–24), when he had studied Raphael, Michelangelo, the Carracci and, more especially, the 16th-century Venetians. The grand but simply conceived "The Combat" (1825; National Gallery of Scotland) brought Etty his first great success. In 1824 he became an associate of the Royal Academy and in 1828 academician, and during the next decade he painted his finest works.

After 1830 religious and classical subjects received equal attention, and Etty became friendly with A. W. N. Pugin and his circle. During the last ten years of his life bad health, economic pressure and unenlightened patronage forced him to concentrate on minor fancy pieces which sold easily. His nude studies are often brilliant pieces of flesh painting and result from continuous attendance at the Royal Academy's life school, but small landscape and still life studies are unexpected additions to his work of this period.

Visits to the Netherlands in 1840 and 1841 fulfilled Etty's ambition to see Rubens' paintings there, and two years later he was again in France. A retrospective exhibition was held at the Society of Arts in 1849. On Nov. 13, 1849, Etty died at York. Etty's promise in youth was much greater than his subsequent achievements. Although admired by the younger generation, he was considered the last of the academic history painters rather than the leader of a new realist school.

BIBLIOGRAPHY.—Autobiography in *Art Journal*, vol. xi (1849); W. Gaunt and F. G. Roe, *Etty and the Nude* (1944); Dennis Farr, *William Etty* (1958). (D. L. FR.)

ÉTUDE, the French term in common use for a musical study. The étude was primarily designed to exercise the player, usually of a keyboard instrument, in a technical problem. Works of this kind are also known as "Lessons" (notably those of Purcell and Handel), "Esercizi" (the original name of the harpsichord sonatas of Domenico Scarlatti) or "Studien" (*e.g.*, the *Charakteristische Studien* of Ignaz Moscheles). Études fall into two main classes. They may either serve as technical exercises, such as the studies of Karl Czerny, Muzio Clementi and J. B. Cramer; or they may far transcend this function, as in the *Études* of Chopin, the *Études d'exécution transcendantes* of Liszt, the *Études symphoniques* of Schumann and the *Douze Études* of Debussy, in which a technical problem itself provides a composer's inspiration, thus exercising the performer's technical and musical faculties.

ETYMOLOGY, that part of linguistics which is concerned with the origin or derivation of words (Gr. *etymos*, "true," and *logos*, "account"). Primitive peoples consider that the name of a person or of a thing is intimately connected, even if not actually identified, with that person or thing: *nomen est omen*. The etymologizing of proper names, which is so conspicuous in the Old Testament, to some degree reflects this primitive concern with names (*cf.* Gen. xvii, 5 and xvii, 15 for the changing of the name Abram to Abraham and that of Sarai to Sarah). Among the ancient Greeks, also, the earliest attempts at etymology assumed that the name by which a thing was called stood in some simple and explicable relationship to the thing itself. To etymologize a name was merely to give a "true account" of its meaning, not to provide a scientific explanation of it in the modern sense. Plato, in his *Cratylus* (with its significant subtitle, *On the Propriety of Names*), pokes fun at the fanciful etymologies of the day; at the claim of Socrates, for example, that *hero* (*hērōs*) is derived "with a slight change of form" from *erōs*, "love," because "the heroes

owed their birth to love." But even Plato, like the Stoics, who paid great attention to etymology, had little sense of historical development and little acquaintance with other languages. The Romans, under Greek influence, continued to regard etymology as an arbitrary pastime. Marcus Terentius Varro knew that *meridiēs* ("noon") stood for a nonclassical *medi-diēs* (*i.e.*, "midday") and adduced older forms where he could; but his etymologies, if not self-evident, are largely wordplay.

Scientific etymology had to wait until the 19th century, when the groundwork of comparative linguistics was laid down. The founders of the science, from Franz Bopp to Karl Brugmann, emphasized the so-called "sound laws"; *i.e.*, regular patterns of phonetic correspondence detected over a given period between related languages or between successive historical phases of the same language. These laws, first used to reconstruct a hypothetical Indo-European parent language from which the historically attested Indo-European languages have been derived, were subsequently applied with signal success to the separate Indo-European languages and to other languages not related to Indo-European. To summarize the general principles involved:

1. The earliest form must be ascertained as well as all parallel and related forms.

2. Every sound of a given form must be compared with that of the form from which it is derived (often called its etymon).

3. Should there be a deviation in the previously established phonetic correspondences, a rational explanation must be provided; *e.g.*, Eng. *feeble* via O.F. *feble, foible* (now *faible*) comes from Lat. *flēbilis*, "lamentable," but with the first *l* lost through simple dissimilation.

4. Shift in meaning (semantic change) must also be explained; *e.g.*, Eng. *stink* from O.E. *stincan*, "have a smell" (whether good or bad), has acquired a pejorative sense.

5. Words which are isolated in the language ("boycott," "hooligan"), which present nonnative sounds or combinations of sounds ("loch," "axolotl"), or which, more commonly, show marked deviation from the usual phonetic correspondences (*e.g.*, "root," "nay," which betray Scandinavian origin as opposed to "wort," "no") are probably borrowed rather than inherited.

Subsequent investigators have greatly increased the exactitude of etymological methods. Thus, Rudolf Meringer stressed the connection between the history of a word and the history of the thing it represents: the etymology of fencing "foil" (from Fr. *feuille*, "sword blade"), for instance, presumes the special knowledge that this weapon was originally a rough sword blade used for hacking rather than for thrusting. Jules Gilliéron, whose magnificent *Atlas linguistique de la France* (with E. Edmont, 1902–10) founded the technique of linguistic geography, drew attention to the vast diversity of dialect and substandard forms upon which a standard language draws, as well as to the often complex factors which determine the choice of rival words. A complete etymology must take into account the full history of a word as well as its position with regard to other items of vocabulary; *e.g.*, "female" via O.F. *femelle* from Lat. *femella*, diminutive of *femina*, "woman," has been influenced by "male."

The term popular etymology or folk etymology is sometimes used for the alteration of a word through a mistaken idea of its true etymology; *e.g.*, O.E. *brȳd-guma*, literally "bride man" (*guma* is cognate with Lat. *homo*, "man") yielded Eng. *bridegroom* under the influence of "groom." *See* also DICTIONARY; LINGUISTICS.

BIBLIOGRAPHY.—The vocabulary of English, which draws upon so many sources, contains many words of which the etymology is dubious or unknown. For information on English words, *see* the Oxford *New English Dictionary*, corrected reissue in 13 vol. (1933). W. W. Skeat, *Etymological Dictionary of the English Language*, rev. ed. (1910), is still extremely valuable. Two useful books for the student are: F. Holthausen, *Etymologisches Wörterbuch d. englischen Sprache* (1949); E. Weekly, *Concise Etymological Dictionary of Modern English*, rev. ed. (1952). (G. M. M.)

EU, a town of Normandy (in the north of the modern French *département* of Seine-Maritime), on the Bresle river 2 mi. E.S.E. of Le Tréport and 64 mi. N.N.E. of Rouen, important as the seat of an ancient countship. The town, said to date from Roman times, grew round the castle, which may have been built in the Carolingian period to protect the region against the Northmen. The first line of counts is supposed to have been descended from Robert I of Normandy. Its heiress Alix (d. 1227) married Raoul de Lusignan; then through their granddaughter Marie's marriage the countship passed to the house of Brienne. When Raoul de Brienne, constable of France, was beheaded (1350), Eu was confiscated by John II of France and given, in 1352, to John of Artois.

The town had obtained a charter as a commune in 1151, but the wars of the 14th and 15th centuries ruined its prosperity. After John of Nevers, a cousin of the duke of Burgundy, had inherited the countship (1472), Louis XI in 1475 had the town burned down lest it should fall into the hands of Burgundy's English allies.

After passing through heiresses to the houses of Cleves (1491) and of Lorraine-Guise (1633; Henry I of Guise and Catherine of Cleves are entombed in the chapel of the old Jesuit college), the countship was sold in 1660 to Anne Marie Louise d'Orléans, duchesse de Montpensier (*q.v.*). She improved the château and employed Le Nôtre to design its gardens, but in 1681 had to cede the Eu to Louis XIV's natural son Louis Auguste, duc du Maine, whose younger son Louis Charles (1701–75) bore the title comte d'Eu. Subsequently the property passed to the house of Penthièvre (Bourbon-Toulouse) and finally, after confiscation during the Revolution, to that of Bourbon-Orléans, in which the title of comte d'Eu was revived in favour of Gaston (1842–1922), eldest son of Louis, duc de Nemours (*see* BOURBON: *Table IV*). Again confiscated under the second empire, the property was restored to the Orléans family after the Franco-German War.

BIBLIOGRAPHY.—*Le Livre rouge d'Eu*, ed. by A. Legris (1911); S. Deck, *Une commune normande au moyen âge; la ville d'Eu* (1924); J. Duhamel, *Le Château d'Eu* (1955). (MI. M.)

EUBOEA (modern Gr. EVVOIA, formerly often NEGROPONTE), the largest island, after Crete, of the Greek archipelago in the Aegean sea, is about 106 mi. from northwest to southeast and from 4 mi. to 30 mi. in breadth. Together with the northern Sporades (Skyros, Skopelos and Skiathos) it forms a *nomos* or prefecture, with its capital at Chalcis (*q.v.*). Pop. (1961 census) 165,758. Area 1,492 sq.mi. (island, 1,398 sq.mi.). The mountains of Euboea prolong the chains bounding Thessaly to east and south, and continue south into the Cyclades. Three principal masses are separated by fertile lowlands producing vines, olives and cereals. In the north the highest peaks are Mt. Telethrion, or Galitsades (3,182 ft.), and Mt. Xiron (3,251 ft.), the former famed for medicinal plants in antiquity and for hot sulfurous springs near the coast at Aidhipsos, called the Baths of Hercules, used in antiquity by Sulla and still frequented. In the northwest is the promontory of Cenaeum (Kinaion), where the highest point, Lithadha (2,211 ft.), retains the old name of the off-lying Lithadhes (Lichades) Islands.

In the centre, northeast of Chalcis, rises Mt. Dhirfis, or Delphi (5,718 ft.), whose bare summit carries snow till May, with pines and firs below and then chestnuts, limes and planes, one of the most conspicuous peaks in eastern Greece. At the south end, Mt. Oche (Okhi or Ayios Ilias, 4,587 ft.) is visible from Chios. The southwest promontory is Geraestus; the southeast is Kafirevs, the Caphareus dreaded in ancient times, overlooking the storm- and current-swept channel toward Andros. The east coast is rocky and harbourless and in ancient times the main traffic from the north Aegean to Athens used the inshore channels.

In Greek belief Euboea was torn from the mainland by earthquake, and ancient writers record tremors and volcanic activity there. The northern (Oreon) strait narrows to 1¼ mi., and at Chalcis the Evripou (Euripus) channel, divided by a rock, is bridged on the Boeotian side by a causeway and on the Euboean (which is navigable) by a swing bridge. The first bridge there was built by Boeotians when Euboea revolted from Athens in 411 B.C.; they first built a mole, leaving only a small channel which was spanned by a wooden bridge.

Euboea has few streams. In the northwest the Budorus (modern Kirevs) river flows into the Aegean from two headwaters. On the north coast near Istiaia issues the Kollias river, whose rich alluvial plain, like that of the Budorus, was occupied by foreign

settlers after the War of Liberation. South of Chalcis flows the Lelantus (modern Vasiliko) river; whose fertile coast plain was in ancient times a breeding place of horses and a standing cause of feud between Chalcis and its rival city Eretria (*q.v.;* modern Nea Psara, 14 mi. S.E.). The mountains have good pasture for sheep and cattle; the latter seem to have given Euboea its name (*Euboia* would appear to mean "rich in cattle"). Forests have been widely devastated by goats, fires and reckless felling. In the mountains were formerly iron and copper mines, and ancient Chalcis had a reputation for metalworking. Magnesite is worked for export; lignite is mined at Kimi on the east coast, at Nea Psara and 15 mi. E. at Aliverion (Tamynae) for fueling an electric power station built with U.S. aid after World War II. From Karistos in the south comes the green and white *cipollino* marble, much used for building in imperial Rome and still exported.

The population is varied: Greeks have immigrated from the mainland and other islands; the south, like much of Andros, is occupied by Albanians; and there is an element of Vlach population in the hill country. (J. Bo.)

History.—The earliest inhabitants of the island were the Abantes who brought a Bronze Age civilization from central Greece and are named in the catalogue of Achaean ships in the *Iliad* of Homer. The island had in literature a number of other names such as Makris and Doliche ("long island"), Abantis and Ellopia. Chalcis was already an important place in Mycenaean times, for it controlled the Euripus channel and the crossing from the mainland. In classical times the north of the island was occupied by Ellopes, the centre by Ionians and the south by Dryopes. The Ionians excelled in seamanship and traded especially in swords; Chalcis led the colonizing movement to the west and Eretria that to the coast of Thrace in 750–700 B.C. Pithecusae (Ischia) and Cumae in Italy and Naxos in Sicily were the first of many colonies planted under the leadership of Chalcis. The Chalcidic alphabet (*see* ALPHABET) and the name "Graecus" (derived from a Euboean tribe, the Graei) were both eventually adopted by the Romans and so by Europe. Eretrians were the first colonists at Corcyra (Corfu), but fled subsequently to the Albanian coast on the arrival of the Corinthians. They led the large-scale colonization of the Thracian peninsulas known later as Chalcidice.

Euboea's prosperity was checked by a war (*c.* 700 B.C.) for possession of the rich Lelantine plain, from which both the victor Chalcis and the vanquished Eretria emerged weak and exhausted. The belligerents and their allies engaged in hand-to-hand fighting, the use of missiles being forbidden by mutual agreement. Thereafter the Euboean states lost the lead in trade and in colonization to the mainland. When Athens grew strong, Chalcis became allied to Boeotia and Sparta. In 506 B.C. the Athenians captured Chalcis, occupied the Lelantine plain with their own citizens as "cleruchs" (*see* CLERUCHY) and introduced a democratic government in place of the aristocratic class called *Hippobotae.* Meanwhile Eretria and Athens remained friendly; both states helped the Ionians in their revolt (*c.* 499) and incurred the wrath of the Persian king Darius I in 490 B.C., when the Persians subjugated Carystus (modern Karistos) in the south and destroyed the town of Eretria; but sufficient population survived for the Eretrians to man seven ships for the battle of Salamis (480). During the counteroffensive the states of Euboea (Carystus under compulsion) joined the Delian league (*q.v.*), and subsequently fell victim to Athenian imperialism. Euboea revolted from Athens in 446 and cleruchies were planted at Chalcis, Eretria and Histiaea (Oreus; modern Istiaia), and tribute was levied on the other states. It again revolted during the Peloponnesian War in 411.

After the collapse of the Athenian empire at the end of the Peloponnesian War the states of Euboea passed from the control of one leading state to that of another (Sparta, Athens, Thebes, Macedonia, Rome, Syria, Pontus and so on) because the island was strategically important and relatively weak. A league of Euboean states, formed in the second half of the 4th century B.C., had a long, if interrupted, existence. Chalcis appeared as its leader in the 3rd century A.D. and in the 6th century the bridge to the mainland from Chalcis was repaired by the Eastern Roman emperor Justinian I. Although in the partition of 1204, whereby the eastern empire was divided after the fourth crusade (*see* CRUSADES), Euboea had been assigned to the Venetians, it was usurped in 1205 by Boniface of Montferrat (then king of Thessalonica) as part of his conquest of Greece, and he divided it into three fiefs (Chalcis, Carystus, Oreus) which he granted to three Veronese nobles as *terzieri,* "triarchs," including Ravano delle Carceri, who in 1209 was recognized by the Venetians as their sole lord. From this period dates the corruption of the name of the Euripus channel (Egripo, Negripo expanded to Negroponte) and its extension to the city of Chalcis and to the island of Euboea as a whole. When at the end of the 14th century the families of the *terzieri* died out, Venice achieved complete control of the island. The Turks took the island from the Venetians in 1470; in 1830 it became part of the kingdom of Greece. (N. G. L. H.)

EUBULIDES (4th century B.C.), a member of the Megarian school of philosophy (*q.v.*), was born in Miletus. He was a contemporary of Aristotle, whom he attacked. Tradition says that Demosthenes was one of his pupils. Eubulides was renowned as an inventor of logical paradoxes, presumably suggested by the arguments of Zeno. One of these, "The Liar" ("Does a man who says that he is now lying, speak truly?"), embodies a more serious difficulty than the rest and has a long history (*see* LOGIC, HISTORY OF). (D. J. A.)

EUBULUS (4th century B.C.), an Athenian statesman, was noted for his able financial administration. He first became prominent in 355 B.C. when the imperialistic party was discredited (*see* GREECE: *Ancient History*). From then until 346 B.C. he was the most influential politician at Athens. His administration as chief commissioner of the theoric fund, which provided free seats at public spectacles, gave him general control over Athenian finances since the commissioners of the fund were now given power in other departments of finance. His astute supervision of state expenditure restored the economic position of Athens without increasing the burden of taxation. Under his guidance the efficiency of the fleet was improved, docks and fortifications repaired, and various public services carried out. His financial policy was based on the assumption that Athens should not be involved in major wars. This explains his pacific attitude toward Philip II of Macedonia, which was bitterly opposed by Demosthenes (*q.v.*).

Eubulus was largely responsible for the overtures to Philip which led to the Peace of Philocrates in 346 B.C. With this peace his political ascendancy ended. His last recorded public activity was in 343 when he helped to secure Aeschines' acquittal in the case brought against him by Demosthenes for treasonable conduct during the negotiations with Philip in 346 B.C.

See G. Glotz, *Histoire Grecque,* vol. iii (1936); N. G. L. Hammond, *A History of Greece,* book v, ch. 3 and 4 and book vi, ch. 1 (1959). (H. L. H.-W.)

EUCALYPTUS, a large genus of trees of the family Myrtaceae (*q.v.*), indigenous, with only a few exceptions, to Australia and Tasmania. In Australia the eucalypti are commonly called "gum trees" or "stringybark" trees. The genus, from the evidence of leaf remains, appears to have been represented by several species in Eocene times. The leaves are leathery, hang obliquely or vertically, and are studded with glands. The petals cohere to form a cap which is discarded when the flower expands. The fruit is surrounded by a woody cup-shaped receptacle and contains very numerous minute seeds. Possibly the largest fruits, from 2 to 2½ in. in diameter, are borne by the shrub *E. macrocarpa,* also known as mottlecah or silverleaf eucalyptus.

BY COURTESY OF AUSTRALIAN NEWS AND INFORMATION BUREAU

FLOWERS OF THE MOTTLECAH (EUCALYPTUS MACROCARPA)

The eucalypti are rapid in growth, and many species are of

BY COURTESY OF AUSTRALIAN NEWS AND INFORMATION BUREAU

SALMON GUM TREE (EUCALYPTUS SALMONOPHLOIA)

great height. *E. regnans,* the giant gum, of Victoria, one of the tallest trees known, attains a height of about 300 ft. and a girth of 25 ft., thus equaling the California big tree *(Sequoia gigantea)* in height though not in massiveness. *E. globulus* is the blue gum tree of Tasmania. The flowers are single or in clusters, and nearly sessile.

After 1854 the eucalyptus was successfully introduced into the south of Europe, Algeria, Egypt, Tahiti, New Caledonia, Natal and India. It is extensively planted in California and throughout South America, often for its durable, pest-resistant timber. The young twigs of *E. cordata* and *E. pulverulenta,* both grown in California, are sold as "blue spiral" by most American florists.

Uses.—The leaf glands of many species of *Eucalyptus,* especially *E. salicifolia* (formerly known as *E. amygdalina*) and *E. globulus,* contain a volatile, aromatic, straw-coloured oil, known as eucalyptus oil. This is extracted by steam distillation and should, according to the American and British pharmacopoeias, contain at least 70% of eucalyptole, which is technically known as cineole. Eucalyptole also contains a crystallizable resin, allied to kino. It has the odour of camphor, is pungent, cooling, and has a spicy taste. It is insoluble in water but soluble in 60% alcohol. Its chief use is as an expectorant, to encourage spitting of mucus or fluid from the trachea and lungs in bronchitis and tuberculosis. Eucalyptus oil is also used in medicine, chiefly as an inhalant to give symptomatic relief in certain respiratory difficulties.

E. globulus, E. resinifera and other species yield what is known as Botany Bay kino, an astringent dark-reddish amorphous resin, which is obtained in a semifluid state by making incisions in the trunks of the trees. *E. gigantea* yields a notable proportion of gum. From the leaves and young bark of *E. mannifera* and *E. viminalis* is procured Australian manna, a hard, opaque, sweet substance, containing melitose. On destructive distillation the leaves yield much gas, 10,000 cu.ft. being obtained from one ton. The wood is extensively used in Australia as fuel and the timber is of remarkable size, strength and durability. Among the many species of timber-yielding eucalypti are *E. salicifolia,* the wood of which splits with remarkable facility, *E. botryoides,* hard, tough and durable and one of the finest timbers for shipbuilding, *E. diversicolor* or karri, *E. globulus, E. leucoxylon* or ironbark, *E. marginata* or jarrah (*see* JARRAH WOOD), *E. obliqua, E. resinifera, E. siderophloia* and others. The timber is often useful for building, fencing, etc.

The bark of different species of eucalyptus has been used in papermaking and tanning.

See also references under "Eucalyptus" in the Index volume.

BIBLIOGRAPHY.—R. T. Baker and H. S. Smith, *A Research on the Eucalyptus Especially in Regard to Their Essential Oils* (1920); R. N. Parker, *Eucalyptus Trials in the Simla Hills, Calcutta* (1925); H. W. Younken, *Text-Book of Pharmacognosy* (1943). W. Y. Blakely in *Key to the Eucalyptus* (1955) lists 522 species and 160 varieties.

(N. TR.; X.)

EUCHARIS, a genus of plants of the family Amaryllidaceae (*q.v.*), containing six species, natives of Colombia. *Eucharis grandiflora,* the Amazon lily, is the most generally cultivated species. It is bulbous, with broad-stalked leaves, and an erect scape 1½ to 2 ft. long, bearing an umbel of three to ten large, white, showy flowers. The flowers resemble the daffodil in having a prominent central cup or corona. The plant is propagated by removing the offsets, which may be done in spring, potting them singly in six-inch pots. It requires good loamy soil, with sand enough to keep the compost open, and should have a good supply of water.

Other kinds found in some gardens are *E. bakeriana, E. mastersi, E. lowi* and *E. sanderi;* mostly mere forms of *E. grandiflora.*

EUCHARIST (from the Greek for "thanksgiving"), the central rite of Christian worship. It is also called the Lord's Supper, Holy Communion, the Mass, the Divine Liturgy, the Blessed Sacrament or simply the Sacrament. Together with prayer, teaching and baptism, it has been an indispensable component of the Christian service since the earliest days of the church.

The piety and the theology of the church have found a variety and richness of meaning in the Eucharist. First one eucharistic theme, then another, has predominated in Christian thought and liturgical practice. It is one of the supreme ironies of Christian history, evident already in the New Testament, that the Eucharist, intended as a means of fostering the unity of the church, has been a source of disunity and contention as well. But this must not be permitted to obscure the real areas of consensus among most Christians in the interpretation of the Eucharist. All Christians would agree that it is a memorial action in which, by eating and drinking, the church calls to remembrance what Jesus Christ was, said and did. There is also universal agreement among Christians that participation in the Eucharist enhances and deepens the communion of believers not only with Christ but also with one another. The breaking of the bread and the pouring out of the wine are recognized by every Christian denomination as the central symbols of the death of Jesus Christ on the cross. Augustine's designation of the sacraments as a "visible word" expresses the idea, shared by Roman Catholics, Orthodox and Protestants, that the Eucharist communicates the same gospel that comes through the written and the spoken forms of the Word of God. And most Christian traditions teach that Jesus Christ is present in the Eucharist in some special way, disagree though they do about the mode, the locus and the time of that presence. In short, there is more of a consensus among Christians about the meaning of the Eucharist than would appear from the confessional debates over the sacramental presence, the effects of the Eucharist and the proper auspices under which it may be celebrated.

Behind this Christian consensus is the witness of the New Testament to the eucharistic faith and practice of the early church. The story of the institution of the Eucharist by Jesus on the last night before his crucifixion is one of the few incidents in his life reported by four books of the New Testament—the Gospels of Matthew (xxvi, 26–28), Mark (xiv, 22–24) and Luke (xxii, 17–29) and I Corinthians (xi, 23–25). The epistles of Paul, especially I Corinthians, and the Book of Acts (ii, 42, 46) make it clear that early Christianity regarded this institution as a mandate to "do this in remembrance of" Christ, gathering around the cup and the broken bread and learning to know, even in this present life, the joys of the heavenly banquet that was to come in the kingdom of

God. "As often as you eat this bread and drink the cup, you proclaim the Lord's death until he comes"; the past, the present and the future came together in the Eucharist. There is some reason to believe, on the basis of Acts, that the early church celebrated the Eucharist daily, but the evidence is not conclusive. On the other hand, the very form of the words of institution in the Gospels and I Corinthians indicates, despite variations in the texts, that a formula of "blessing" or consecration was an irreducible minimum in the earliest celebration. As part of the Christian mysteries, the Eucharist was kept away from the eyes of the unbelieving world. Often, if not always, it was celebrated in the setting of a common meal. Nowhere in the New Testament is eucharistic doctrine isolated from the practice of obedience to the institution of Christ; nowhere is the practice observed or urged without reference to the doctrine of Christ. Such an interrelation between doctrine and practice characterizes the approach of all the major Christian groups to the meaning of the Eucharist, as an examination of these groups will disclose.

According to the eucharistic doctrine of Roman Catholicism, the elements of bread and wine are "transubstantiated" into the body and blood of Christ; *i.e.*, their whole substance is converted into the whole substance of the body and blood, although the outward appearances of the elements, their "accidents," remain (*see* TRANSUBSTANTIATION). This eucharistic doctrine gives support to certain eucharistic practices and is supported by them in turn. For example, the adoration and reservation of the host follow from the belief that the whole Christ is truly present in his body and blood under the forms of bread and wine. Similarly, the presence of the whole Christ in either of the consecrated elements justifies the practice of distributing only the host, not the chalice, to the laity. From the affirmation of the real presence comes substantiation for the sacrifice of the Mass: Christ sacrificed himself on Calvary once, but the Mass presents that one sacrifice again—*i.e.*, "re-presents" it. Both by reasserting the doctrine of transubstantiation and by redefining the meaning of the eucharistic sacrifice with greater precision, the Council of Trent continued the theological work of Thomas Aquinas and laid down the lines for further theological development. During the 19th and 20th centuries the Roman Catholic liturgical movement (*q.v.*) put new emphasis upon the frequency of communion, upon the participation of the entire congregation in the priestly service and upon the real presence of Christ in the church as the fundamental presupposition for the real presence in the Eucharist. These were not new ideas, but they received new life through the issuance of *Mystici Corporis* in 1943 and *Mediator Dei* in 1947, both by Pope Pius XII. So closely are eucharistic practice and eucharistic doctrine intertwined in Roman Catholicism that the liturgy of the Mass has been interpreted as a commentary upon the image of the church as the mystical body of Christ.

The eucharistic beliefs and practices of Eastern Orthodoxy have much in common with those of Roman Catholicism. One of the differences in practice, the use of leavened rather than of unleavened bread, is based upon differing interpretations of when the Eucharist was instituted, whether on the last day of leavened bread or on the first day of unleavened bread in the Jewish calendar. Eastern Orthodox faithful receive both the consecrated elements in the Eucharist, but by the method of intinction; *i.e.*, the dipping of the consecrated bread in the cup. While Roman Catholic theology maintains that the recitation of the words of institution constitutes the Eucharist as a sacrament, Eastern theology has taught that the invocation of the Holy Spirit upon the elements (Gr. *epiklesis*) is the essential form of the Eucharist. In the area of eucharistic doctrine, Eastern Orthodox theology has not achieved the elaboration and precision of Western doctrine. There is a Greek term corresponding to transubstantiation (*metousiosis*), but it is evident that the term was adopted by Eastern Orthodoxy in the course of theological discussions with the Western Church. The eucharistic liturgies of Eastern Christendom are a highly dramatic representation of the incarnation and resurrection of Christ as the transformation of the created world through the victory of life over death. It is probably accurate to say that the principal differences between Eastern Orthodoxy and Roman Catholicism on the Eucharist are in the area of piety and liturgy rather than in doctrine.

Of the various theological tendencies in non-Roman Catholic Western Christendom, the two that adhere most closely to the traditions of Catholic eucharistic doctrine and practice, Eastern and Western, are the Anglican and the Lutheran. Early Anglican theology vigorously opposed Roman Catholic teaching on the sacraments, sometimes even identifying Anglican with Reformed theology (*see* below). But from the beginning, and especially since the Oxford movement, Anglican liturgical practice has kept more of the Catholic tradition than has Reformed practice, and in the 19th century Anglican theology moved toward traditional ways of speaking about the Eucharist. The theology of Lutheranism in the 16th century unequivocally affirmed the real presence of the body and blood of Christ "in, with, and under" the bread and wine in the Eucharist. The term "consubstantiation," while not officially approved by Lutheran theologians, did summarize the Lutheran alternative to the idea of transubstantiation. In their liturgies both Anglicanism and Lutheranism worked within the framework of the Western liturgy of the Mass, adopting certain elements and rejecting others; the liturgical movements in both communions during the 19th and 20th centuries restored additional elements of the eucharistic tradition, even though the theological interpretation of the Lord's Supper continued to display great variety.

More radical in its rejection of traditional definitions of the eucharistic presence and in its suspicion of traditional liturgical practices, Reformed Christianity replaced the altar with the communion table and subordinated the sacraments to the preached and written word of the gospel. In the theology and practice of Huldreich Zwingli the memorial aspect of the eucharistic celebration received new emphasis, overshadowing if not excluding other aspects. But the more characteristically Reformed doctrine was that of John Calvin, who taught a "real but spiritual presence" of the living Christ, but in the sacramental action rather than in the elements (*see* CALVIN, JOHN: *Theology*). The liturgical observance introduced by Calvin was intended to eliminate from the Lord's Supper what Calvin regarded as the superstitious practices of the Mass and to restore the simplicity of the gospel. The doctrine of the Lord's Supper became the occasion for the most serious of the confessional debates between Reformed and Lutheran churches. From the Colloquy at Marburg in 1529 to the Barmen declaration in 1934 the controversy over the Eucharist has been a dominant theme of Lutheran-Reformed discussions.

Other "Free church" traditions have drawn upon Reformed thought and practice but have been even more thoroughgoing in their attempt to reform eucharistic doctrine and practice. The sacraments have become "ordinances," not channels of grace but expressions of the faith and obedience of the Christian community. Among Baptists the practice of "close communion" has restricted the ordinance to those who are baptized properly; *i.e.*, as adults upon a personal profession of faith. The Society of Friends dropped the use of the Eucharist altogether, as well as the practice of baptism, in its reaction against formalism.

As a result of all this variation among Protestants in both doctrine and practice, the Eucharist has been a central issue in the discussions and deliberations of the ecumenical movement. Strict Baptists cannot receive the Eucharist outside the community of the properly baptized; strict Lutherans cannot receive it where the real presence of the body and blood of Christ in the elements is not affirmed; strict Anglicans cannot receive it where the minister has not been properly (*i.e.*, episcopally) ordained. Together with the social upheavals after World War II, the ecumenical movement compelled reconsideration of the problem of intercommunion, which was one of the chief subjects at the third World Conference on Faith and Order held in Lund, Swed., in Aug. 1952. In 1958 the commission on the Eucharist appointed by the Evangelical Church in Germany presented its report, the "Arnoldshain Theses," attempting to transcend the traditional differences between Reformed and Lutheran doctrine by a new appeal to biblical thought and language. And at the various meetings of the World Council of Churches several of the churches have sponsored communion services, to which others have been

invited as observers and sometimes as recipients.

Thus both the unity of the church and the disunity of the churches continue to be symbolized by the Eucharist, as they have been since the beginnings of Christianity.

See Mass; Sacrament; *see* also references under "Eucharist" in the Index volume.

Bibliography.—Yngve Brilioth, *Eucharistic Faith and Practice, Evangelical and Catholic* (1934); Anscar Vonier, *A Key to the Doctrine of the Eucharist* (1960); Marie Joseph Nicolas, *What Is the Eucharist?* (1960); Donald Baillie and John Marsh (eds.), *Intercommunion* (1952).
(J. J. Pn.)

EUCHRE, a card game played with 32 cards; the standard 52-card deck with all ranks from deuce to six, inclusive, eliminated. For the quarter century preceding the introduction of auction bridge (*c.* 1905) euchre was the most popular family card game in the U.S., principally in the four-hand partnership game here described.

In general, the cards rank: A (high), K, Q, J, 10, 9, 8, 7; but in the trump suit the knave (jack), called the "right bower," is the highest trump, and the knave of the other suit of the same colour, called "left bower," becomes the second highest trump. Thus the trump suit always contains nine cards; the next suit (same colour as the trump suit) contains seven; each cross suit (opposite colour to trumps) contains eight.

Five cards are dealt to each player in rounds of three and two at a time. The next card is turned face up. Each player in turn, beginning with the eldest hand (player on the dealer's left), has the right to accept or reject the suit of the turned card as trump. If any player accepts, the dealer discards one card from his hand face down, and takes the turned card instead. If all four players pass, rejecting the proposed trump, it is turned down under the pack. Thereafter, beginning with eldest hand, each player in turn has the right to "make it" (name another suit as trumps) or to pass. If all four players pass in the second round, the hands are thrown in.

The player who makes the trump has the right to declare, "I play alone," whereupon his partner turns his cards face down and takes no part in the play but shares in the winnings. Eldest hand makes the first lead, regardless of the position of the maker, except that if the maker plays alone the opening lead is made by the player at the maker's left.

The rules of play are as in whist (*q.v.*). Any card may be led. Each hand must follow suit if able, but if unable may trump or discard at will. A trick is won by the highest trump, or by the highest card of the suit led. Winner of each trick leads to the next.

The object of play is to win at least three tricks. The maker's side is euchred if it fails to do so. Winning all five tricks is "march." Maker's side scores 1 point for winning three or four tricks; 2 points for march; 4 points for march won when playing alone. For euchre, opponents score 2 points. Game is won by first side to score 5 points, but sometimes by agreement game is increased to 7 or 10.

Cutthroat Euchre is a three-handed game in which the maker always plays alone against the others in temporary partnership. The scoring is the same as in the four-handed game except that maker scores only 3 points for march.

Two-handed Euchre is played with 24 cards, the normal euchre pack being reduced by eliminating the sevens and eights. The 24-card pack also is sometimes used for the four-handed game, speeding up the play.

Railroad Euchre is a form of euchre in which local rules and variations are introduced. These might include the use of the joker as a tenth, and highest, trump; allowing one of the opponents to oppose alone; allowing lone players to discard a bad card and replace it with the best one from his partner's unused hand; permitting any score above five points to be carried forward into the next game, etc.

Auction Euchre is played with five, six or seven hands. Highest bidder makes trumps and selects one or more partners, depending on the size of his bid.

See A. H. Morehead, R. L. Frey, G. Mott-Smith, *The New Complete Hoyle* (1956).
(R. L. Fy.)

EUCKEN, RUDOLF CHRISTOPH (1846–1926), German idealistic philosopher and Nobel prize winner whose early works deal with the philosophy of Aristotle, but whose later ones are mainly concerned with ethical and religious problems, was born on Jan. 5, 1846, at Aurich in East Friesland. He studied at Göttingen under R. H. Lotze and at Berlin under F. A. Trendelenburg, whose ethical tendencies and historical treatment of philosophy attracted him. Eucken was professor of philosophy at the University of Basel, Switz., from 1871 to 1874 and at Jena, Ger., from 1874 to 1920. In 1908 he was awarded the Nobel prize for literature. He died at Jena on Sept. 14, 1926.

Rejecting naturalism (on the ground that the feeling of *ought* is inexplicable if man is only a derivative of natural processes) and distrusting pure intellectualism, Eucken maintains that man is the meeting place of nature and of spirit and that it is his duty and privilege to overcome his nonspiritual nature by incessant active striving after the spiritual life which involves all faculties, especially will and intuition. Historical religion is often merely a means of consolation, but true culture advances both form and energy and seeks a total coherency.

The Nobel prize diploma's eulogy mentioned Eucken's "warmth and strength in presentation with which in his numerous works he has vindicated and developed an idealistic philosophy of life." However, he was regarded by many persons as a weak selection.

EUCLASE, a rare mineral, occasionally cut as a gem stone. Like beryl, it is a silicate of beryllium and aluminum, but hydrogen is also present; the formula is $Be(AlOH)SiO_4$. The colour is generally pale blue or green, though sometimes the mineral is colourless. When cut it resembles certain kinds of beryl (aquamarine) and topaz, from which it may be distinguished by its higher specific gravity (3.1). Its hardness (7.5) is rather less than that of topaz.

Euclase crystallizes in the monoclinic system as striated prisms terminated by acute pyramids. Cleavage is perfectly parallel to the plane of symmetry, and this renders the stone fragile with a tendency to chip, hence its name from the Greek words for "easily fractured."

Euclase occurs with topaz near Ouro Preto, Braz., in the Austrian Alps and with topaz and chrysoberyl in the gold-bearing gravels of the Sanarka river in the southern Urals.

EUCLID (Eukleides) (*c.* 450–368 B.C.), founder of the Megarian school of philosophy, was born at Megara and lived and taught there. He was present at Socrates' death in Athens (Plato, *Phaedo* 59 c); and some members of the Socratic circle, including Plato, are said to have withdrawn afterward to Megara. Euclid and his companion Terpsion appear in the prologue to Plato's *Theaetetus,* from which reference it is probable that he was still living in 368 B.C.

The main inspiration of Euclid's philosophy was clearly Socratic, but he and his school were sometimes ranked as followers of Parmenides on account of their insistence on the unity of all goodness and the high value which they attached to dialectical skill. *See* Megarian School of Philosophy. (D. J. A.)

EUCLID (in Greek Eukleides), Greek mathematician (fl. 300 B.C.) whose text was one of the great successes of world literature. Of his life it is only known that he taught and founded a school at Alexandria in the time of Ptolemy I, who reigned from 306 to 283 B.C. Medieval translators and editors mostly called him Megarensis, through confusion with the philosopher Eukleides of Megara, Plato's contemporary. Proclus tells the story of Euclid's reply to King Ptolemy, who asked whether there was any shorter way in geometry than that of the *Elements*—"There is no royal road to geometry." Another anecdote relates that a pupil, after learning the very first proposition in geometry, wanted to know what he would get by learning these things, whereupon Euclid called his slave and said, "Give him threepence since he must needs make gain by what he learns."

Euclid's great work is the *Elements* (*see* Geometry), in 13 books; of the books formerly purporting to be books xiv, xv, the first, by Hypsicles (2nd century B.C.), adds some interesting theorems about the regular solids, two of which it attributes to Aristaeus and Apollonius respectively; the second was written, at

least in part, by a pupil of Isidorus of Miletus in the 6th century A.D.

The names of some earlier compilers of *Elements* are handed down; the first was Hippocrates of Chios (5th century B.C.), famous for his quadrature of certain lunes; the latest before Euclid was Theudius, whose textbook was that in use in the Academy and was probably that from which Aristotle drew his illustrations. The older elements were at once superseded by Euclid's and then forgotten. For his subject matter Euclid doubtless drew upon all his predecessors; but it is clear that the whole design of his work was his own; he evidently altered the arrangement of whole books, redistributed propositions among them and invented new proofs where the new order made the earlier proofs inapplicable. Thus, while book x is mainly due to Theaetetus, the proofs of several theorems in this book had to be changed in order to adapt them to the new definition of proportion due to Eudoxus.

We know from the commentary of Proclus that Euclid incorporated into his work many discoveries of Eudoxus and Theaetetus. Most probably books v and xii are essentially due to Eudoxus, x and xiii to Theaetetus. Book v expounds a wonderful theory of proportion applicable to commensurable and incommensurable magnitudes alike. The main theorems of book xii state that circles are to one another as the squares, the spheres as the cubes of their diameters, and that the volume of a pyramid is one-third of the volume of a prism having the same basis and the same height. These theorems are certainly due to Eudoxus; they are proved by the "method of exhaustion." Book x deals with irrationals of different classes. Apart from some new proofs and additions, the contents of book x is due to Theaetetus, and so is the whole book xiii, in which the five regular solids are constructed. Euclid seems to have incorporated a finished treatise of Theaetetus on the regular solids into his *Elements*. Just so, book vii, dealing with the foundation of arithmetic, is a self-consistent treatise, written most probably before 400 B.C.

Other books of the *Elements* are not on this high mathematical level. In book viii, the second of the three arithmetical books, we find cumbersome enunciations, needless repetitions and even logical fallacies. Apparently Euclid's exposition is excellent only in those parts where he had excellent sources at his disposal. Euclid may not have been a first-class mathematician. However, he certainly was a first-rate teacher of mathematics. His textbook has remained in use practically unchanged for more than 2,000 years. It was not until the middle of the 19th century that a so-called "away from Euclid" movement began, which led to the appearance of a multitude of rival textbooks of geometry giving the substance of Euclid's early books in so many different forms as to produce a state of chaos in geometrical teaching.

The following are the other extant works of Euclid; the first two belong to elementary geometry:

1. The *Data*, containing 94 propositions, has for its object to prove that if certain elements in a figure are given, then other things are given, *i.e.*, can be determined. The following is an example: If, in a parallelogram with a given angle, the sum or the difference of the two sides and the area are given, both sides will be given.

2. A book *On divisions (of figures)* was discovered in Arabic at Paris and edited by F. Woepcke in 1851. John Dee was the first to find (in Latin) a similar treatise by one Mohammed Bagdadinus which, in 1563, he handed to Commandinus, who published it in Dee's name and his own in 1570. The genuine treatise has now been restored and edited by R. C. Archibald (1915). The type of problem dealt with is that of dividing a given figure (*e.g.*, a triangle, a parallelogram, a quadrilateral, a circle or a figure bounded by an arc of a circle and two straight lines) by one or more straight lines into parts equal, or having given ratios, to one another or to other given areas.

3. The *Optics* of Euclid is extant in Greek in two forms, one being Euclid's own treatise and the other a recension by Theon. The *Catoptrica* (edited by J. L. Heiberg in the same volume) is not by Euclid but is a later compilation from ancient works on the subject.

4. The *Phaenomena*, extant in Greek, is a treatise on the geometry of the sphere intended for use in astronomy, and is similar in content to the work of Autolycus *On the Moving Sphere*.

5. The work *Elements of Music* is attributed to Euclid by Proclus and Marinus. Of two extant treatises of the kind the first, *Sectio Canonis*, gives the Pythagorean theory of music with some later additions. The *Introductio harmonica* is not by Euclid but by Cleonides, a pupil of Aristoxenus.

Of lost geometrical works by Euclid all except one, the *Pseudaria*, belonged to higher geometry.

1. The purpose of the *Pseudaria* was, it is said, to distinguish, and to warn beginners against, different types of fallacies to which they are liable in geometrical reasoning.

2. The *Porisms*, in three books, was an advanced work of which Pappus gives a summary account. (*See* PORISMS.)

3. The *Conics*, in four books, corresponded in content to the first four books of Apollonius' *Conics*, though Apollonius added new theorems and generalized the treatment throughout. Euclid still called the conics by their old names, "Sections of a right-angled cone, an obtuse-angled cone, and an acute-angled cone" respectively; it was Apollonius who first gave them the names "parabola," "hyperbola," "ellipse," arising out of his generation of them all.

4. Regarding the *Surface-Loci*, in two books, mentioned by Pappus, we can only conjecture that the loci dealt with were loci on surfaces, perhaps also loci which are surfaces. Conics would appear to have entered into the subject, for one of the lemmas to the treatise given by Pappus contains a complete proof of the focus-directrix property of all three conics.

A fragment in Latin, *De levi et ponderoso*, included in Gregory's edition of Euclid, contains a statement of the principles of Aristotle's dynamics, but is not by Euclid.

Translations and Editions.—This notice of Euclid would be incomplete without some account of the earliest and the most important translations and editions of the *Elements*. In ancient times Heron and Pappus of Alexandria, Proclus and Simplicius, all wrote commentaries. Theon of Alexandria (4th century A.D.) brought out a new recension of the work, with textual changes and some additions; Theon's version was the basis of all published Greek texts and translations until, early in the 19th century, Peyrard discovered in the Vatican the great Greek manuscript 190 containing an ante-Theonine text. Arabic translations were made: (1) by al-Hajjaj ben Yusuf ben Matar, first for Harun al-Rashid (786–809) and again for Abdullah al-Ma'Mun (813–833); (2) by Ishaq ben Hunain; this translation was revised by Thabit ibn Kurra (d. 901); (3) by Nasir al-Din al-Tusi (b. 1201).

The first extant Latin translation of the *Elements* was made (about 1120) by Adelard of Bath, who obtained a copy of an Arabic version in Spain, where he went disguised as a Moslem student. Adelard also composed an abridged version and an edition with commentary. Hermann of Carinthia translated books i–xii from the same Arabic version. Gerard of Cremona translated the "15 books of Euclid" from the Ishaq-Thabit version. The first Latin translation to be printed was that of Johannes Campanus (13th century).

The first translation from the Greek was made by Bartolomeo Zamberti and appeared at Venice in 1505. The most important Latin translation is that of Commandinus (1572). The *editio princeps* of the Greek text was brought out at Basle in 1533 by Simon Grynaeus.

The first edition of the complete works of Euclid was the Oxford edition (1703), in Greek and Latin, by David Gregory. All texts are now superseded by *Euclidis opera omnia* (8 vol. and a supplement, 1883–1916), edited by J. L. Heiberg and H. Menge. The first English translation of the *Elements* was that of Sir Henry Billingsley.

The many later editions include Robert Simson's in Latin and English, containing the *Elements*, books i–vi, xi, xii, and the *Data* (1756), and T. L. Heath's *The Thirteen Books of Euclid's Elements*, with introduction and commentary, 3 vol. (1908; 2nd ed., 1926).

An excellent German translation with commentary by G. Thaer appeared in Ostwald's *Klassiker der exakten Wissenschaften*, vol.

235, 236, 240, 241, 243. *See* also references under "Euclid" in the Index volume.

BIBLIOGRAPHY.—T. L. Heath, *History of Greek Mathematics*, vol. i, pp. 354–446 (1921); E. J. Dijksterhuis, *De Elementen van Euclides*, 2 vol. (1929–30); A. Lejeune, *Euclide et Ptolémée, deux stades de l'optique géométrique gracque* (1948); M. Clagett, "The Medieval Latin Translations From the Arabic of the Elements of Euclid," *Isis*, vol. 44 (1953); B. L. van der Waerden, *Science Awakening* (1954).
(T. L. H.; B. L. v. D. W.)

EUCLID, a city of Cuyahoga county in northern Ohio, U.S., 11 mi. N.E. of Cleveland (*q.v.*), and a part of the Cleveland standard metropolitan statistical area. Pop. (1960) 62,706; for comparative population figures *see* table in OHIO: *Population.*

The second settlement (1798) in the Connecticut Western Reserve, Euclid is located on the narrow plain between the Appalachian plateau and Lake Erie; roads built on two beaches formed by receding glacial lakes provided east-west access. It was named for the Greek mathematician by surveyors in the party of Moses Cleaveland, an agent of the Connecticut Land company who directed the survey of the Western Reserve.

Outstripped by Cleveland's advantages in transportation, Euclid remained rural, noted chiefly for its grape culture, until the electric interurban and automobile encouraged suburban development. It became a city in 1931.

Large tracts of level land, unlimited water supply, two railroads and five highways offered opportunities for rapid industrial and urban growth after 1940. Chief manufactures are earth-moving equipment, farm tractors, copper and brass products, electric welders, business machines, airplane and automobile parts.

Euclid has a city library, a county library and a hospital. Parks include Euclid beach, a privately-owned amusement centre of 90 ac.; three city parks and the Euclid Creek reservation in the Cleveland metropolitan park system. (M. M. WA.)

EUCRATIDES, king of Bactria in the 2nd century B.C. As for the rest of the history of the Bactrian kingdom, the sources for the reign of this king are meagre; there are a number of coins, and short and vague notices in two ancient authors (Strabo and Justin). According to Justin, Eucratides came to power in Bactria at about the same time as Mithradates I in Parthia. This does not enable one to fix a date even approximately; but for other reasons some time between the years 170 and 165 seems probable. On one of his coins figure the portraits of a man and wife, Heliocles and Laodice; it is agreed that these are the parents of Eucratides. W. W. Tarn argued that Laodice (who wears the royal diadem on the coin) was a Seleucid princess and that Heliocles had been governor of the upper satrapies with headquarters in Babylonia. After thus identifying Eucratides as a cousin of Antiochus IV Tarn explained his career as that of an agent of the Seleucid king; his coming to power in Bactria would be, on this supposition, the result of an invasion. Other scholars have assumed that he was a Bactrian official (possibly related to the Bactrian royal house through his mother) who raised a rebellion within the country during the absence of the king Demetrius in India, just as Euthydemus (*q.v.*) had rebelled against Diodotus II. The evidence is so slight that it can be made to fit either of these theories. It may be noted that if Eucratides was in fact a Bactrian rebel and unrelated to the Seleucids, that need not exclude the possibility that he enjoyed the support of Antiochus IV and acted to some extent as an instrument of that king's policies (*see* ANTIOCHUS). Whether it is a question of a rebellion or an invasion, the movement against Demetrius presupposes support inside Bactria.

The further course of events becomes clearer: Demetrius returned from India to combat Eucratides in Bactria; Eucratides and 300 followers were besieged by an army of (it is said) 60,000 for some months; but finally Eucratides succeeded not only in breaking the siege but in seizing the kingdom. Once in control of Bactria Eucratides passed into India to claim possession of Demetrius' territory there. During his absence it seems likely that the Parthians annexed two of the northwestern Bactrian provinces. When returning from India to challenge them he is said to have been murdered by his son. His death is probably to be dated to about 159.

See BACTRIA; DEMETRIUS (king of Bactria). (R. H. SI.)

EUCRYPHIA, the only genus of the family Eucryphiaceae, comprises several trees or shrubs used as ornamentals. They have simple or compound opposite leaves, showy white hypogynous flowers with four sepals and four petals, an abundance of distinct stamens in many whorls and a single pistil whose ovary is composed of 5–18 carpels. The fruit is a woody or leathery capsule containing many winged seeds.

The presence of this genus in South America and in Australia suggests a vegetation present long ago when there was either a direct or an interrupted land contact between these two continents.

See W. J. Dress, "A Review of the Genus Eucryphia," *Baileya*, 4:116–127 (1956). (G. H. M. L.; X.)

EUDAEMONISM, in ethics, the name applied to theories which find the chief good for man in happiness. The Greek word *eudaimonia* means literally the state of being under the protection of a benign spirit, a "good genius"; and "happiness" is not at all an adequate translation of this word. Happiness, indeed, is usually thought of as a state of mind which results from or accompanies some of our actions; but Aristotle's answers to the question "What is *eudaimonia*?" (namely, that it is "activity in accordance with virtue"; or that it is "contemplation") show that for him *eudaimonia* is not a state of mind consequent on or accompanying certain activities but a name for these activities themselves. "What is *eudaimonia*?" is the same question as "What are the best activities of which man is capable?"

To say that man's chief good is *eudaimonia* is a tautology and not an answer to an ethical question. We should say that a man lives under a good spirit if he lives a good life without frustration in reasonable material surroundings, with friends and a family. So to live *is* the same thing as *eudaimonia*. Aristotle, on the other hand, sharply distinguished pleasure from the activities which it accompanies.

Later moralists, however, for instance Jeremy Bentham and J. S. Mill, defined happiness as pleasure and the absence of pain (*see* HEDONISM); and there is some justification for this in the English usage of the word "happiness." Others, still regarding happiness as a state of mind resulting from certain activities, tried to distinguish it from pleasure on the grounds that it is mental not bodily, enduring not transitory, rational not emotional. But none of these distinctions is sound. Pleasure may accompany mental as well as bodily conditions; it may continue and not perish; it may be found in rational pursuits as well as in the satisfaction of emotional needs. In fact "happiness" is one of the most elusive of words. It may even be held that happiness is not a state of mind of which we are introspectively aware, resulting from or accompanying certain actions. To say that a man is happy is to say that he is fully absorbed in what he is doing and is acting without frustration or difficulty. We should not say "digging or fishing make him happy." "Happily" is (like "readily," "keenly," "eagerly") an adverb qualifying the way in which a person does something (*e.g.*, thinks or digs or fishes) and not a resultant or by-product of his doing it. As moralists came to recognize these difficulties of definition and the divergence of meaning between the Greek *eudaimonia* and the English "happiness," they tended to drop "eudaemonism" as an ethical technical term on the grounds that it is both confusing and uninformative.

See also ETHICS; ETHICS, HISTORY OF. (J. D. M.)

EUDEMOS OF RHODES (fl. before 300 B.C.), Greek philosopher, was a pupil of Aristotle and a friend of Theophrastus (*q.v.*), with whom he completed his master's philosophy from the point of view of systematization. The fragments of his *Physics* (preserved by Simplicius) and his *Analytics* paraphrase those of Aristotle; and it was he who edited or revised the *Eudemian Ethics* which are preserved under Aristotle's name. His history of geometry, arithmetic and astronomy completed the *Doctrines of the Natural Scientists* of Theophrastus.

See the fragments ed. by F. Wehrli, *Eudemos von Rhodos* (1955), being pt. viii of *Die Schule des Aristoteles: Texte und Kommentar.* (F. WI.)

EUDES, SAINT JOHN (1601–1680), French Roman Catholic missionary, was born at Ri, a country parish near

Argentan in Normandy, on Nov. 14, 1601. After completing his college education under the Jesuits at Caen, he entered the Oratory of Jesus, a society of priests founded by the distinguished Pierre Cardinal de Bérulle. John Eudes was ordained to the priesthood in 1625. Endowed with brilliant talents as an orator, he devoted 50 years of his life, until his death (at Caen, on Aug. 19, 1680), to the preaching of parish missions. In 1641 he established the Order of Our Lady of Charity for the education and rehabilitation of delinquent girls and women. Two years later he left the Oratory and founded the Congregation of Jesus and Mary, now generally known as the Eudist Fathers, for the training of candidates to the priesthood and for the preaching of missions to the faithful. For persons aspiring to Christian perfection in the world he organized the Society of the Admirable Heart of Mary. Greatest among his achievements was the preaching and spreading of the devotion to the Sacred Hearts of Jesus and Mary. He established the first feasts in their honour, for which he wrote both mass and office. On April 25, 1909, in the decree of beatification, Pope Pius X proclaimed John Eudes "The Father, Doctor and Apostle of the Devotion to the Sacred Hearts," and Pope Pius XI further confirmed these titles in his solemn canonization on May 31, 1925. His feast day is Aug. 19.

See the biography by Daniel Sargent, *Their Hearts Be Praised* (1949). (W. Mt.)

EUDES (Odo) (d. 898) king of France, or rather of Francia Occidentalis, the West Frankish kingdom, from 888, was a son of Robert (*q.v.*) the Strong. The holder of extensive lands in northern France and count of Paris, which town he had successfully defended against a Norman siege (885–886), he was chosen king by the West Frankish magnates after the deposition of the emperor Charles III the Fat by the East Franks (Nov. 887). Fulk, archbishop of Reims, objected to this choice, but Eudes was anointed king at Compiègne by Walter, archbishop of Sens, on Feb. 29, 888. He won a brilliant victory over the Normans at Montfaucon (June 24, 888), but had scant authority, especially south of the Loire river, and was obliged to solicit the support and admit the supremacy of the German or East Frankish king Arnulf, Charles the Fat's successor. He had constantly to repel Norman raiders and was hindered by the insubordination of Baldwin II of Flanders. While Eudes was struggling with a revolt in Aquitaine (893), Fulk consecrated Charles III the Simple as king at Reims. Eudes fought Charles with varying degrees of success; he had, however, to cede some lands between the Seine and the Meuse to him in 896, when the Normans were attacking again. He died, thoroughly discredited, at La Fère-sur-Oise on Jan. 1, 898.

See E. Lavisse, *Histoire de France,* vol. ii, part i (1903). (J. De.)

EUDES (Odo) (d. 735), duke of Aquitaine from about 714. The Neustrians sought his help in 718 against Charles Martel (*q.v.*), mayor of the palace in Austrasia, probably offering to recognize him as king of Aquitaine. After two years' warfare, however, he was obliged to come to terms with Charles (720), but retained his independence and repelled unaided a Muslim raid on Toulouse (721). His intrigues with the Muslim governor of Saragossa, a rebel against Abd-al-Rahman, governor of Córdoba, perhaps precipitated the Muslim invasion of 732. Eudes then sought assistance from Charles Martel and, after Charles's victory over the Muslims near Poitiers, swore allegiance to him and was allowed to retain his duchy. Eudes died in 735 and was succeeded by his son Hunald.

In the 16th and 17th centuries forged genealogies of Eudes, suggesting that he came from Aragon, were employed to justify the claims upon French territory made by the Spanish successors of the emperor Charles V; but nothing certain is known of his origins.

See J.-F. Bladé, "Eudes duc d'Aquitaine," *Annales du Midi,* vol. iv (1892). (J. De.)

EUDOCIA (Aelia) (d. A.D. 460), wife of the eastern Roman emperor Theodosius II (*q.v.*), was originally called Athenais and was a native of Athens, where her father Leontius was a pagan philosopher. She was introduced to Theodosius by his sister Pulcheria and changed her name when baptized by the patriarch of Constantinople. Her marriage took place on June 7, 421. In 422 she bore a daughter, Licinia Eudoxia, who married the western emperor Valentinian III (*q.v.*) in 437. In 438 she went on a pilgrimage to Jerusalem and returned to Constantinople in the following year. But in 443 she left the capital a second time to spend the rest of her life in Jerusalem, where she rebuilt the fortifications and built splendid churches. She died in Jerusalem on Oct. 20, 460. For a time she was associated with Monophysitism (*see* Monophysites) but died a Catholic. Her poems included a panegyric on the Roman victory over the Persians in 422 and several religious works. (E. A. T.)

EUDOCIA MACREMBOLITISSA (*c.* 1021–1096), daughter of John Macrembolites and niece of the patriarch Michael Cerularius, was the wife of the Byzantine emperor Constantine X Ducas. After his death (1067), to protect her young sons and to defend the eastern frontier against the Seljuks, she married Romanus IV Diogenes, who became emperor on Jan. 1, 1068. On his capture by the Seljuks in 1071 and subsequent deposition, she soon had to retire in favour of her son Michael VII, and she entered a convent. (J. M. Hy.)

EUDOXIA (Aelia) (d. A.D. 404), wife of the eastern Roman emperor Arcadius (*q.v.*), was the daughter of a Frank named Bauto, consul in 385. It was she who caused the downfall of the eunuch Eutropius (*q.v.*), although he had arranged her marriage with the emperor, which took place on April 27, 395. She was proclaimed *Augusta* on January 9, 400, and for the rest of her life she exercised great influence over her feeble husband. She was an earnest Catholic, but quarreled bitterly with John Chrysostom (*see* Chrysostom, Saint John), patriarch of Constantinople, who attacked her and the frivolity of the court ladies in the most outspoken terms. She procured his deposition in 403, and after a short period of reconciliation had him expelled from his see and exiled in 404. She died from a miscarriage in October of that same year and was buried in the Church of the Apostles in Constantinople. She bore Arcadius four daughters and one son, Theodosius, later emperor. (E. A. T.)

EUDOXIA (Evdokiya Fedorovna Lopukhina) (1669–1731), Russian tsaritsa, the first wife of Peter the Great, was a daughter of the boyar Fedor (Ilarion) Lopukhin. On Feb. 6 (new style; Jan. 27, old style), 1689, she was given in marriage to Peter, a reluctant bridegroom of only 17. The young tsar with his passionate interest in the mechanics of war on sea and land, and Eudoxia, bred in the seclusion of the *terem* (women's quarter), endowed with beauty but not with brains or ambition, had practically nothing in common. Already by 1695 Peter had begun to neglect her, and by 1697 he wanted her out of the way. In 1698 he obliged her to retire to the Pokrovski monastery at Suzdal, where in 1699 she took the veil and the name of Elena but after six months reverted to living as a laywoman. The liaison between Eudoxia and Maj. Stepan Glebov, who was in charge of recruiting in the district, begun in 1709 continued intermittently after Glebov's departure in 1710. The facts of her misconduct came to light at the trial of her son the tsarevitch Alexis (*q.v.*) in 1718. At the same time it transpired that Alexis, who had visited Eudoxia in 1706 and had now and again corresponded with her and sent her money, had not received any help from her in his flight to Austria. After the trial, Eudoxia was transferred to the Uspenski convent on Lake Ladoga and subsequently to the fortress of Schlüsselburg. In 1728, on the accession of her grandson Peter II, she was released and officially rehabilitated. In 1729 she was installed at the Novodevichi convent in Moscow and provided with a generous allowance. Her candidature for the succession to Peter II, who died in 1730, was not widely canvassed and received no support. Eudoxia died in Moscow on Sept. 7 (N.S.; Aug. 27, O.S.), 1731. (L. R. Lr.)

EUDOXUS of Cnidus (*c.* 400–*c.* 347 B.C.), Greek scientist, was equally outstanding as mathematician, as astronomer and as geographer. Though his biography by Diogenes Laërtius (viii, 86*ff.*) is inaccurate and confused, it is reasonably certain that he was a member of Plato's Academy, possibly its acting head when Aristotle joined it in Plato's absence in 367; that either after an earlier stay in Athens in 377 or after having taught at the Academy (*i.e., c.* 364), he spent about 16 months in Egypt; and that he later had a school of his own at Cyzicus.

Eudoxus gave a new definition of proportion (Euclid v, def. 5) which made it possible to develop the theory of proportions, especially in application to incommensurables, far beyond what had been possible before. He developed the method of exhaustion and applied it to many geometrical and stereometrical problems and theorems, which could thus for the first time be rigorously proved. Probably most of Euclid's book xii is essentially the work of Eudoxus. He first solved the problem of the duplication of the cube in a way that later satisfied the standards of Eratosthenes.

Eudoxus made a map of the stars (*enoptron*) that remained authoritative for centuries. To explain the motions of the stars he constructed a model of 27 spheres. The motion of each planet was explained as resulting from the motions of four spheres fixed one within the other so that each interior sphere revolved around its own axis while participating in the motions of the external spheres revolving around different axes. The sun and the moon required three spheres each; the fixed stars only one. He divided the sky into degrees of longitude and latitude, gave a better approximation of the length of the solar year and improved the calendar. He estimated the earth's circumference at 400,000 stadia, divided the earth into zones, drew a new map of the known parts of the earth and wrote a comprehensive geography in seven books.

While a member of the Academy, Eudoxus gave an explanation of the theory of ideas from a physicist's point of view. Through his theory that pleasure is the highest good he exercised a great influence on Aristotle, who discusses this theory in book x of his *Nicomachean Ethics*. Eudoxus may also have had some influence on Aristotle's theory of the prime mover.

BIBLIOGRAPHY.—Kurt von Fritz, "Die Lebenszeit des Eudoxos von Knidos," *Philologus,* 85:478 (1930); O. Becker, "Eudoxosstudien," *Quellen und Studien zur Geschichte der Mathematik, Astronomie und Physik,* pt. B, 2:311–333, 369–387 (1933–34), 3:236–244, 340–410 (1936); H. Karpp, *Untersuchungen zur Philosophie des Eudoxos von Knidos* (1933); F. Gisinger, *Die Erdbeschreibung des Eudoxos von Knidos* (1921). (K. V. F.)

EUDOXUS OF CYZICUS (fl. 130 B.C.), Greek navigator who was employed by Ptolemy Euergetes II to command a fleet exploring the Arabian sea for the sea route to India. After two successful voyages, Eudoxus left the Egyptian service and went to Cádiz, Sp., to fit out an expedition for the circumnavigation of Africa. In the first attempt he was driven aground south of Morocco. He sailed again down the west coast of Africa, but the entire expedition was lost.

BIBLIOGRAPHY.—E. H. Bunbury, *History of Ancient Geography,* vol. ii (1879); P. Gaffarel, *Eudoxe de Cyzique* (1873); M. Cary and E. H. Warmington, *The Ancient Explorers* (1929). (WM. C. B.)

EUGENE OF SAVOY (1663–1736), a prince of the Carignan line of the house of Savoy who in the Austrian service made his name as the greatest soldier of his generation and was also a statesman of remarkable foresight, was born in Paris on Oct. 18, 1663, the younger son of Eugene Maurice of Savoy-Carignan, comte de Soissons, and his wife Olympia Mancini, a niece of Cardinal Mazarin. His father, a cousin of the duke of Savoy, Charles Emmanuel II, had held high positions under the French crown, but after his death (1673) the comtesse de Soissons, who had designs on Louis XIV's affections, became compromised in the scandalous Affair of the Poisons (*see* POISONS, AFFAIR OF THE) and, in 1680, had to flee from France to Brussels, leaving the young Eugene in Paris with his grandmother. Ugly and of poor physique, Eugene was neglected by his relatives and destined for an ecclesiastical career. From an early age, however, he had read the classic treatises on war and studied geometry. Reluctant to enter the church and humiliated at court, he finally applied to serve in the French army, but Louis XIV refused him—partly because of his mother's disgrace. Then, in July 1683, Eugene fled to the Austrian Netherlands, whence he made his way to the Holy Roman emperor Leopold I at Passau.

Eugene had hoped to receive the regiment of his elder brother Louis Jules, who had been killed at the battle of Petronell on July 13, 1683, but Leopold had already bestowed this command elsewhere, so that Eugene had to content himself with serving as a volunteer under Charles V of Lorraine in the relief of Vienna from the Turkish siege. In Dec. 1683, however, Eugene received the Kuefstein dragoon regiment. During the campaigns for the reconquest of Hungary from 1684 to 1688 he distinguished himself by extraordinary bravery. He was promoted major-general in 1685 and lieutenant field marshal in 1688. Early in 1686 he had visited Madrid with his mother, but her intrigues to secure for him a place in Spanish affairs and an advantageous marriage failed, though he became a grandee and a knight of the Golden Fleece.

CULVER PICTURES, INC.

"PRINCE EUGENE OF SAVOY" BY JAKOB VAN SCHUPPEN (1670–1751). IN THE RIJKSMUSEUM, AMSTERDAM

From that time on he resolved to devote himself entirely to the imperial service. He was impressed by the strict military discipline of Charles of Lorraine and also by the vehement offensive spirit shown by his gifted cousins Maximilian II Emmanuel of Bavaria and Louis of Baden.

After the outbreak of the War of the Grand Alliance (*q.v.*) Eugene succeeded in 1690 in winning Victor Amadeus of Savoy to the allied side. After three months' service on the Rhine he was sent to Italy in 1690 as general of cavalry, where, despite the relief of Coni, the capture of Carmagnola and victory over the French general Nicolas Catinat, he was hindered by his superiors Antonio Caraffa and the duke of Savoy from gaining any decisive success. His part in the otherwise disastrous invasion of Dauphiné (capture of Embrun and Gap) gained him appointment as field marshal in 1693 and as imperial commander in chief in Italy in 1694. Now, however, he was hampered by the duke of Savoy, who was to go over to the French in 1696.

The emperor Leopold, though he was skeptical about the young Paris-educated prince, nevertheless in 1697, at the request of Rüdiger von Starhemberg, president of the war council, appointed Eugene *adlatus* to the imperial commander in chief in Hungary, Frederick Augustus I of Saxony, in consideration of his experience and his popularity with the troops; and when Frederick Augustus became king of Poland as Augustus I, Eugene succeeded him as commander in chief. He reorganized the dilapidated and disintegrating army in Hungary and, at Zenta on Sept. 11, 1697, crushed the main Turkish army, three times larger than his own.

After this, by the treaty of Carlowitz in 1699, the Turks had to cede Hungary and Transylvania to the Habsburgs. The battle of Zenta made Eugene the most famous commander in Europe.

The War of the Spanish Succession.—Eugene, together with the archduke Joseph (the future emperor Joseph I), played a great part in persuading Leopold to begin the War of the Spanish Succession (*q.v.*), though the empire was militarily and diplomatically unprepared. In 1701 Eugene with a small élite army crossed the Tirolean Alps into upper Italy, where a French army under Catinat held an extended front between Milan and the Adige river. After his victory at Carpi on the Adige (July 9) he drove Catinat back to the Oglio river. He then defeated Catinat's successor Villeroi in a series of engagements and, in Feb. 1702, took him prisoner in a night attack on Cremona. However, he received inadequate supplies, could not win against Vendôme at the battle of Luzzara on the Po in Aug. 1702 and returned, at the end of the year, to Vienna. There he ruthlessly exposed the empire's weak conduct of the war and its inadequate general policy. When the French, the Bavarians and the Hungarian rebels under F. Rakoczy were threatening Vienna, Eugene was nominated president of the imperial war council and of a commission to purge the administration. He mobilized all resources and, on Marlborough's arrival in southern Germany, joined forces with him. At the battle of Blenheim (*q.v.*) on Aug. 13, 1704, by holding the greatly superior forces under Ferdinand de Marsin and Maximilian Emmanuel and by sending a squadron of cuirassiers that he could ill spare to assist Marlborough, he enabled the latter to prepare the final and victorious offensive. After this great victory, which liberated Germany, the two commanders became close friends, and their subsequent joint successes were the fruit of an amicable partnership unique in military history. (*See* MARLBOROUGH, JOHN CHURCHILL, 1st Duke of.)

The war in Italy was now resumed. The battle of Cassano (Aug. 1705) was inconclusive, though for propaganda reasons Eugene claimed a victory. In 1706, however, he outflanked the French and in September attacked their army besieging Turin. The French greatly outnumbered him, but Eugene dealt them such a blow that the bulk of their forces left Italy. He was then designated lieutenant-general, field marshal of the empire and regent of Milan. His operations against Toulon in 1707 came however to nothing. Sent to the Netherlands in 1708, Eugene hastened ahead of his own army so as to be present with Marlborough in the battle of Oudenaarde (*q.v.*) on July 11. The victory led to the capture of the southern Belgian fortresses, besides Ghent and Bruges. The capture of Lille in Dec. 1708 broke Vauban's defense system, hitherto regarded as impregnable.

At the peace negotiations at The Hague in May 1709 Eugene demanded that the French should not only renounce the Spanish succession but also give up the territories that they had taken from Germany in Alsace since 1648, so that these might form a bulwark to defend the empire against France. He assisted the English demand that the French king should join the allies in driving his grandson Philip from Spain. Louis XIV's refusal led to the bloody battle of Malplaquet (*q.v.*) on Sept. 11 which, though not militarily decisive, strengthened Eugene's intention to march on Paris. When operations were delayed by the new Tory administration in England and by the recall of Marlborough, Eugene visited London in 1712 in a vain attempt to hold the English to the alliance. Despite popular demonstrations in his favour, he could only delay the measures against Marlborough.

The Dutch defeat at Denain and the end of the English subsidies for auxiliary German troops caused Eugene to advise the new emperor Charles VI to accept the peace terms proposed at Utrecht in 1713. After the loss of Landau (July) and Freiburg (November), Charles agreed. In the peace negotiations with the marshal de Villars, at Rastatt, Eugene was able to achieve, on March 6, 1714, remarkably good conditions for the emperor.

The Turkish War of 1715–18.—Formerly the sworn enemy of France, Eugene now advocated a close understanding between the emperor and the French regent, Philippe duc d'Orléans. This left the empire free to assist Venice against the Turks. Eugene drove the Turks back by his victory of Peterwardein in Aug. 1716 and took Belgrade 12 months later, having destroyed the superior Turkish relief army in a risky double-fronted movement. As he opposed the idea of marching on Constantinople, the war was concluded with the peace of Passarowitz (1718), by which Austria gained Belgrade, northern Serbia and Little Walachia. Eugene now brilliantly organized the new territories by bringing in German, Lotharingian and Walloon settlers who developed a wilderness into a fruitful country and a stronghold for the empire against the Turks.

The Patron of the Enlightenment.—In the following years of peace Eugene remained an influential adviser to the emperor Charles VI as he had been to Joseph I, but he lived a withdrawn life with a small circle of liberal-minded friends such as Eleonore Batthyány-Strattmann and the imperial vice-chancellor Friedrich Carl von Schönborn. Employing the great architects J. B. Fischer von Erlach and L. von Hildebrandt and patronizing the sculptor Balthasar Permoser, he decorated both his palace in Vienna (the Belvedere) and his country houses in a baroque style refined by the influence of French classicism. He also patronized French writers such as J. B. Rousseau and was an exponent of French political theories and of the ideas of the Enlightenment in Vienna. Leibniz dedicated his *Monadologie* to him. Eugene moreover, as governor of the Netherlands (1716–24), favoured the spread of Jansenism there through his representative the marquis de Prié (Hercule Turinetti); this influenced the development of the Catholic Enlightenment in Germany.

The Statesman.—Prematurely aged from more than 13 wounds and from physical and emotional strain, Eugene exhausted himself in opposing Charles VI in his dream of re-establishing Charles V's universal monarchy. Eugene advocated, rather, the consolidation of the Habsburg lands as a basis of imperial power in Germany; but he achieved no major reform other than in the war department to this end. In 1719 he foiled the plot against him by G. P. Tedeschi and J. F. von Nimptsch, who had wanted to turn the emperor against him. He could not, however, prevent the emperor's promotion of the Ostend company (*q.v.*) nor the signing of the Austro-Spanish treaty of Vienna in 1725, which, as he correctly saw, cost the emperor the friendship of England. When England in 1725 made an alliance with France and Prussia, Eugene persuaded Charles to abandon his "universal" policy. Alliances with Russia and Prussia in 1726 and 1728 in Eugene's opinion gave Austria an unassailable position in Europe and strengthened the Habsburgs' standing in Germany. Charles VI now gained an almost universal recognition of the pragmatic sanction but did not follow Eugene's advice to ensure its fulfillment by building up the army to 180,000 men. Thus when in 1733, contrary to Eugene's expectations, the war party came to power in Paris and the War of the Polish Succession broke out, Eugene could not hinder French successes in Italy or on the Rhine, though by masterly defense he repulsed a French advance on southern Germany. Now indeed the emperor, following Eugene's advice, came to terms with France. He would not however marry Maria Theresa, his daughter and heiress, to the Bavarian elector, a marriage which Eugene believed would assure the Austrian position in Germany by securing Bavaria.

Eugene's Achievement.—Eugene's military reputation rested on his bravery but still more on his mastery of the art of movement, which dominated contemporary warfare. All its possibilities were exploited: the frontal breach (Blenheim, Malplaquet, Belgrade), the circumvention (Turin), the use of interior lines to remove military strong points as necessary (Oudenarde) and the surprise attack (crossing of the Alps and Cremona). A master of psychology, he was able to inspire his badly equipped troops, partly pressed into service, as well as to calculate his enemies' intentions. Acutely observant and gifted with insight, he could act quickly and daringly at the crucial moment. He was one of the world's great commanders: both Maurice de Saxe and Frederick the Great acknowledged themselves his pupils. By his victories he raised Austria, the leading German state, to the rank of a European power. The empire, which in 1648 had almost succumbed to the attacks of the French, the Swedes and the Turks, was re-established mainly by his victories. Thus the development of a supra-denominational German nation outlined in 1648 was made possible, and

also the development of German statehood within the boundaries of the empire. He was the hero of the resurrected German idea of empire and was celebrated as such both in the song *Prinz Eugen der edle Ritter* ("Prince Eugene the noble knight") and by German poets. His policy, which aimed to secure Bavaria by dynastic alliance and to use Prussia on the lower Rhine against France, was given up after his death but nevertheless proved to be correct; his plans to exchange the Netherlands for Bavaria later became vitally relevant.

Eugene the man, whose personality is revealed neither in conversation nor in his letters and whose private correspondence is lost, represents the ideal of the *honnête homme:* his passions were subject to a will ruled by his sovereign, his honour and his deistically conceived God. Culturally a Frenchman, by fate and career bound to Germany and Austria, full of admiration for Italian art and for the English economy, he was the ideal representative of a European culture even to his contemporaries, who played on his name to call him E*uropae* Genius. He died in Vienna in the night of April 20–21, 1736, and was buried in St. Stephen's cathedral.

Bibliography.—A. von Arneth, *Prinz Eugen von Savoyen*, 3 vol. (1888); *Feldzüge des Prinzen Eugen von Savoyen*, 20 vol. (1876–92); H. Rössler, *Der Soldat des Reichs, Prinz Eugen* (1934); E. Ritter, *Politik und Kriegführung, ihre Beherrschung durch Prinz Eugen 1704* (1934); H. Oehler, *Prinz Eugen im Urteil Europas* (1944), *Prinz Eugen in Volkslied und Flugschrift* (1941); W. S. Churchill, *Marlborough*, vol. ii (1934); M. Braubach, *Geschichte und Abenteuer, Gestalten um den Prinzen Eugen* (1950). (He. G. R.)

EUGENE, a city of western Oregon, U.S., at the head of the Willamette valley, 125 mi. S. of Portland; the seat of Lane county. Named after Eugene Skinner, an early settler (1846), it was incorporated as Eugene City in 1862 and as Eugene in 1864.

It is the major lumber and plywood producing and shipping point of western Oregon and serves as the wholesale distribution centre for all areas of the state south and west. Light industry and agricultural processing are other major businesses. Railroad maintenance shops and a major division point are located there. The public utilities of Eugene are municipally owned and operated.

Eugene has a council-manager system of government, in effect since 1945. It is the seat of the University of Oregon (founded 1872), with its Museum of Art and Museum of Natural History, and the location of Northwest Christian college, a liberal arts college founded in 1895 and supervised by the Disciples of Christ. Eugene is also a medical centre, with two hospitals and several specialized clinics.

The town is noted for its proximity to a variety of recreational attractions, ranging from hunting and fishing to mountaineering.

Eugene is the central city of a standard metropolitan statistical area (SMSA) comprising Lane county. Pop. (1960) city, 50,977; SMSA, 162,890. For comparative population figures *see* table in Oregon: *Population.* (M. Sc.)

EUGENIA, a genus of more than 600 species of generally aromatic trees and shrubs, chiefly tropical, of the myrtle family (Myrtaceae), many of which are of economic importance both for their often edible fruits, which are used for jellies and confections, and for their ornamental value in gardens and landscapes. The tree that produces clove spice is often included in the *Eugenia* (*see* Clove). They have opposite, evergreen leaves, mostly small, solitary flowers in the leaf axils or in few flowered clusters, and the fruit is a berry.

J. HORACE MCFARLAND CO.

ROSE APPLE OR JAMBU (EUGENIA JAMBOS) FLOWER

The rose apple (*E. jambos*), also known as jambu, an ornamental tree up to 30 ft. high, is much planted in southern California; the brush cherry (*E. paniculata*) is also a popular tree in California, while its variety *australis*, the Australian brush cherry, a bushier form, is popular for hedges and topiary pieces. The pitanga, or Surinam cherry (*E. uniflora*), indigenous to Brazil, is a popular garden plant in Florida where it fruits abundantly. (J. M. Bl.; X.)

EUGENICS, an applied biological science concerned with increasing, from one generation to another, the proportion of persons with better than average genetic endowment. It draws from genetics for an understanding of the processes of heredity; from psychology for analysis of the part played by differences in heredity and environment in the development of personality, intelligence, and character; from medicine and medical genetics for information on hereditary defects and susceptibility to disease; and from sociology, and more particularly demography, for the rates of births and deaths, the mating habits, and the social and physical factors that determine the relative increase or decline of persons with different characteristics. In its scientific aspect it is essentially the study of trends and causal factors in human evolution.

HISTORICAL DEVELOPMENT

Early History.—The idea of applying knowledge of heredity to the improvement of the human race goes back to earliest times. References to eugenic ideals appear in the Old Testament, and Plato's *Republic* idealizes a society in which there is constant selection for the improvement of the human stocks. The modern eugenic movement began with Sir Francis Galton (*q.v.*; 1822–1911), a cousin of Charles Darwin. Galton was among the first to see that the theory of evolution implied the possibility that man, in part, might direct his own evolutionary future. His first important work, *Hereditary Genius* (1869), contained the results of his studies of the families of eminent men as evidence for his belief that "it would be quite practical to produce a highly gifted race of men by judicious marriages during several consecutive generations." He also recognized the effect of social influences on the distribution of births. Much of Galton's long life was spent gathering further evidence to test these beliefs. In 1883 he published *Inquiries into Human Faculty*, in which he coined the word eugenics. In *Natural Inheritance* (1889), Galton pioneered the development and application of advanced statistical methods to the study of man.

While Galton's thinking was much ahead of his time, he retained some of the prejudices of an English gentleman in regard to social class and race. In his studies of the families of eminent men Galton underestimated the effects of a favourable environment. But his faults in this respect were far less than those of many of his followers. As a scientist, Galton was pioneering in a new field. He developed and used the normal curve of distribution and the coefficient of correlation and was the first to recognize the value of twin studies for research in heredity. He repudiated the Lamarkian view, then widely held, that acquired characteristics were inherited. He also developed the law of filial regression.

While Galton attached religious significance to the eugenic movement, he did not think of it in revolutionary terms, but as ". . .the new duty which is supposed to be exercised concurrently with and not in opposition to, the old ones upon which the social fabric depends."

For all his wisdom and high ability, Galton was not able to gain any wide acceptance for eugenics, largely because much of the scientific and technical foundation was lacking. Psychology was not sufficiently advanced to help him in his studies of individual differences or in assessing the importance of the environment; there was no cultural anthropology to provide an outline of the growth of civilizations. Perhaps most important of all, birth control was so little recognized or accepted that Galton did not take it into account as a possible factor in controlling births.

From Galton to World War II.—Galton had endowed a research fellowship in eugenics in 1904 and, in his will, provided funds for a chair of eugenics at University College, London University. The fellowship and later the chair at University College were both occupied by Karl Pearson (*q.v.*), a brilliant mathematician who helped to create the science of biometry (*q.v.*). Pearson was a controversial figure who believed that environment had little to do with the development of mental or emotional qualities. He

felt that the high birth rate of the poor was a threat to civilization and that the "higher" races must supplant the "lower." His views gave countenance to those who believed in racial and class superiorities. Thus, Pearson shares the blame for the discredit later brought on eugenics in the United States and for making possible the dreadful misuse of the word eugenics in Hitler's propaganda. The English Eugenics Society, founded by Galton in 1907 as the Eugenics Education Society, opposed Pearson's views, but was itself slow in throwing off the incubus of the pseudoscience of that time.

In the United States the American Eugenics Society had been founded in 1905 by Madison Grant, Henry H. Laughlin, Irving Fisher, Fairfield Osborn, and Henry Crampton. Their views were coloured by the times in which they lived. They believed that the white race was superior to other races, and that the "Nordic" white was superior to other whites. They lacked entirely the Mendelian concept of race; they thought of races as "pure" groups, quite separate from each other. They did not know, as is clearly known today, that all races are mixtures of many types, the distribution of genes among the races varying in proportions rather than in kind, as evidenced by the distribution of the various blood groups among all the races. They believed that the upper classes had superior hereditary qualities that justified their being the ruling class. The science of that day supported extreme views on feeblemindedness and "criminal types." Intelligence tests, introduced in the early 1900s by Alfred Binet, were thought, contrary to Binet's views, to be measures of innate, genetic intelligence. People whose IQ gave them a mental age of 12 or under could be classified as feebleminded or morons, even though the term would then include most of those brought up in deprived environments, such as the southern Appalachian highlands or the areas occupied by the descendants of slaves. Criminality was generally considered a concomitant of feeblemindedness. Studies by R. L. Dugdale, Lothrop Stoddard, and others on degenerate family stocks—the Jukes and Kallikaks of the eugenic literature of that day—were taken to prove that hundreds of persons in each of these families were feebleminded or criminal types because of the inheritance they received from a single ancestor five or six generations back. Claims were made that immigrants from southern and eastern Europe, besides being socially inferior, included a remarkable number of criminal and defective stocks. There was much in the intellectual atmosphere of the U.S. that fostered an extreme hereditarian racist view.

Even before 1905 eugenists had been active in the effort to restrict immigration. Their arguments, along with those of others who advocated restrictions for different reasons, culminated in 1924 with the passage of the National Origins Quota Law limiting quota immigrants to about 150,000 annually, with no more than 2% of each nationality according to the number of persons of their national origin in the U.S. as of 1890, and providing that beginning July 1, 1929, the quota of any country shall have the same ratio to 150,000 as the number of persons of that national origin living in the U.S. had to the total population of the U.S., as determined from the 1920 census of population. In later years it became clear that the material the eugenists had presented to congressional hearings had little scientific foundation.

Eugenists at this time laid great stress on the importance of sterilizing defective persons. By 1931 sterilization laws had been enacted by 27 states in the U.S., and by 1935 sterilization laws had been passed in Denmark, Switzerland, Germany, Norway, and Sweden. Most of these laws provided for the voluntary or compulsory sterilization of certain classes of people thought to be insane, idiotic, imbecilic, feebleminded, and epileptic; some applied equally to habitual criminals, moral perverts, or the feebleminded. In most cases the purpose was clearly eugenic, though some laws permitted sterilization for social rather than genetic reasons. In most states the laws were not put to much use, and the number of sterilizations under the law was seldom more than 100 per year. Exceptions were California, where sterilizations averaged over 350 cases per year and a total of 9,931 by 1935, and some of the southern states, with a fairly high rate relative to the size of their populations.

With the advance of science the background for these laws began to be questioned. In 1935 a committee of the American Neurological Association reported on an extensive investigation and evaluation of the "facts and theories which constitute the subject matter of the inheritance of the mental diseases, feeblemindedness, epilepsy and crime." The committee concluded that there was serious question whether "many of the eugenic proposals now current take into account the newer genetic data" and that "though the avowed purpose of such sterilization invariably starts out by being a eugenic one, it often ends by becoming a social one." They exposed many fallacies in the reasoning behind the sterilization laws and concluded that there was no sound basis for sterilization because of immorality or character defect. They recommended sterilization only with the consent of the patient or those responsible for him, and only in the case of certain specified diseases or defects, namely: (1) Huntington's chorea, hereditary optic atrophy, familial cases of Friedreich's ataxia, and a few other rare defects known to be hereditary; (2) feeblemindedness of familial type; (3) schizophrenia (dementia praecox); (4) with serious reservations, manic depressive psychosis; and (5) epilepsy, mainly on the ground of its social aspects. Their recommendation for concerted, coordinated, and planned long-term research was not seriously followed until after 1945, when medical schools began serious work in medical genetics.

By 1935 American scientists had in general repudiated much of the work of such men as H. H. Laughlin and C. B. Davenport at the Cold Spring Harbor (N.Y.) Eugenics Record Office, and the leadership of the American Eugenics Society had passed from persons with strong racist leanings to those who sought a eugenics program that would be both scientifically valid and socially acceptable.

In Europe the studies of criminal types were being discredited, and serious work in eugenics continued in the form of family studies by government institutes. Only in Germany did eugenics continue to serve racial and political purposes. There Hitler, exploiting the concepts of the master race and the superman as expounded by some German scholars, developed eugenics-based racism into an extraordinary political weapon.

During the late 1930s, with eugenics at its lowest ebb, demographic changes of great eugenic significance were taking place almost unnoticed. The inverse relation between fertility and education or income reached and passed its peak. Birth-control practices had begun to extend to the lower-income groups, and there were signs that differentials were leveling off at the top. If this change were taking place between economic groups, something of the same sort might take place between individuals of different types of ability within each group. These trends seemed to suggest possibilities for new and practical eugenic applications.

Eugenics After 1945.—With the development of atomic weapons, the general public and medical men alike became educated to the dangers of genetic mutations produced by radioactive fallout from nuclear explosions. X rays used in medicine came under new scrutiny, and the findings of studies on hereditary factors in diabetes and other diseases received new attention. Medicine and public health, professions previously little interested in heredity, recognized a new responsibility. By the 1960s the major medical schools in the United States had included geneticists on their research staffs, and a number had set up departments of medical genetics. Also, heredity clinics had been made available to advise couples on the chances of their having defective offspring. In the same period the experience of adoption agencies led to a more careful selection of foster homes appropriate to the genetic and familial background of the child being adopted.

Still other factors arose to increase public interest in heredity. Of these the "baby boom" was perhaps the most important. Births, which had averaged around 2,000,000 per year in the U.S. before the war, rose to over 4,000,000 by the mid-1950s and remained at a high level through the following decade. A similar increase in births took place in Europe, but it was of shorter duration.

During the post-war period there were large increases in crime, juvenile delinquency, and illegitimacy. Questions arose as to

the children of the next generation: what would be their home environment, what their heredity? Birth control, given worldwide publicity as a solution to the population problem, raised similar questions in the public mind; would its use be restricted to the people who could afford it and who presumably would provide a good environment for their children and possibly a good heredity, or should public agencies provide free birth-control facilities for people at the lowest income levels, who might not be able to provide their children a good environment, and possibly not a good heredity? Also during the post-war period considerable publicity was given to the practice of artificially inseminating women whose husbands were not able to give them children, leading to questions of the heredity of a child born by such insemination. Meantime, the scientific basis for eugenics was being strengthened by advances in psychology, in demography, and especially in human genetics. As a result of all these factors, public opinion, having swung from a considerable hereditarianism in the first quarter of the 20th century to an extreme environmentalism in the second, was beginning to take a more balanced view of the contributions made by both heredity and environment to qualities of character and intelligence.

The atmosphere by the 1960s was appropriate to a reconsideration of eugenic problems. L'Institut National d'Études Démographiques in France had for some time sponsored important work in the field of population. In England the Royal Commission on Population made its report in 1949, and the Population Investigation Committee at the London School of Economics was giving continuing attention to quantitative and qualitative aspects of population trends in its publication, *Population Studies*. In the United States a number of the larger universities had set up facilities for training and research in demography, and there had been several nationwide surveys on factors affecting fertility. Genetics, too, made advances, moving on from the gene to the genetic code in the nucleic acid of the cell nucleus and opening up a whole new field of the biochemistry of gene structure and gene behaviour. Advanced studies on the genetics of defect were carried on in government institutes in Europe and in the National Institutes of Health in the U.S.

BASE LINES OF MODERN EUGENICS

Physical and Mental Defects.—The first genetic differences between individuals to be studied by scientists were departures from the normal in the form of defective development, susceptibility to disease, and mental defects arising from constitutional causes. Scientists have found that genes that cause defects are not limited to afflicted families, but are spread widely through the population. The genetic consequences of artificial radiation, of lifesaving advances in medicine, and of new migration and marriage patterns are reflected in an accumulation of deleterious genes over and above the subtle equilibrium previously established by natural forces.

It is generally accepted that genes causing defects originate in mutations that occasionally take place in one or another of the many thousands of genes present in pairs in every human cell. Usually the mutated gene is recessive, that is, it does not have a harmful effect unless the other gene in the pair has the same characteristics. The chances of two such genes meeting as a pair depend on the number of such genes distributed throughout the intermarrying population. Thus under random mating, if a particular deleterious gene is carried by one person in a hundred, the chances of the mating of persons with such genes are 1 in 10,000. (In a large population, such as that of the United States, such a ratio results in a considerable number of defective persons.)

With an increase in the proportion of carriers of a particular defective gene, there is an even greater increase in the likelihood of a mating in which two such genes will be paired in the same fertilized cell. When this happens, there will be a defect. If the defect is lethal, the two deleterious genes will be taken out of circulation. If the defect is minor, but of a sort to make marriage or reproduction less likely, the genes are to that extent less likely to survive. Thus at some point nature establishes a balance in which, for every new deleterious gene brought into circulation, a similar gene is lost from circulation. Geneticists believe that most people carry at least a few deleterious genes; some estimate the average to be as high as 8%, and at least 2% of the babies born will carry all their lives some major or minor genetic defect.

Most scientists believe that the proportion of carriers of deleterious genes may now be increasing, though the results in the form of an increase in defect would not be apparent for a number of generations. Exposure to X rays and to fallout from atomic explosions is accelerating the mutation rate, while humanitarian medicine is carrying an increasing number of persons with some kind of defect through a period of successful childbearing. Diabetics, for instance, formerly died at an early age, and the genes for diabetes were lost with them. After an insulin treatment was found, these people led longer useful lives, but were not generally able to have children. Now, safe delivery of the children of diabetic mothers is commonplace. The deleterious genes remain in circulation, while new ones are constantly added by mutations. With good medical care, diabetics live normally and bear diabetic children of their own, but the accumulation of certain types of deleterious genes could conceivably reach a point that might seriously threaten man's future.

With strong control measures in effect, defects due to dominant genes could be reduced to a minimum in a short time. But most defects are due to recessive genes, and these are hard to locate because carriers often do not show the defect. Methods being devised for identifying carriers would facilitate control, but defective genes will never be completely eliminated. The best that can be hoped for is that they might be reduced to those that are constantly arising from new mutations. The possibility of "correcting" nature's mistakes at their source, the deoxyribonucleic acid (DNA) molecule, is a hope for the future. It is a guarded hope, however, for the control of the gene—the engineering of human development—is a terrible power, equivalent to the harnessing of the atom. How man uses it should stir concern among all thinking persons.

While present social and medical conditions favour the increase of many genes that cause physical defects and susceptibility to certain diseases, other genes that affect mental disease and mental defect are probably diminishing as a result of the low marriage rate and the institutionalization of the carriers. It is probable that efforts at control will be strengthened by increased use of sterilization in some areas, by increased use of more effective means of birth control, and by a more eugenic use of heredity counseling clinics. Each year an increasing number of defects are being defined, the type of transmission understood, and in many cases means found for spotting carriers.

Psychological Traits.—With responsibility for work on mental and physical defects transferred to the professions of medicine and public health, eugenists are now giving major attention to the genetic factors that provide the base for the development of such multifactorial traits as intelligence, character, and personality. The genetic base for these psychological traits is provided by multiple genes in various combinations. They may be broken up and recombined with other genes from one generation to another, making them difficult to trace in family lines. It will take time to locate these genes and trace the processes involved in their transmission. Fortunately, the efficacy of selection does not depend on a full knowledge of the genetic mechanism. For the present and perhaps for generations to come, eugenists will be guided by familial incidence of intelligence and personality and by the survival of family stocks of different types.

Massive evidence from studies on identical and fraternal twins, nontwin siblings, and adopted children reared away from their natural parents indicates that differences in genetic factors play an important part in psychological differences between individuals. These were Galton's findings, and they have been confirmed by larger and more carefully controlled studies. Identical twins are extremely similar; even when reared apart from infancy, they resemble each other in physical as well as mental characteristics to a degree far greater than that shown by fraternal twins reared together.

Children tend to be like their parents in intelligence because of

their common heredity, as well as because they usually share a common environment. The evidence on differences in personality is not so clear, largely because these traits are more difficult to measure, but all studies indicate that personality, like intelligence, is the product of the interaction between the environment and the genetic base. This is clearly shown in the extremes of personality, like schizophrenia, which are quite susceptible to measurement.

Environmental and Socioeconomic Factors.—As work in the field of individual differences has progressed, scientists have come to feel that the old dichotomy between heredity and environment is no longer useful; both are a part of the processes that begin when the cell is fertilized and starts its separate life; both operate in all the processes of growth and change and decay leading to the death of the individual; and both contribute to making each individual different from every other individual. Genetic factors can be given value only in relation to specific environments, since some qualities that are socially valuable in one environment may be harmful in another.

There is no scientific basis for ranking any one race genetically superior to another. The geneticist finds that no races are "pure," as shown, for instance, by the finding that blood groups do not in general differ in kind among the races, but only in the proportion of each type. Racial groups differ in their average response to intelligence tests, but these tests invariably involve environmental factors that favour those culturally more fortunate. Even when differences in such factors as place of residence, income, or length of schooling can be controlled, the results may still be influenced by more factors affecting motivation and behaviour. The anthropologist finds no evidence that physical differences between the major races denote a higher stage of development in one race than in another.

In studies of socioeconomic, or other, groups, there is no evidence of genetic differences, except perhaps in the case of comparisons between certain of the professional classes and the rest of the population, in which there is some indication that the genetic factors for response to intelligence tests may be slightly superior. Differences between socioeconomic classes in intelligence-test response are obviously strongly influenced by differences in environment. It is as yet impossible to measure precisely the relative genetic potentialities of any large population groups. Eugenists today are concerned with increasing the proportion of able individual and family stocks regardless of the race or social grouping to which they belong.

Until very recently, from 30–50% of all children died before maturity, and death was an important factor in the survival of particular stocks. The evidence from some areas where people are still living under primitive conditions seems to favour the belief that under conditions of life as it was everywhere prior to the mid-18th century, the general tendency was toward the survival of the more competent stocks. If so, this situation changed in the industrialized countries, when the death rate began an accelerated decline, dropping to 12 or less per thousand by 1940. In the mid-20th century, about 95% of all children lived to age 30, and differential deaths had little effect on differential survival.

While this change was taking place, social class differences in births were widening with the introduction of birth-control measures, which reached first into the more educated and better-off classes. With survival from death no longer favouring the more educated and differential births greatly favouring the less educated, who made little effective use of birth control, the inverse relation between births and education or socioeconomic status widened rapidly.

In the United States this inverse relationship reached its peak with the postponement of births and marriages in the last years of the depression. For the period 1935–40 women with one to three years of college education had a net reproduction rate 33% below that required for replacement of their numbers, and women with four years of high school had a rate 26% below replacement, while women at the elementary school level had a rate twice that of the college women and well above that required for replacement. The upsurge of births after World War II changed this trend. By 1960 among women 35 to 40 years old, who had practically completed their childbearing, high school women and women with one to three years of college had borne more children than the number needed for replacement. Women with less than eight years of elementary school education had had about the same number of children as those who had completed their childbearing during the depression.

The trend toward a reversal of differentials seems likely to continue. Extensive studies made about 1960 on national samples of the U.S. population indicate that the number of children actually desired by couples in the "blue collar" group is less than the number desired by couples in the "white collar" group. When both the husband and the wife were college graduates, only 7% of the last pregnancies were unwanted, while among the couples at grade school level 33% had not wanted their last pregnancy. Of the college couples 92% effectively used contraception, while only 69% of the grade school couples used contraception, and used it less effectively.

It is expected that contraception may soon be used as widely and as effectively at the lowest educational levels as at the highest. In such case, present trends suggest that educational and social class differentials in births may soon bear a direct relation to size of family. The public generally, and many scientists, have assumed that an inverse relation between size of family and education or socioeconomic class must represent an unfavourable, or dysgenic, trend, even though there is little evidence for genetic differences between classes. If the relation between size of family and educational level becomes direct, with the more educated groups having the greater proportion of children, the trend would be generally taken as favourable, or eugenic. But its greatest significance would be an increase in the proportion of children brought up in above-average home surroundings. The major opportunity for genetic improvement will not be found in selection by groups, but in selection of individuals within each group.

There is some limited evidence that at the present time in the United States the abler and the more intelligent members of any group tend to have the most children. In studies made as early as 1927 on graduates of Yale and Harvard, and again on Princeton graduates in 1939, the men were rated by a committee of classmates for degree of success some years after graduation. In each group the highest-rated married men averaged over two children apiece, while the lowest-rated married men averaged less than one and one-half children. J. V. Higgins and S. C. and Elizabeth Reed reported in 1962 on the intelligence and family size of a group of 4,000 persons in Michigan with known IQ values. They found that, while there was a slight negative correlation between test intelligence and size of family, this was offset by the larger proportion married at the higher intelligence levels, leaving the IQ potential of the population near equilibrium. A 1963 Michigan study of native whites born in 1916–17 reported a positive, though small, correlation between IQ and fertility. Of the individuals in the IQ range 69–79, 30% left no offspring. Carl J. Bajema in a study made in Ann Arbor, Mich., in 1964, relating individual education with fertility, showed, for women over 50 years of age, a significant negative correlation between years of schooling and fertility; but among the younger groups of women there was no relation between fertility and years of schooling, indicating a trend similar to that which took place between educational groups in the United States in the same period.

There is thus some evidence that in attempting to increase the comparative fertility of the more intelligent and successful individuals, eugenists are working with and not against present trends.

A number of changes taking place in industrial societies today are favourable to individual selection. There are more opportunities for finding work appropriate to one's genetic capacities. There is a wider choice of mates in our highly mobile society, which favours the production of novel types and offers greater possibilities if there is selection. The trend toward more education for all classes would help make birth control selective within each class of society. Furthermore, the continuing improvement of contraceptive methods can be expected to improve voluntary forms of selection at even the lowest levels of education. To further these favourable trends, the eugenist would urge better training for

choice of mate and for marriage, a more general individual and community sense of responsibility for the kind of parents who have more than one or two children, and a continuing study of all the psychological, social, and economic pressures and rewards that would tend toward a voluntary and favourable selection of births within every occupational group.

THE RANGE OF EUGENICS PROPOSALS

There is no general agreement among eugenists as to criteria for selection. The English Eugenics Society has conducted studies designed to identify superior families in every class of society who might be given special incentives to have large families. In the U.S. the Nobel Prize winner H. J. Muller proposed that selection be from men and women of the highest intellectual and social attainments, using artificial insemination to swell the proportion of their children, a proposal that would obviously meet with considerable social, legal, and religious objections. Most American eugenists, however, believe that selection can properly be based on the comparative success or achievement of each individual within the particular group in which he is found. On such a basis of selection, the eugenic policy that would evolve is best described in a report of the American Eugenics Society made in 1961:

> One course open to eugenics is to find, and to further, those institutional arrangements, those social, economic and psychological pressures, which would result in a tendency for individuals to have more or fewer children in proportion to their success or achievement in their particular environment. Eugenic policies of this sort would reach into each social, economic and occupational group, and improve the quality of each. The result would be, not a segregation into superior and inferior classes, but a genetic diversification in the process of adapting to the diverse elements of our society. One could then expect a greater variety of talents, and hopefully, a constantly higher level in those qualities which have a common tendency toward success in all environments . . . . It can probably be assumed that couples with a better than average genetic inheritance will, more often than not, provide a better than average home environment, by whatever criteria quality is to be judged. If this is indeed the case, eugenic policies would work hand in hand with an improved environment to raise the level of human qualities. There is no conflict between those desiring to raise the level of our genetic inheritance and those desiring to raise the level of our social inheritance.

Modern eugenics seems little interested in authoritarian controls. Rather it is hoping to shape the social and economic environment in such a way as to influence a eugenic distribution of births throughout entire populations in a voluntary and largely unconscious process of selection.

See GENETICS, HUMAN; HEREDITY; INTELLIGENCE.

BIBLIOGRAPHY.—M. H. Haller, *Eugenics* (1963); C. P. Blacker, *Eugenics: Galton and After* (1952); F. H. Osborn, *Preface to Eugenics*, 2nd ed. (1951); J. Sutter, *L'Eugénique* (1950); American Neurological Association, *Eugenical Sterilization: a Reorientation of the Problem*, by Abraham Myerson *et al.* (1937); A. Anastasi, *Differential Psychology*, 3rd ed. (1958); C. F. Westoff *et al.*, *Family Growth in Metropolitan America* (1961); C. F. Westoff, R. G. Potter, Jr., and P. C. Sagi, *The Third Child* (1963); R. Freedman, P. K. Whelpton, and A. A. Campbell, *Family Planning, Sterility and Population Growth* (1959); T. G. Dobzhansky, *Mankind Evolving* (1962); F. H. Osborn, "Overpopulation and Genetic Selection," in F. Osborn (ed.), *Our Crowded Planet* (1962); G. Wolstenholme (ed.), *Man and His Future* (1963). (F. O.)

EUGÉNIE (1826–1920), empress of the French as consort, from 1853, of Napoleon III, was born at Granada on May 5, 1826, the younger daughter of the Spanish grandee Don Cipriano Guzmán y Porto Carrero, conde de Teba and subsequently conde de Montijo, and his wife Maria Manuela Kirkpatrick, a daughter of the U.S. consul at Málaga. Before her marriage she used the surname de Montijo and the title condesa de Teba. Her father had fought on the French side in the Peninsular War, and as a child she had listened to Stendhal's stories of Napoleon. She grew up with a romantic interest in the Napoleonic legend and saw Louis Napoleon for the first time in Paris after his ill-fated Strasbourg attempt (1836). When he became president of the second republic she appeared frequently with her mother at the balls that he gave at the Élysée. He was strongly attracted to her auburn-haired blue-eyed beauty but she made it clear to him that he must think of her as a possible consort rather than as a mistress. His projects for a royal marriage had fallen through, and five weeks after the proclamation of the second empire (Dec. 1852) he proposed marriage to Eugénie. On Jan. 22 he formally announced his engagement and justified what was considered a *mésalliance*. The marriage was celebrated with great pomp at Notre Dame on Jan. 29, 1853. On March 16, 1856, the empress gave birth to a son, who received the title of prince imperial. By her beauty, elegance and charm of manner she contributed to the brilliance of the imperial regime, and when the end came, she was, as the official *enquête* made by her enemies proved, one of the few who showed calmness and courage in face of the rising tide of revolution.

The empress acted three times as regent during the absence of the emperor—in 1859, 1865 and 1870—and she was generally consulted on important questions. She was brave and intelligent, but willful and impatient. Her Catholic sympathies influenced French policy in the affair of the Holy Places (1853), the Roman question (1858–59) and the Mexican enterprise (1861). She also contended that the development of the liberal empire, forced on Napoleon by the failure of his foreign policy, should be postponed until the prince imperial should come of age. In July 1870 she strongly supported the decision, which precipitated the Franco-German War, to demand assurances from Prussia that the Hohenzollern candidature to the throne of Spain should never be renewed. On the collapse of the empire (Sept. 1870) she fled to England, and settled with the emperor and her son at Chislehurst. Widowed in 1873, she found a friend in Queen Victoria, whose own bereavements deepened her sympathy with Eugénie. In 1879 her son was killed in the Zulu War, and in 1880 she made a pilgrimage to the spot where he had fallen. Thereafter she lived either at Farnborough (where she built a church as a mausoleum for Napoleon III and her son) or at Cap Martin on the Riviera but continued to follow closely the course of events, especially the conclusion of the Versailles treaty of 1919. She died in Madrid on July 11, 1920, on a visit to the queen of Spain, her favourite goddaughter. She was herself buried at Farnborough.

See R. Sencourt, *La Vie de l'impératrice Eugénie*, 4th ed. (1933), being a French translation with certain factual corrections of *The Life of the Empress Eugénie* (1931). (F. M. H. M.)

EUGENIUS, the name of four popes.

SAINT EUGENIUS I (d. 657), pope from 654 to 657, was a Roman. He was elected on Aug. 10, 654, during the lifetime of the exiled Martin I, who acknowledged him to be the legitimate pope. Emperor Constans II urged Eugenius to recognize the Monothelite patriarch of Constantinople, Peter, but he declined to do so. He died June 2, 657, and his feast is observed on the anniversary of his death.

EUGENIUS II (d. 827), pope from 824 to 827, was a Roman. A disputed election followed the death of Paschal I, and Eugenius, candidate of the nobles, carried the day. The *Constitutum* of the co-emperor Lothair I (824) vested the papal election in the Roman clergy and nobles, subject to imperial confirmation. Eugenius died Aug. 27, 827. (A. G. Bi.)

EUGENIUS III (Bernardo Paganelli) (d. 1153), pope from 1145 to 1153, the first Cistercian pope, an ardent reformer and a disciple of St. Bernard of Clairvaux, was elected on Feb. 15, 1145. Eugenius, like the entire west, was shocked by the news of the fall of Edessa (Dec. 1144) a few months after he succeeded Lucius II. The pope, who was on cordial terms with Conrad, first Hohenstaufen king of Germany, and Louis VII of France, urged the French king to lead a crusade, naming St. Bernard to preach it and working tirelessly to promote it himself. The second crusade, most impressive of all in scope, ended in failure, largely because of Christian rivalries (*see* CRUSADES). Out of Rome under its hostile new senate during much of his reign, Eugenius held many councils. He concluded the treaty of Constance (1153) with Frederick I, fixing conditions for his imperial coronation, but died, on July 8, 1153, before the king could come to Italy. His cult was approved Oct. 3, 1872, and his feast day is July 8. (J. J. RN.)

EUGENIUS IV (Gabriele Condulmaro) (1383–1447), pope from 1431 to 1447, was unanimously elected on March 3, 1431, to succeed Martin V. His pontificate was dominated by his struggle with the Council of Basel, which opened in 1431. When Eugenius sought to dissolve it, because of its hostility to his authority, its

members affirmed the superiority of the council over the pope. Eugenius decreed the closure of the council and called a new council to meet at Ferrara in 1438. A plague forced its transfer to Florence, where an ephemeral union with the Greek Church was concluded in 1439. His pontificate dealt the death blow to the conciliar movement and restored papal sovereignty to the church. Eugenius died on Feb. 23, 1447.

BIBLIOGRAPHY.—Philip Hughes, *A History of the Church* (1947); Ludwig Pastor, *The History of the Popes From the Close of the Middle Ages* (1891); A. Fliche and V. Martin (eds.), *Histoire de l'Église depuis des origines jusqu' à nous jours* (1935 *et seq.*).

(J. A. Cr.)

EUGENIUS (EUGENIUS II) (*c.* 1130–*c.* 1202), royal admiral (*'αμηρᾶς* or *amiratus*) of Sicily, was not only a brilliant scholar at the learned court of the Norman kings of Sicily but also a great civil servant who played his part in politics. Sprung from a noble Greco-Sicilian family, several being *amirati,* he was knowledgeable in Greek, Latin and Arabic and became famous in mathematical physics through his Latin translations, from Arabic and Greek respectively, of Ptolemy's *Optics* and *Almagest.* He produced also a Latin translation and expansion of the *Prophecy of the Erythraean Sibyl* and revised an existing Greek translation of the *Fable of Kalila and Dimna.* Further, he wrote distinguished Greek poems, containing much biographical matter and revealing his own character as well as his familiarity with classical Greek literature, Platonic philosophy and observation of natural phenomena. From 1174 to 1189, as *magister regie duane baronum,* he controlled the financial administration in the south Italian provinces of the Norman kingdom. After the death of William II (1189), he supported the illegitimate Tancred of Lecce against the legitimate heiress Constance and her husband, the emperor Henry VI. He was rewarded with the office of *regius amiratus,* which gave him the supreme direction of finance (command of the navy being vested in the *amiratus regii stolii*). Involved in the final ruin of Tancred's family, Eugenius suffered a brief imprisonment in Germany, but he soon returned to fiscal duties in southern Italy under successive rulers till 1202.

Eugenius' translations helped the advance of physical science and also fostered the apocalyptic enthusiasm of the later Joachites. The emperor Frederick II owed much to his political philosophy, psychology and fresh scientific method. Moreover, if Eugenius indeed wrote, as has been suggested, the famous Latin *Historia* and the *Epistola ad Petrum* ascribed to "Hugo Falcandus," it is through his eyes that Norman Sicily has been largely evaluated. Both works make extensive use of *Kalila and Dimna* and show many similarities with Eugenius' poems in political thought, scientific outlook and analysis of character. An acquaintance with administrative technique is evident, as well as linguistic genius.

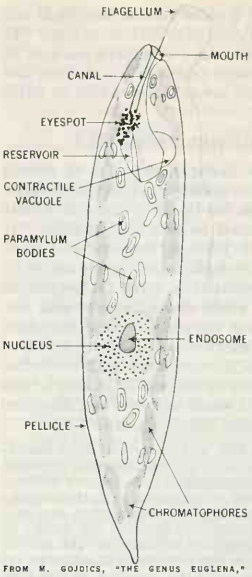
BIBLIOGRAPHY.—E. Jamison, *Admiral Eugenius of Sicily,* . . . (1957). *See* also C. H. Haskins, *Studies in the History of Medieval Science,* 2nd ed. (1929); M. Amari, *Storia dei Musulmani di Sicilia,* vol. iii, 2nd ed. (1939); L. R. Ménager, *Amiratus—'Αμηρᾶς: l'émirat et les origines de l'amirauté* (1960).

(E. M. Ja.)

EUGLENA, a genus of one-celled organisms usually placed in the order Euglenida of the animal phylum Protozoa (*q.v.*), but, because of certain plantlike characteristics, consigned by some biologists to the division Euglenophyta of the algae (*q.v.*). Species of *Euglena* are flagellates characterized by an elongated body containing a single nucleus and one to many grass-green (chlorophyll-containing) pigment bodies (chromatophores). Species measure 15–500 microns (.015–.5 mm.). An anterior flask-shaped reservoir opens through a canal ("gullet"). A whiplike swimming flagellum extends from a granule (blepharoplast) near the base of the reservoir but, in exceptional cases (*e.g., E. mutabilis*), does not project beyond the canal.

Food is absorbed directly through the cell surface. Nutritional requirements are known for several species in pure culture in chemically defined media. Although growth is stimulated by simple organic foods, certain species (like plants) grow in light in inorganic media supplemented with vitamins B_1 (thiamine) and B_{12}. In darkness, an organic energy source (*e.g.,* acetate) is needed. *E. gracilis* is a good assay organism for determining the presence of forms of B_{12} in biological materials and natural waters.

Different euglenas live in fresh and brackish waters and in mud, especially with moderate organic pollution. Some can develop tremendous populations as green or red "blooms" in ponds or lakes. Several species produce resting cysts that can withstand drying for at least several weeks. Reproduction is asexual, by binary fission; sexual reproduction is unknown.



FROM M. GOJDICS, "THE GENUS EUGLENA," 1953; REPRODUCED WITH PERMISSION OF THE REGENTS OF THE UNIVERSITY OF WISCONSIN, UNIVERSITY OF WISCONSIN PRESS

CELL STRUCTURE OF EUGLENA GRACILIS

The flagellar core (axoneme), in electron micrographs, shows the usual filaments (*see* FLAGELLUM). The flagellar sheath shows delicate lateral filaments (mastigonemes). A few species have a short second flagellum within the reservoir; more often the second flagellum adheres distally to the swimming flagellum, simulating a forking. A paraflagellar body on the flagellum may be the light-sensitive structure responsible for phototaxis. The eyespot (stigma) on the reservoir, opposite the paraflagellar body, is a discoid aggregate of reddish granules, possibly a light filter for the paraflagellar body. The contractile vacuole, during contraction, fuses with the reservoir on the side opposite the flagellar attachments.

The cell pellicle, typically spirally striated, may be thin or thick. Species with a thin pellicle can undergo marked changes in shape (metaboly); those (*e.g., E. acus*) with a thicker pellicle bend but do not show metaboly.

Paramylum bodies (containing a glucose polymer) may be large and few or small and numerous. These inclusions may be rodlike, platelike or cup-shaped. Chromatophores are often associated with colourless pyrenoid bodies (sheathed with paramylum or naked). In electron micrographs chromatophores contain alternate dense and less dense layers, the former presumably containing lipids (fatty substances) or lipoproteins. Loss of chromatophores can be induced in *E. gracilis* following exposure to streptomycin or Pyribenzamine. Such chromatophore-lacking strains grow well in suitable media. Temporary bleaching may also occur in darkness; restored to light, the euglenas become green again. Certain species (*e.g., E. rubida*) contain so much pigment as to appear red when in sunlight. This so-called greed-red shift is stimulated by high temperature and intense illumination.

The approximately spherical nucleus shows a central structure, the endosome (or more than one), which apparently lacks nucleic acid. The chromatin (chromosomal material) is usually seen as strings of granules.

(R. P. H.)

EUHEMERUS (EUEMERUS, EVEMERUS) (fl. *c.* 300 B.C.), Greek mythographer, is chiefly known by his *Sacred History,* a philosophic romance based upon archaic inscriptions that he claimed to have found during his travels in various parts of Greece. It is thought that he was born at Messina, though some claim that he was born at Chios, Tegea or Messene in the Peloponnesus. He lived at the court of Cassander, king of Macedonia, from 301 to 297 B.C.

Euhemerus relies particularly upon an account of early history that he said he discovered on a golden pillar in a temple on the island of "Panchaea" when on a voyage around the coast of Arabia, undertaken at the request of Cassander, who was his friend and patron. There is apparently no doubt that this island is imaginary. In his work he systematized for the first time an old oriental (perhaps Phoenician) method of interpreting the popular myths, asserting that the gods who formed the chief objects of popular worship had been originally heroes and conquerors who had thus earned

a claim to the veneration of their subjects. This system spread widely, and the early Christians especially appealed to it as a confirmation of their belief that ancient mythology was merely an aggregate of fables of human invention. Euhemerus was a firm upholder of the Cyrenaic philosophy (*see* CYRENAICS), and many ancient writers regarded him as an atheist.

His work was translated by Quintus Ennius into Latin, but was itself lost. Of the translation, only a few fragments survive.

The word "euhemeristic" is applied to such explanations of primitive myths. There is no doubt that his explanation contains an element of truth, for among the Romans the gradual deification of ancestors and the apotheosis (*q.v.*) of emperors were prominent features of religious development. Among primitive peoples, moreover, it is sometimes possible to trace the evolution of family and tribal gods from great chiefs and warriors.

All theories of religion that give prominence to ancestor worship and the cult of the dead are to a certain extent euhemeristic. But as the sole explanation of the origin of the idea of gods it is not accepted by students of comparative religion. It had, however, considerable vogue in France, and in the 18th century the *abbé* Banier, in his *Mythologie et la fable expliquées par l'histoire,* was frankly euhemeristic; to a great extent, Herbert Spencer also was a euhemerist.

EULALIUS (d. 423), antipope from Dec. 418 to April 419 against Boniface I. The rivalry led to the first interference of the temporal authorities in papal elections. For disobeying the emperor's orders Eulalius, till then imperial favourite, was finally rejected. According to one report he was subsequently banished to Campania, according to another he was created bishop of Nepi. *See* also BONIFACE; PAPACY. (JN. SR.)

EULENBURG (UND HERTEFELD), PHILIPP, PRINCE ZU (1847–1921), German diplomat and intimate friend and adviser of the emperor William II, was born in Königsberg on Feb. 12, 1847. He fought in the campaigns of 1866 and 1870–71 but left the army and in 1877 entered the diplomatic service. From 1881 to 1888 he was secretary to the Prussian mission in Munich, where he delighted in the city's artistic life; he returned in 1891 as Prussian minister.

A close friend of William II since 1886, after Bismarck's fall (1890) Eulenburg became the emperor's most influential adviser—not strong enough to control William's erratic personality but for years one of the few who were outspoken. In 1894 he refused the chancellorship but went as ambassador to Vienna, where he remained until 1902. His distinguished career ended tragically when, in 1906, the publicist Maximilian Harden, in his paper *Die Zukunft,* printed a series of attacks on Eulenburg's private life. He was arraigned on a charge of homosexuality, but his health broke and the trial was suspended. Though the charges were never proven, the scandal seriously damaged the monarchy's prestige. Eulenburg died at Liebenberg on Sept. 17, 1921. He wrote *Aus 50 Jahren* (1923), *Mit dem Kaiser als Staatsmann und Freund auf Nordlandreisen,* two volumes (1931), and *Erlebnisse,* two volumes (1934), the last two works being edited by his widow.

See J. Haller, *Philipp Eulenburg; the Kaiser's Friend,* Eng. trans. (1930); M. Baumont, *L'Affaire Eulenburg et les origines de la guerre mondiale* (1933).

EULENSPIEGEL (ULENSPIEGEL), **TILL,** a popular German peasant jester said to have been born at Kneitlingen, Brunswick, and to have died in 1350 at Mölln, Schleswig-Holstein, where his gravestone has been pointed out since the 16th century. By the end of the 15th century many anecdotes had gathered around his name. These seem to have been first collected about 1500 (perhaps earlier) and printed in one or more Low German versions. But the first extant text of the chapbook is a High German version, *Ein kurtzweilig lesen von Dyl Vlenspiegel geboren vss dem land zu Brunsswick,* published in Antwerp in 1515; the sole surviving copy is in the British museum, London. A second text, containing like the first many Low German words and forms, came from the same press in 1519. The jests and practical jokes, which generally depend on a pun, a deliberate literal interpretation of some metaphorical command, are broadly farcical, often brutal, sometimes obscene; but they have a serious theme. In the figure of Eulenspiegel, the individual gets his own back on society, the stupid yet cunning peasant demonstrates his superiority to the narrow, dishonest, condescending townsman, as well as to the clergy and nobility.

The Low German text was also translated into Dutch (*Van Vlenspieghels leuen,* after 1519) and English (fragments of a text, 1516–20?; *Here beginneth a merye Jest of a man that was called Howleglas, c.* 1560). The Dutch version is the basis of the first translation into French (1532) and into Latin verse (1558). There are also translations into several other European languages, as well as numerous editions, modern versions and adaptations both in and outside Germany. Eulenspiegel, who has become the representative of Low German humour, has been the subject of many literary and musical works; *e.g.,* Johann Fischart's *Eulenspiegel reimensweis* (1572), Richard Strauss's symphonic poem *Till Eulenspiegels lustige Streiche* (1894) and Gerhart Hauptmann's epic in hexameters *Till Eulenspiegel* (1928).

BIBLIOGRAPHY.—The 1515 text was edited by H. Knust in *Neudrucke deutscher Literaturwerke des XVI. und XVII. Jahrhunderts,* no. 55–56 (1884), and reproduced in facsimile by E. Schröder (1911). The 1519 text was reprinted as *Dr. T. Murners Ulenspiegel,* ed. by J. M. Lappenberg (1854). *See* also F. W. D. Brie, *Eulenspiegel in England* (1903); E. A. Roloff, *Ewiger Eulenspiegel* (1940); the article "Vlenspegel" in *Die deutsche Literatur des Mittelalters. Verfasserlexikon,* ed. by W. Stammler, vol. 4 (1953). (J. R. WE.)

EULER, LEONHARD (1707–1783), Swiss mathematician, who made his greatest contributions to modern analysis, of which he is considered one of the founders, was born at Basel, Switz., on April 15, 1707, and died at St. Petersburg, Russia, on Sept. 18, 1783.

A student under Jean Bernoulli, Euler in 1727 became, upon the invitation of Catherine I, an associate of the Academy of Sciences at St. Petersburg and a colleague of Daniel and Nicolas Bernoulli, sons of Jean. In 1730 he became professor of physics, and in 1733 he succeeded Daniel Bernoulli in the chair of mathematics. The author of innumerable papers, Euler overtaxed himself and in 1735 lost the sight of one eye. Invited by Frederick the Great, Euler in 1741 became a member of the Academy of Sciences at Berlin, where for 25 years he poured forth a steady stream of publications. When Frederick became less cordial, Euler in 1766 accepted the invitation of Catherine the Great to return to Russia. Soon after his arrival at St. Petersburg a cataract formed in the other eye, and Euler spent the last years of his life in total blindness. Despite this and other tragedies, his productivity continued undiminished, sustained by an uncommon memory and a remarkable facility in mental computation.

Euler's interests were broad, and his *Lettres à une princesse d'Allemagne* (3 vol., 1768–1772), an exposition of principles in physics and astronomy, was widely read; but his best work lay in the field of pure mathematics. Not a classroom teacher, Euler nevertheless had a more pervasive pedagogical influence than any other modern mathematician. From elementary geometry (where the line through the orthocentre, the circumcentre and the barycentre of a triangle bears his name) to advanced calculus (where the Eulerian integrals define the beta and gamma functions) Euler is remembered for familiar notations and important theorems. It was through his work that the symbols e, π and i came into common use, and it was he who related them through the equation $e^{\pi i}+1=0$.

Euler was responsible for the analytic treatment of the trigonometric functions as numerical ratios (rather than as geometric lines) and for relating them, through the Euler identities, with imaginary exponentials; and he gave the algorithmic treatment of logarithms as exponents, including the discovery that each number has infinitely many natural logarithms.

Euler's *Introductio in analysin infinitorum* (2 vol., 1748) did for modern analysis what the *Elements* of Euclid had done for ancient geometry, and the resulting tendency to arithmetize mathematics and physics has continued ever since. Euler's great textbooks in the calculus, *Institutiones calculi differentialis* (1755) and *Institutiones calculi integralis* (3 vol., 1768–1770, supp. 1794), have served as prototypes down to the 20th century. But Euler was far more than a textbook writer, and he enriched mathematics

with beautiful new results. He proved the addition theorem for elliptic integrals; he found the sum to infinity of the reciprocals of the squares of the integers and showed that the sum of the reciprocals of the primes less than n is asymptotic to $ln\ ln\ n$ as n increases indefinitely.

In the theory of numbers, Euler discovered the law of quadratic reciprocity (1772). Differential geometry got its first real start in Euler's study of lines of curvature (1760); and the calculus of variations took on independent status when Euler, in 1736, gave his differential equation expressing a necessary condition for a minimizing curve.

In the effort to replace synthetic methods by analytic, Euler was succeeded by J. L. Lagrange. But where Euler had delighted in special concrete cases, Lagrange sought for abstract generality; and while Euler incautiously manipulated divergent series, Lagrange attempted to place infinite processes upon a sound basis. Hence Lagrange, rather than Euler, often is regarded as the greatest mathematician of the 18th century; but Euler never has been excelled either in productivity or in the skillful and imaginative use of algorithmic devices.

See also references under "Euler, Leonhard" in the Index.

BIBLIOGRAPHY.—An account of Euler's achievements is found in the *Leben und Wirken* by O. Spiess (1929), and in Eric T. Bell, *Men of Mathematics* (1937). It was estimated that a complete edition of Euler's works would comprise at least 60 large quarto volumes. The *Opera omnia*, begun at Leipzig in 1911 and continued since 1942 at Lausanne, appeared in three series: (1) *Opera mathematica;* (2) *Opera mechanica et astronomica;* and (3) *Opera physica.* (C. B. BR.)

EULER-CHELPIN, HANS KARL AUGUST SIMON VON (1873–1964), German biochemist who in 1929 was awarded jointly with Sir Arthur Harden the Nobel prize for chemistry in recognition of their work on enzymes and fermentation of sugar. Euler-Chelpin was born on Feb. 15, 1873, at Augsburg, Bavaria. As a young man he abandoned painting for science and studied in Munich, at the universities of Würzburg, Berlin and Göttingen, and in the Pasteur institute, Paris. His teachers included W. Nernst and J. H. van't Hoff (*qq.v.*). In 1897 he became assistant in physics to S. A. Arrhenius in Stockholm and in 1900 he was appointed docent for physical chemistry, becoming professor of general and inorganic chemistry in 1906, and in 1929 being appointed director of the new biochemical institute there. He died in Stockholm on Nov. 7, 1964.

His researches on enzymes showed that when an enzyme acts on another substance, called the substrate, the chemical linkage between enzyme and substrate is a bond between an acid group and an alkaline group, and he studied more particularly the coenzymes, which act or "coact" with certain enzymes and which are necessary for their action. In the case of zymase, he isolated the coenzyme, cozymase, determined its chemical nature and showed it to be a nucleotide, containing phosphoric acid with a sugar residue and a purine residue, closely related to adenylic acid, which occurs in another form in muscle; and he was able to devise a method of measuring the purity of the coenzyme by its coaction. Knowledge of the biochemistry of sugars and phosphates was thus greatly advanced. The fermentation of sugar required that it should first be combined with phosphate by the action of the enzyme phosphorylase, the complex formed being then split by the action of the enzyme phosphatese, with exchanges of phosphate between the substrate and the different enzymes involved before the muscle obtained the energy available from the sugar. (D. McK.)

EULER NUMBERS (EULERIAN NUMBERS) are the coefficients of the expansion:

$$\sec x = \frac{1}{\cos x} = 1 + A_1x^2 + A_2x^4 + A_3x^6 + \ldots$$

(*See* TRIGONOMETRY: *Trigonometric Functions,* for an explanation of symbols.) They were so named by the German mathematician H. F. Scherk after the man who first discovered their significance in analysis. Although not of such wide application as the Bernoulli numbers, they are interesting because of their intimate connection with the former and because of their various theoretical properties. They are also used in the summation of certain series. The first nine of these numbers were computed by Leonhard Euler ($q.v.$). The ninth was found to be erroneous by R. Rothe, who gave the correct value. Scherk (1825) computed 6 more, J. W. L. Glaisher extended the list to 27, and S. A. Joffe computed 23 more, making 50 in all. Euler used the recurrent formula

$$\frac{A_n}{1} - \frac{A_{n-1}}{2!} + \frac{A_{n-2}}{4!} - \frac{A_{n-3}}{6!} + \ldots = 0$$

and independent representations in the form of determinants (*see* DETERMINANT) were given by Glaisher:

$$E_n = \begin{vmatrix} \frac{1}{2!} & 1 & \ldots & 0 & 0 \\ \frac{1}{4!} & \frac{1}{2!} & \ldots & 0 & 0 \\ \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \ldots & 1 & 0 \\ & & & \frac{1}{2!} & \\ \cdot & \cdot & \ldots & \cdot & 1 \\ \frac{1}{2n!} & \frac{1}{2n-2!} & \cdots & \frac{1}{4!} & \frac{1}{2!} \end{vmatrix}$$

and by M. Haussner as follows:

$$E_n = \begin{vmatrix} 1 & 1 & 0 & \ldots & 0 \\ 1 & \binom{4}{2} & 1 & \ldots & 0 \\ 1 & \binom{6}{2} & \binom{6}{4} & \ldots & 0 \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ 1 & \binom{2m-2}{2} & \binom{2m-2}{4} & \ldots & 1 \\ 1 & \binom{2m}{2} & \binom{2m}{4} & \ldots & \binom{2m}{2m-2} \end{vmatrix}$$

The following formula does not seem to have been given in the memoirs on the subject:

$$E_n = (2n)! - B_1(2n-2)! + B_2(2n-4)! + \ldots$$

where B_1 represents the sum of the products of the squares of the first $n - i + 1$ odd numbers taken as products i at a time with repetitions; *e.g.*,

$$\begin{aligned} E_4 = {} & 8! - 6!\,(1^2 + 3^2 + 5^2 + 7^2) \\ & + 4!\,(1^4 + 3^4 + 5^4 + 1^2.3^2 + 1^2.5^2 + 3^2.5^2) \\ & - 2!\,(1^6 + 1^4.3^2 + 1^2.3^4 + 3^6) + 1. \end{aligned}$$

J. Binet made the observation that the nth Eulerian number is equal to the number of permutations that can be formed of $2n$ elements, $a_1, a_2, a_3 \ldots a_{2n}$, such that the index of any element is either larger or smaller than each of the two adjacent indices; *e.g.*, out of the 24 permutations of the four elements, a_1, a_2, a_3, a_4, the following satisfy the above condition:

$$\begin{matrix} a_2\ a_3\ a_1\ a_4 \\ a_3\ a_1\ a_4\ a_2 \\ a_4\ a_2\ a_3\ a_1 \\ a_1\ a_4\ a_2\ a_3 \\ a_2\ a_1\ a_4\ a_3 \end{matrix}$$

totaling five distinct permutations, and hence $E_2 = 5$.

Of the theoretical properties of the Eulerian numbers the following are of interest:

1. Every Eulerian number is a positive odd integer.
2. The sum of any two successive Eulerian numbers is divisible by 3.
3. When n is even, $E_n + 1 \equiv 0 \pmod 3$.
4. When n is odd, $E_n - 1 \equiv 0 \pmod 3$.
5. E_n always ends in 1 or 5.

The first 10 Euler numbers are 1, 5, 61, 1385, 50521, 2702765, 199360981, 19391512145, 2404879675441, 370371188237525; and E_{50} is a number of 127 figures. (J. GI.)

EUMENES, the name of two rulers of Pergamum (*q.v.*).

EUMENES I (ruled 263–241 B.C.) succeeded his uncle Philetaerus in 263 B.C., won a victory near Sardis over Antiochus I of Syria in the next year, and thereby exchanged his former position as a vassal of the Seleucids for the position of an independent ruler. He died in 241.

EUMENES II (ruled 197–160 or 159 B.C.) was the eldest son of Attalus I, whom he succeeded as king in 197. He continued his father's policy of co-operation with Rome. At the battle of Magnesia (winter 190/189) he led the right wing and contributed largely to the Roman victory over the Seleucid king Antiochus III (*see* ANTIOCHUS). The peace of Apamea (188) brought him substantial reward: the Thracian Chersonese (Gallipoli peninsula) and the bulk of the former Seleucid possessions in Asia Minor. He was indeed easily the chief beneficiary of the peace. Yet he had the good sense to realize that for all the enhancement of his power he was still totally dependent upon Roman favour, and his policy continued to be one of treating the interests of Rome and Pergamum as identical.

He fought wars with the kings of Bithynia and Pontus; in both cases he enjoyed the moral support of Rome. In 172 he visited Rome and denounced the king of Macedonia, Perseus (*q.v.*), for having aggressive designs; and when Rome declared war on Perseus in the next year, Eumenes led a Pergamene force to fight alongside the Romans. When this war was protracted, it was rumoured that Eumenes had begun negotiations with Perseus. Whatever the truth of the rumour, the mere suspicion of double-dealing cost Eumenes the favour of Rome, and he spent the rest of his life under the shadow of Roman displeasure. He died in 160 or 159.

A brilliant statesman, he brought his kingdom through perils to its highest point of power. In doing so he inevitably made many enemies, and it is understandable that in ancient tradition he acquired the reputation of being an unscrupulous man. He was responsible for the construction of nearly all the main public buildings on the acropolis at Pergamum; more than any other of the Attalid kings he made the city a great cultural centre.

See works cited under ATTALID DYNASTY. (R. H. SI.)

EUMENES (*c.* 362–316 B.C.), Greek general from Cardia in the Thracian Chersonese, is best remembered for his part in the civil war within the Macedonian empire after the death in 323 of Alexander III the Great. Before then he had been Alexander's principal secretary. In the distribution of satrapies at Babylon after Alexander's death Cappadocia was assigned to him, and when the regent Perdiccas (*q.v.*) was challenged by his Macedonian enemies Antipater, Craterus and Ptolemy, Eumenes gave him valuable aid. The death of Perdiccas left him isolated, and he was condemned to death by the victorious generals gathered at Triparadisus (321), Antigonus Monophthalmus (*q.v.*) being charged with the execution of the sentence. Though severely defeated by him in 320, Eumenes escaped; having in 319 been recognized by the new regent Polyperchon as the royal general in Asia, he collected an army in Cilicia. With this he marched toward the eastern provinces pursued by Antigonus. At Susa the assembled governors of the eastern satrapies joined hands with him (318). In a long and hard campaign on the Iranian plateau Eumenes held Antigonus in check very successfully, but at the end he was betrayed to the enemy by his own men and put to death.

Eumenes was an extremely able general, and his achievements are all the more remarkable for the fact that in all his battles he commanded soldiers whose loyalty was in doubt. As a Greek, he was disliked by the Macedonians, who occupied nearly all the leading positions in the empire. For this reason he may have felt that his safest course would be to show unswerving loyalty to the Macedonian royal house, rather than to offer his support to any individual, such as Antigonus. Whatever his motive, his complete devotion to the legitimist cause cannot be disputed; to this he sacrificed his life.

Plutarch, in composing his biographies, very appropriately chose Eumenes as comparable with the Roman soldier Sertorius (*q.v.*) in nobility of character. The sympathetic portrayal of Eumenes in the ancient sources derives from the lost history of Hieronymus, who was a fellow-townsman and close personal friend.

See H. D. Westlake, "Eumenes of Cardia," in *Bulletin of the John Rylands Library,* vol. 37, 1 (1954). (R. H. SI.)

EUMENIDES (Greek *eumenēs,* "kindly"), ancient goddesses, also called *semnai* ("reverend ones"), worshiped at the foot of the Areopagus at Athens, at the village of Colonus just north of Athens and in places outside Attica. Their cult closely resembled that of Gaea (Earth), and they were probably earth spirits, largely concerned with fertility, but having also certain moral and social functions. By Aeschylus and later writers they were generally identified, although wrongly, with the Erinyes (*q.v.*).

See L. R. Farnell, *The Cults of the Greek States,* vol. v, p. 440 (1896–1909).

EUMENIUS (fl. *c.* A.D. 300), Roman orator and teacher of rhetoric, born at Autun, was the author of the oration *Pro instaurandis scholis* ("For the Restoration of the Schools"), delivered at Augustodunum (Autun) in 298 in the presence of the governor of the province of Gallia Lugdunensis. Ostensibly a plea for restoration of the "local university college" (*Maenianae scholae*), it is really a panegyric of the emperor Constantius Chlorus in appreciation of plans already made for restoration of the college and steps taken to repair the damage suffered by the city during the disturbances and siege of 269. Eumenius was a professor of rhetoric, born (but not educated or previously employed) at Autun, where his grandfather, a Greek immigrant from Athens, had, after a successful career at Rome, been professor of rhetoric until well into his 80s. In 298 the principal of the college had died, and part of Constantius' reconstruction scheme was the appointment of Eumenius as principal. Eumenius had been Constantius' private secretary and, to maintain his status, was appointed principal at double the normal salary. These emoluments, however, he declared he would donate to the reconstruction fund.

Among much rhetorical and mythological preciosity, he gives some interesting details of the teaching of geography by means of maps; an ancient map inscribed on marble, found at Autun and subsequently buried in the course of building, has been conjecturally associated with these. From the 16th century onward other contemporary speeches in the collection of *Panegyrici Latini* have been ascribed to Eumenius, but without evidence or probability. There is no sign of Christian belief in Eumenius; but the panegyrists are all noncommittal in this respect. In him the frigid affectations of panegyric are enlivened by personal enthusiasms.

BIBLIOGRAPHY.—For *Pro instaurandis scholis* see W. Baehrens (ed.), *XII Panegyrici Latini* (1911). *See* also E. Galletier, *Panégyriques Latins,* vol. i (1949); W. S. Maguinness in *Greece and Rome,* vol. xxi, no. 63 (1952). (W. S. MA.)

EUMOLPUS, mythical ancestor of the priestly clan of the Eumolpidae at Eleusis (*see* MYSTERY), in ancient Greece. The name ("good" or "strong singer"; *i.e.,* priest who can chant his litanies clearly and well) is an obvious personification of their hereditary functions. As might be expected in the case of so shadowy a figure, his legend fluctuates greatly, so much so that three Eumolpi have been assumed: (1) Being a "sweet singer," he is naturally connected with Thrace, the country of Orpheus (*q.v.*). He is the son of Poseidon and Chione ("Snow girl"), daughter of Boreas; after sundry adventures he becomes king in Thrace, is invited to help the Eleusinians in their war with Erechtheus and is killed in the war. (2) As one of the originators of the Eleusinian mysteries, he is an Eleusinian, a son of Earth, father of Keryx, the mythical ancestor of the Kerykes (heralds), taught the mysteries by Demeter herself. (3) Since Orpheus and all his following were closely connected with mysteries of all sorts, Eumolpus is son, father or pupil of Musaeus, a mythical singer closely allied with Orpheus.

Further confusion is introduced by his being sometimes called son, not father, of Keryx, possibly a reflection of some old dispute over precedence between the two clans. As a priest, he purifies Heracles from the blood of the Centaurs and also initiates him; as connected with the Orphic circle, he is a culture-hero, maker of various advances in the useful and fine arts. His tomb was shown both at Eleusis and in Athens.

EUNAPIUS (*c.* 345–after 414), Greek rhetorician and historiographer whose writings supply valuable information on contemporary Neoplatonism, was born at Sardis in Lydia (western Asia Minor). Educated at Athens under the rhetorician Proaeresius, he was initiated into the Eleusinian mysteries and became a priest of that cult. Hostile to Christianity, he sympathized with the Neoplatonists and included much about them in his *Lives of the Philosophers and Sophists* (ed. with Lat. trans. by J. F. Boissonade, 1849; with Eng. trans. by W. C. Wright, *Philostratus and Eunapius*, 1922). He also wrote a supplement to the *Chronological History* of Publius Herennius Dexippus, continuing the story from A.D. 270 down to 404. Of this work only fragments remain (ed. by C. W. Müller, *Fragmenta historicorum Graecorum*, iv, 1851; and by L. Dindorf, *Historici Graeci minores*, i, 1870), but the 5th-century historian Zosimus used it extensively in compiling his own work.

EUNUCH, a castrated human male. Castration as a form of punishment was practised in Assyria in the 2nd millennium B.C., and from remote antiquity, in the far and middle east, eunuchs were employed as harem officials (*see* HAREM). Their position as the personal servants of a ruler's mother might well lead to considerable personal influence over the ruler himself, and they came to be used as the confidential assistants of rulers. Eunuchs were almost always castrated either for a crime or in childhood after having been sold by poor parents; they were therefore people who were completely dependent on and owed everything to their rulers. Eunuchs were used as political advisers in China during the Chou period, and also under the Han, T'ang, Ming and Sung emperors. The Achaemenid Persians employed political eunuchs. The Roman emperors Claudius, Nero, Vitellius and Titus also employed many political eunuchs, as did most of the later orientalized emperors of Byzantium. Besides acting as confidential advisers, eunuchs were used also as bodyguards and as generals and admirals. Many of the patriarchs of Constantinople were eunuchs. Political eunuchism also flourished in the centres of Muslim power during and after the Abbasid caliphate. It did not occur to any great extent in Hindu India, nor was it found in America. Eunuchs were used, for both domestic and political purposes, in several west African states (Hausa, Yoruba, Nupe, etc.). The trade of castrating boys to be sold as eunuchs for Muslim harems has continued to modern times, the principal districts from which they are taken being north-central Africa (Baguirmi, etc.). The Italian practice of castrating boys in order to train them as adult soprano singers (*castrati*) ended with the accession of Pope Leo XIII (1878).

Voluntary eunuchs who emasculated themselves, or caused the operation to be performed on them, for the avoidance of sexual sin or temptation, have appeared in several Christian periods, acting on the texts Matt. xix, 12; v, 28–30. Origen's case is the most celebrated example, and by the 3rd century A.D. there had arisen a sect of eunuchs, the Valesii, who castrated themselves and their guests in the belief that they were thereby serving God.

(J. F. M. M.)

EUNUS, the leader of a slave revolt in Sicily against the Romans in 135 B.C. A Syrian by birth, he was a slave at Enna, where he gained the confidence of other slaves in a revolt against their masters. Before long 70,000 slaves were organized into a fighting force. Enna was captured, and Eunus, who called himself King Antiochus, soon controlled much of central and eastern Sicily. The first armies sent by Rome were badly defeated. L. Calpurnius Piso, consul in 133, however, managed to reach Enna, while his successor P. Rupilius in 132 stamped out the revolt. Eunus, who was captured, died in prison. (H. H. Sd.)

JOHN H. GERARD

BERRIES AND LEAVES OF EASTERN WAHOO (EUONYMUS ATROPURPUREUS)

EUONYMUS (EVONYMUS), a genus of deciduous or evergreen shrubs or small trees of the staff-tree family (Celastraceae). It comprises more than 100 species, widely distributed in the north temperate zone. A common species is the Eurasian *E. europaeus*, the spindle tree. It is a shrub or small tree growing in copses or hedges, with a gray smooth bark, four-angled green twigs, opposite leaves and loose clusters of small greenish-white flowers. The ripe fruit is a pale crimson colour and splits into four lobes, exposing the bright orange-coloured seed. Besides the spindle tree, which has become naturalized in the eastern United States, the genus is represented in North America by *E. americanus* (strawberry bush), *E. obovatus* (running strawberry bush), *E. atropurpureus* (eastern wahoo) and *E. occidentalis* (western wahoo). Numerous species are in cultivation, among which are *E. japonicus*, a handsome evergreen often with variegated leaves, and *E. fortunei*, a hardy climber for walls, especially the variety *vegetus*, which bears profuse persistent fruits. The flowering was said to foretell plague. *See* SPINDLE TREE; WAHOO.

EUPATORIUM, a large genus of plants of the family Compositae (*q.v.*), comprising 500–700 species, nearly all American and found chiefly in tropical South America, the West Indies and Mexico. They are mostly perennial herbs, a few are annuals and many tropical species are shrubby or treelike. The leaves are usually opposite and the purplish, rose or white flowers are borne in rayless heads disposed usually in flat-topped clusters. Many species are found in the U.S. and Canada, widely distributed but mostly in the eastern, southern and southwestern states, very few occurring on the Pacific coast. Representative North American species are *E. perfoliatum*, boneset (*q.v.*); *E. purpureum*, joe-pye weed; *E. rugosum*, white snakeroot; and *E. coelestinum*, mistflower. Of the few old-world species, one, *E. cannabinum*, the hemp agrimony, occurs in Great Britain. Various tropical species are in greenhouse cultivation and several hardy species are grown as border plants.

JOHN H. GERARD

JOE-PYE WEED (EUPATORIUM PURPUREUM)

EUPATRIDAE in ancient Greece meant the old nobility of Attica. The Greek word means "sons of noble fathers," and is so used in a nontechnical sense in literature. At Athens it had a technical sense as the name of a definable noble caste; it was also, confusingly, the name of one particular noble clan (*genos*).

In classical times the court at the Prytaneum, which conducted a formal trial of unknown murderers and of animals or inanimate objects which had caused the death of a man, was composed of the *phylobasileis* ("tribal kings") under the presidency of the *basileus* (the archon called "king"; *see* ARCHON) and these *phylobasileis* were chosen from the *eupatridae:* they were also concerned with certain sacrifices. There were also *exegetae* ("expounders" of ritual) chosen from the *eupatridae*, whose inherited knowledge of sacred law was not published in written form till Roman times. These are mere vestiges: in classical times the nobility of certain famous houses was of political importance, but not mere membership of the eupatrid class; and on religious questions of public importance Athens then took the advice of professional seers (*manteis*) rather than the *exegetae*.

In 580, when Damasias was deposed after illegally retaining the archonship for over two years, the archonship was somehow shared between five *eupatridae*, three "farmers" (*agroikoi, georgoi, geomoroi*) and two "craftsmen" (*demiourgoi*): the mechanics of this are not clear, but it probably represents a half-and-half compromise between nobles and non-nobles. This is the only recorded appearance of these classes in practical politics, but it is reasonable to suppose they were of great importance in the aristocratic period before Solon, of which few details are known. Plutarch describes their creation by Theseus (*q.v.*), but his account seems to be

based on a late theoretical construction. There are signs that the *phylobasileis* once had a more extended competence, and the *exegetae* will have been important in the period before laws were written down. It is likely that public office was then in practice confined to the *eupatridae*, and that they had a political monopoly comparable to that of other Greek aristocracies in the archaic period. Solon's reforms (594), which established property qualifications for office, were an important stage in the limitation of their power, and after 580 these archaic classes in practice disappeared.

The clan Eupatridae included Alcibiades among its members, and survived to a late period. It was excluded from the sacrifices offered to the Eumenides (*q.v.*) by the *genos* Hesychidae: there is no other evidence for the hypothesis, framed to explain this exclusion, that it claimed descent from or connection with Orestes (*q.v.*), and the origin of the name remains obscure.

BIBLIOGRAPHY.—Aristotle, *Constitution of Athens*, 57.4, 13.2; Plutarch, *Theseus*, 25; material on the *exegetae* is collected in F. Jacoby, *Atthis*, ch. 1 (1949); for the *genos see* Isocrates 16.25 and Polemon fr. 49. *See* also G. Busolt and H. Swoboda, *Griechisches Staatsrecht*, vol. ii, p. 772 (1926); C. Hignett, *History of the Athenian Constitution*, ch. 3–4 (1952). (A. As.)

EUPEN-ET-MALMÉDY, a region along the Belgo-German frontier, belongs to the Verviers *arrondissement* in Liège province, Belg., which covers the so-called *cantons rédimés* ("redeemed cantons") of Eupen, Malmédy and St. Vith.

Until 1794 the region was part of the duchy of Limbourg, the ecclesiastical principality of Stavelot-Malmédy and the duchy of Luxembourg. From 1794 until 1814, under French rule, it belonged to the Ourthe *département* (the present province of Liège). The annexation by Prussia (treaty of Vienna, May 31, 1815) was ratified at the Vienna congress of June 9, 1815. Some of the people were Walloon and French speaking and some belonged to the Netherlands language group among whom German became the dominant tongue. The territory of Moresnet was much contested because of the then important zinc mines. One part was allocated to Prussia, another to the kingdom of the Netherlands and the third part became a condominium known as neutral Moresnet. After World War I the Versailles treaty assigned Prussian Moresnet (west of the Liège-Aachen road) and neutral Moresnet to Belgium together with Eupen, the region of Malmédy and St. Vith. It also provided for registers to be opened in the region (voting to take place between Jan. 23 and July 23, 1920) in order that the 63,000 inhabitants of the cantons could say whether they wished the territories to remain as a whole or partly under German sovereignty. Only 271 inhabitants voted to remain German and the council of the League of Nations' proposal that districts should be transferred to Belgium was accepted unanimously by the League's general assembly. In 1925, by the Locarno treaty, Germany recognized the new frontier.

After the German invasion of Belgium in 1940 Hitler transferred the territories to Germany, but this was undone by the liberation of Belgium at the end of 1944. After World War II frontier adjustments were demanded. In 1949 the Council of Foreign Ministers allowed the transfer to Belgium of 11.5 sq.mi. of enclaves along the Belgo-German frontier. Belgium, however, took possession of only 7.75 sq.mi., and on Sept. 24, 1956, a Belgo-German agreement was signed in Brussels returning four villages and two forests to Germany. The agreement came into force on Aug. 28, 1958. After these readjustments the territories covered an area of approximately 410 sq.mi.

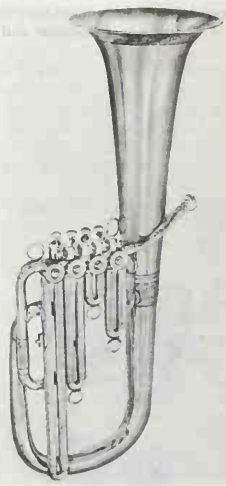
Most of the cultivated areas are pastureland and St. Vith is known for its cattle market. For centuries there was a famous leather industry and, in the Eupen region, a woolen industry connected with Verviers. Eupen also produces electric cables, metal wires, chocolates and soap. Malmédy has a prosperous paper industry and brewery. St. Vith produces furniture.

On Jan. 1, 1960, the Eupen canton included one city (Eupen) and eight villages; Malmédy, one city (Malmédy) and nine villages; St. Vith, one city (St. Vith) and ten villages. The population of the cantons was Eupen 26,216; Malmédy 22,198; St. Vith 17,678; or in total 66,092 as compared with 62,855 in 1947 when the previous census was taken. At that time, 44,921 inhabitants spoke only or mostly German. Crombach in St. Vith canton contained the highest percentage of German-speaking people (98%) and Bervercé in Malmédy canton the lowest (6%). (M. H. St.)

EUPHEMIA, SAINT (3rd/4th century), an early Christian martyred at Chalcedon probably during the persecution of Diocletian (*c.* 303). No credence should be given to the fantastic stories of her life and death in the Greek accounts of her passion. In 451 the Council of Chalcedon, which defined the doctrine of the two natures of Christ and condemned the monophysite heresy, met in the basilica dedicated to her and thereafter her fame became widespread. Churches were built in her honour, particularly in Constantinople and Rome, her relics were distributed, she was portrayed on mosaics, for instance at Ravenna and Parenzo (Porec), scenes from her martyrdom were depicted on frescoes and children were named after her. To her two festivals on April 13 and Sept. 16, attested in both east and west, the Greeks added a third on July 11 to commemorate the Chalcedonian definition, which Euphemia had traditionally supported by a miraculous indication. It was also claimed that at certain times liquefied blood which worked miracles could be collected at her tomb.

BIBLIOGRAPHY.—For editions of Greek accounts of her martyrdom, etc., see *Bibliotheca hagiographica Graeca*, 3rd ed. by F. Halkin, nos. 619–624n (1957). *See* also A. M. Schneider, "Sankt Euphemia und das Konzil von Chalkedon," in A. Grillmeier and H. Bacht (eds.), *Das Konzil von Chalkedon*, vol. i, pp. 291–302 (1951), and "Das Martyrion der hl. Euphemia beim Hippodrom zu Konstantinopel," *Byzantinische Zeitschrift*, 42:178–185 (1943); H. Grégoire, "Ste. Euphémie et l'empereur Maurice," *Le Muséon*, 59:295–302 (1946). (F. Ha.)

EUPHONIUM, in the U.S. called baritone, a brass wind instrument with valves, pitched in B flat an octave below the cornet or trumpet, with a wide tubalike bore and often in the U.S. with the bell turned forward. In military and brass bands it is the leading instrument in the tenor-bass range, with a role corresponding to that of the cello in the orchestra. It normally carries a fourth valve in addition to the essential three, in order to take the compass continuously down to the fundamental pitches below the bass stave, from which the total compass rises to the middle of the treble stave. Its notation is in the bass clef at actual pitch in military bands and in the treble clef a ninth above actual sounds in brass bands. In "duplex" models, an alternative bell is fitted; it is switched in by a valve to supply a lighter, tenor-horn tone quality. *See* TUBA; WIND INSTRUMENTS. (A. C. BA.)



BY COURTESY OF BOOSEY & HAWKES LTD.

EUPHONIUM WITH FOUR ROTARY VALVES, BUILT BY HENRY DISTIN, ABOUT 1860. IN THE BOOSEY AND HAWKES COLLECTION

EUPHORBIA, a large genus of plants from which the spurge family (Euphorbiaceae; *q.v.*) takes its name. It includes about 1,600 species and is of almost world-wide distribution. It is represented in North America and Great Britain by the spurges—plants with simple leaves and inconspicuous flowers arranged in small cuplike heads (cyathia). The cyathium is characteristic of the genus and consists of a number of tiny male flowers, surrounding a central female flower, the group of flowers being enveloped in a conspicuous cup formed by the union of four or five bracts. In some species the bracts and petallike extensions from nectar-bearing glands between the bracts are often brilliantly coloured, giving the whole inflorescence the appearance of a single flower, as in the poinsettia (*q.v.*). All *Euphorbia* have a milky juice, or latex, which in some species is poisonous, as, for example, in the popular garden annual snow-on-the-mountain (*E. marginata*).

Many south African species closely resemble cacti, having a thick succulent stem and branches with the leaves very small or reduced to a small wartlike excrescence. These occur in the desert and are much cultivated for hedges in tropical America.

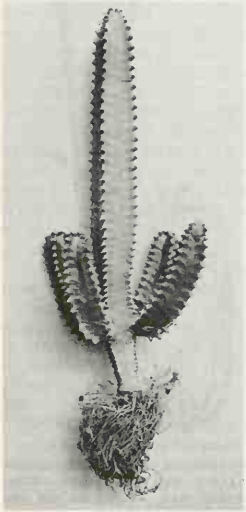
EDGAR AUBERT DE LA RÜE

ARBORESCENT EUPHORBIA (E. ABYSSINICA), A CACTUSLIKE PLANT OF FRENCH SOMALILAND

EUPHORBIACEAE, a large family of flowering plants containing over 300 genera with 7,000 or more species, chiefly tropical, but occurring worldwide, except in the arctic and cold alpine zones. They are represented in the temperate zones primarily by a number of herbaceous species, but many species are woody, often trees.

In size the species range from small, prostrate herbs to large trees of the tropical forest. Many African species are cactuslike succulents; woody and herbaceous vines and floating aquatics are also represented. The leaves are commonly simple and pinnately veined, but are palmately veined or divided in some, or palmately compound, as in the Pará rubber tree (*Hevea braziliensis*). The leaves are reduced to inconspicuous scales in many succulent forms and in some species of *Phyllanthus* that have wide, flat branches of leaflike form and function. Many other species of *Phyllanthus* bear deciduous branches that simulate pinnately compound leaves.

The functional stamens and pistils of the Euphorbiaceae are borne in separate flowers on the same or different plants. The ovary is always borne above the other flower parts. The sepals and especially the petals are often small or absent, but there are species with large and showy petals, as in *Aleurites* and *Jatropha*. The stamens vary from one to many in number and are variously united in some species. The castor-oil plant is unusual in bearing branched stamens. The ovary is composed of from one to several united carpels, but the preponderant number is three, often supplying a useful recognition feature for the family. One or two seeds are borne in each carpel, the frequently oily seeds often bearing a fleshy outgrowth, or caruncle. Some species bear drupes or berries, but the fruit is usually a capsule which explodes with considerable force, throwing the seeds yards away.



BY COURTESY OF L. CUTAK

EUPHORBIA CANARIENSIS, A CACTUSLIKE PLANT OF THE CANARY ISLANDS

The basic pattern of inflorescence for the family is the dichasium, that is, a terminal flower beneath which are borne two buds, each developing into a flower subtended by two buds, etc. There are many modifications of this pattern, but the most interesting is the cyathium found in *Euphorbia* and related genera: this is a compact flower cluster that closely simulates a simple flower. In the centre of the cyathium is a terminal female flower, which is naked or bears only a scalelike calyx, resembling the pistil of an ordinary flower. This female flower is surrounded by five clusters of tiny male flowers, each consisting of a single naked stamen. Except for two African genera that bear small calyxes on the male flowers, these are only distinguished from simple stamens by a joint between the flower and the pedicel. Each cluster of male flowers is enfolded by a sepal-like bract. Nectar-bearing glands are usually borne between the bracts and, in many species, each gland develops a coloured petal-like structure. In these species the resemblance between the whole inflorescence and a simple flower is especially striking.

Many members of the family produce a milky juice or latex, which exudes when the plant is injured and which contains starch, resins, sometimes a poison and often rubber.

Cross-pollination is the rule in the Euphorbiaceae, as the stamens and pistil are borne by different flowers. In *Mercurialis* and others with inconspicuous flowers pollination is by the wind, but in many cases insects are attracted to the flower by highly coloured bracts or flower parts; the presence of nectar is also frequently an attraction, as in the glands of *Euphorbia* and other genera.

Economically the most important species of the family is the Pará rubber tree. Rubber may also be obtained from other species of *Hevea* and from *Manihot, Euphorbia, Sapium* and some other latex-producing genera, but in many the quantity is small or the resin content too high. *Manihot esculenta* (manioc or cassava; *q.v.*) is an important tropical food plant; its thick, starchy, tuberous root is the source of tapioca. Castor oil (*q.v.*) is obtained from the seeds of *Ricinus communis*. The seeds of *Aleurites fordii* (tung tree), *A. moluccana* and *A. cordata* yield important commercial oils, while the waxy coating of the seeds of *Sapium sebiferum* is used for various purposes. Resin is obtained from species of *Croton* and *Euphorbia*. The fruits of various species of *Baccaurea, Phyllanthus* and *Antidesma* are edible and some of them are cultivated for their fruits. Many of the species are poisonous; *e.g.*, the south African *Toxicodendron* and the tropical American *Hippomane*, or machineel. Many, such as *Euphorbia, Mercurialis, Croton, Jatropha* and *Tragia*, have been, or still are, used as medicines. Species of *Acalypha, Codiaeum, Euphorbia, Jatropha, Phyllanthus, Poinsettia* and others are used as ornamental plants.

The relationships of the Euphorbiaceae among other plant families are unclear. The small family Buxaceae (including the boxwood, *Buxus*) is closely related and has at times been included in the Euphorbiaceae. The Malvaceae and Geraniaceae have been considered as possible close relatives.

See F. Pax and K. Hoffmann, "Euphorbiaceae" in Engler and Prantl, *Die natürlichen Pflanzenfamilien*, 2nd ed., 19c:11–232 (1931) for a general treatment of the Euphorbiaceae; A. White, R. A. Dyer and B. L. Sloane, *The Succulent Euphorbiaceae*, 2 vol. (1941).

(R. L. D.)

EUPHORION (born *c.* 275 B.C.), Greek poet and grammarian, was highly esteemed by the contemporaries of Catullus and their successors, who were dubbed *cantores Euphorionis* by Cicero. Born about 275 B.C. at Chalcis, in Euboea, Euphorion studied philosophy at Athens, where he spent much of his life and amassed wealth, it was said, by a liaison with an elderly widow. Soon after 223 B.C. Antiochus the Great made him director of the royal library at Antioch, in Syria. His poetry comprised mythological epics, *e.g.*, *Mopsopia*, a collection of Attic legends, and *Thrax*, a hotchpotch of myths from various sources; invectives, including the *Curses* on a man who had stolen his goblet; and epigrams, two of which are preserved in the Palatine anthology. He also composed prose works on antiquities and grammar.

Increased knowledge of Euphorion as a poet derived from the papyri has made it no easier to understand the reputation he enjoyed in Rome. He was a habitual plagiarist, his style is affected and his language obscure and he preferred subjects that were horrific and morbid.

BIBLIOGRAPHY.—J. U. Powell, *Collectanea Alexandrina* (1925); D. L. Page, *Greek Literary Papyri, i, Poetry*, Loeb series (1942); V. Bartoletti (ed.), *Papiri Greci e Latini*, vol. 14 (1957).

(E. A. B.)

EUPHRANOR, of Corinth (dated by Pliny in 364 B.C.), a Greek sculptor and painter. In the Stoa Basileios at Athens he painted the "Twelve Gods," "Theseus With Democracy and Demos" and the cavalry engagement at Mantinea (362 B.C.); at Ephesus the feigned madness of Odysseus. A statue found in the Agora at Athens has been identified with his "Apollo Patroos," mentioned by Pausanias.

Other recorded statues are "Leto With Her Children Apollo and Artemis," "Philip" and "Paris and Alexander in Chariots." He wrote on proportion and colour.

For attempts to identify copies *see* A. Furtwängler, *Masterpieces of Greek Sculpture*, pp. 348 ff. (1893).

EUPHRATES (Arab. NAHR AL FURAT; Turk. FIRAT NEHRI), the largest river of western Asia and one of the historic rivers of the world. Rising on the Armenian plateau in Turkey, it flows generally southeastward across Syria and southern Iraq where it joins the Tigris to form the Shatt al Arab (*q.v.*) and thence to the Persian gulf, a total length of 2,235 mi. Its course may be divided into three sections: an upper, essentially within the Turkish highlands and extending as far as the village of Samsat; a middle, which crosses the eastward-sloping Syrian plateau to Hit; and a lower, which flows out onto the alluvial lowlands of Iraq to its junction with the Tigris at Al Qurnah. The importance of the Euphrates and Tigris together cannot be overemphasized—not only in their interriver relations but as the entire basis for the geography of ancient Mesopotamia as well as modern Iraq. In a desert climate, the people of the region depend on these twin rivers for their very existence, and their degree of economic development depends upon the effectiveness of their utilization and control of these rivers.

The Upper Euphrates.—The upper Euphrates begins with two principal tributaries, the Kara Su (or "muddy"), the more northerly branch, and the Murat Su (or "clear"), the easterly branch. These two streams flow off the high Armenian plateau in a series of relatively wide valleys linked by narrow, deep gorges. They unite about 30 mi. N.W. of the town of Elazig. The Kara Su rises on the Dumlu Dagi just north of Erzurum at an altitude of nearly 9,000 ft. above sea level. Its upper valley communicates by a low, easy pass, commanded by the defenses of Erzurum, with the head of the Aras (ancient Araxes) valley, thus becoming a part of the ancient commercial and military east-west route across the Armenian plateau. The Kara Su receives a number of minor tributaries before its junction with the Murat Su approximately 275 mi. S.W. of its source. In the last part of its course it cuts through the Munzur Silsilesi in a gorge noted for its wild scenery and precipitous walls.

The Murat Su rises on the northern side of the Ala Dagh at an altitude of more than 10,000 ft. above sea level. Its valley has never been an important thoroughfare, partly because of the presence of narrow, rocky gorges in its lower course and partly because of its isolation. Only its uppermost valley, where it is joined by the Sharian Su near Karakose, forms a trough on the south side of the Agri Dagi (Mt. Ararat) through which passes the route between Erzurum and Iran. Near Elazig the Murat Su is joined by its most important tributary, the Perisuyu, which rises on the Bingol Daglari near the source of the Aras and drains the mountainous region between the two main branches of the upper Euphrates. As the Murat Su joins the Kara Su it is the longer, being approximately 415 mi. from its source.

From the Kara-Murat confluence the Euphrates proper continues in a great reverse curve between the Malatya and the Ziyaret Daglari, major ranges of the Taurus mountain system of southern Turkey. It is in this region that the Euphrates flows around a lava mass erupted from the Karacali volcano and through several spectacular gorges as it cuts across anticlines in the Taurus system. Especially noteworthy are the Kemer Khan canyon above Tilek, with 12 mi. of perpendicular walls of basalt and marble, and the Gergez gorge below it. From the junction of the two main tributaries to the point where it emerges from the highlands to the level of the Syrian plateau at Samsat, a distance of about 110 mi., the Euphrates drops nearly 1,000 ft. or nearly 10 ft. per mile. Boiling rapids and cataracts fill portions of the river. The course of the Euphrates in the lower part of this section is controlled by the northward continuation of the so-called Syrian trench, the rift valley in which the Gulf of Aqaba, the Dead sea and the Jordan-Litani-Orontes rivers are located.

The Middle Euphrates.—The middle portion of the river, from its emergence onto the Syrian plateau at Samsat to the Iraqi lowlands at Hit, is approximately 900 mi. long. In this distance the Euphrates drops from about 650 ft. above sea level to about 220 ft., a gradient averaging nearly seven inches per mile. The river occupies a typically steep-sided valley cut several hundred feet into the plateau surface through strata of marls and gypsum overlaid with sandstone and topped with the breccia of the desert surface. The flood plain is seldom less than two, or more than four, miles wide, and the river winds across it.

It is there in its middle course that the Euphrates receives its only important tributaries between the Kara-Murat junction and its confluence with the Tigris. The Sajur joins it on the right bank about 15 mi. downstream from Jerablus, and the Belikh enters on the left bank about halfway between Jerablus and Deir-ez-Zor. Below Deir-ez-Zor the river bed broadens out somewhat and contains numerous rocky stretches, shallow rapids and islands. The last important tributary, the Khabur, joins the Euphrates on the left bank about 20 mi. below Deir-ez-Zor. It rises on the south-facing slopes of the Taurus mountains in Turkey and drains northeastern Syria. No tributaries of any size are received on the right bank below the Sajur, but a number of dry river beds indicate that in relatively recent geologic time a number of tributaries were received by the Euphrates from what are now the Syrian and Arabian deserts.

The riverine alluvium on the valley floor of the middle Euphrates is irrigated somewhat by simple lift techniques and occasionally in the lower portion between Anah and Hit, by great wheels called *naurahs*. These clumsy mechanisms, some of them 40 ft. in diameter, are equipped with jars or paddles which carry water upward and empty it into aqueducts. Such irrigation supports a sparse rural population. Where a major route crosses the Euphrates, or where it is joined by a tributary, there is sufficient support for a town. Such is the case for Birecik on the Urfa-Gaziantep road in Turkey; Jerablus on the Turkish-Syrian frontier where the Euphrates is crossed by the Baghdad railway; Meskene, where the river road leaves the valley and goes westward to Aleppo; Raqqa, near the mouth of the Belikh; Deir-ez-Zor, at the meeting of the western road via Palmyra and the eastern road to Mosul, as well as the river road, and where there is a highway bridge across the river; and Abu Kemal, on the Syrian-Iraqi frontier.

The Lower Euphrates.—As the Euphrates emerges at Hit from its entrenched valley in the Syrian plateau and spreads out onto the plains of Iraq it decreases in both volume and velocity. In the dry climate of that region more water is lost by evaporation from the surfaces of the river and overflow marshes and by irrigation than is contributed by the upstream tributaries. At Hit the Euphrates may be said to enter its delta. From there on the river approaches a base-level gradient, falling on the average less than four inches per mile in the 700-mi. distance to its confluence with the Tigris. In the last 100 mi. of this part of its course the Euphrates drops only one inch per mile.

This rapid decrease in the velocity of the river as it levels out has caused widespread deposition of sediment on the delta plain. Lower Iraq is constructed of the materials laid down by both the Euphrates and the Tigris. With the accumulation of these alluvial deposits the drainage of lower Iraq has become increasingly poor and is held in a delicate balance between the rate of deposition and the ability of the river to flow off the land into the sea. This difficult drainage is accompanied by the extensive development of braided channels, marshes and shallow permanent lakes. These absorb much of the flow of the Euphrates and fluctuate greatly according to the season, receding during low water in the fall and expanding at high water in the spring, the area inundated depending on the size of the flood.

From Hit to Al Musayyib there is a single channel. Between these points, at Al Fallujah, the Euphrates approaches most closely to the Tigris, giving support from ancient times to a city of some sort, today Baghdad. A few miles below Al Musayyib the river divides into two branches, the more easterly, the Shatt al Hillah, being a former main channel and the more westerly, the Shatt al Hindiyah, carrying the present main stream. Below Al Kifl the Hindiyah branch splits in two and, after passing through marshes, re-forms just above Ash Shinafiyah. The Hillah and Hindiyah

branches reunite near Samawah, 110 mi. from their beginning, and continue as a single stream to An Nasiriyah. There the Euphrates divides into numerous channels and spreads out into marshy land and the Hawr al Hammar, a permanent lake at the eastern end of which it merges with the Tigris. From this point onward the combined rivers flow 111 mi. as the Shatt al Arab to the Persian gulf.

Considerable discussion has taken place concerning the growth of the delta of the Euphrates and Tigris. The usual earlier interpretation was that the large quantities of sediment carried by these rivers, with significant additions from the Karun, have gradually filled up the northern end of the Persian gulf depression. A great delta, the Iraqi lowland, about 400 mi. long and 100–150 mi. wide, has thus been built by deposition. It has been estimated that enough material is carried by the several rivers to cause the rate of growth of their common delta at the sea's edge to be as much as a mile and a half every 100 years. Consequently, many archaeologists and historians assumed that some of the ancient Mesopotamian capitals, such as Ur, now 150 mi. inland, were at one time seaports. Later geologic evidence, however, proved quite conclusively that the head of the Persian gulf, at least in historic times, was never very far from its present position and might indeed have been farther seaward. (See *History*, below.)

Regime of the Euphrates.—During the summer months there is almost no rainfall anywhere in the catchment basin of the Euphrates. The river is consequently at its lowest level at the end of this season, in September and October. With the coming of the winter rains, particularly to the upper section, the level of the river rises, reaching its maximum flood stage at the end of the season in April and May. In the winter minor flood peaks occur irregularly as a result of short periods of heavier than usual rain. The big spring flood is due to rain and the melting of the snow that has accumulated in the Turkish highlands. The spring flood is relatively predictable in terms of arrival time and height.

Although the rainy season in the Euphrates basin occurs regularly each winter, the actual amount of rainfall varies greatly from year to year. This is characteristic of arid and semiarid climates, the percentage variation from the normal becoming greater with increasing aridity. The flow of the Euphrates and its relation to water supply is thus affected by the yearly variation in rainfall, some years having an abnormally high rate of discharge and others a subnormally low rate. For example, in 1929 the maximum flood discharge rate was 4,700 cu.m. per second, while the maximum the following year was only 650. The same degree of variability also occurs in the minimum discharge rates.

Flood Control and Irrigation.—The irregularity in the seasonal and annual flow of the Euphrates has made the problem of controlling floods and providing adequate irrigation facilities especially difficult in Iraq. The objective in the development of any river system is the efficient utilization of the total resource, which consists principally of reducing the level at flood time and increasing it at low water. A number of improvements toward the accomplishment of this objective have been made.

Because of the near base-level gradient in its lower course the Euphrates is confined between artificial embankments to protect the adjacent farmlands from flooding at high water. The dikes, however, have no control over the volume of discharge.

To assist in reducing the flood hazard on the Euphrates and also to improve irrigation, a low barrage at Ramadi enables a portion of a flood to be diverted through the Warrar inlet channel into Lake Habbaniyah, a large natural basin close to the right bank of the river. Dikes around Lake Habbaniyah create a large reservoir. If necessary, overflow from Lake Habbaniyah may be absorbed by the still larger Abu Dibs depression south of the lake through the Mujarra escape regulator. Water diverted into Lake Habbaniyah may be returned to the Euphrates through the Dhibba regulator, which enters the river about 25 mi. downstream from the Ramadi barrage, for irrigation in late summer when the water level is low. The original Habbaniyah scheme was recommended by Sir William Willcocks in 1911, but wars and other interruptions delayed its final completion to 1956.

Between 1865 and 1890 the main channel of the Euphrates shifted from the Hillah to the Hindiyah branch, threatening the Hillah with complete desiccation. A weir was constructed at the head of the Hindiyah branch to raise the level of the water upstream, thereby ensuring the feeding of the Hillah. A new barrage at Hindiyah was built in 1911–13 as a part of a revised scheme for irrigation and flood control designed by Willcocks. The Hindiyah barrage gives perennial irrigation along both channels of the Euphrates through a system of take-off canals and permits the slightly lower, western Hindiyah branch to be used for flood diversion in times of emergency.

Between Ramadi and Hindiyah there are a number of perennial irrigation canals which depend on the natural levels of the river for their supply of water. They run at somewhat less than capacity during the low-water period in late summer. Their winter flow, on the other hand, can be controlled according to their capacity and the need. These canals follow roughly parallel courses between the Euphrates and the Tigris and irrigate some of the best land in the upper part of the "delta." Other irrigation canals and improved drainage facilities occur near Al Kufah and Shamiyah in the marshy region between the Hillah and Hindiyah branches.

The Euphrates is not navigable for river boats with a draft of more than three or four feet. Some traffic is carried in the lower section of the river by means of native craft of various types.

(D. D. Cr.)

History.—The earliest recorded form of the name Euphrates is the Sumerian Buranunu ("Great River"), found in documents of the 3rd millennium B.C. This became Purattu in Babylonian and Assyrian and Ufrat in Old Persian. The present Greco-Latin form is first found in Herodotus (*c.* 450 B.C.).

The theory, current until 1952, that the earliest settlements in southern Mesopotamia were established on the newly formed delta of the two rivers was disproved by later geologic evidence. Shells found in Holocene deposits northwest of the present shore line of the Persian gulf indicate a fresh-water fauna except in areas of local and temporary marine inundation, and the episodic subsidence of the Mesopotamian basin, a continuing effect of the mountain-building movements of western Iran, has counteracted the rise in land level caused by alluvial sediments. This subsidence is the probable cause of the local marine inundations and of the continued existence of marshy depressions which would, under static conditions, have been rapidly filled in by wind- and water-borne deposits.

The course of the Euphrates in the alluvial plain below Ramadi and the extent of the marshes have varied considerably in historic times. In the 3rd millennium B.C. the main stream ran east of its present course, passing by the Sumerian cities of Sippar, Shuruppak, Nippur, Larsa and Ur, whence it flowed through marshes to the gulf. An eastern branch, Iturungal, watered Adab and Umma before rejoining the Euphrates near Larsa, and a third channel, Sirara, served Lagash and continued through marshes to the sea.

Total dependence on the Euphrates soon proved unsatisfactory. About 2400 B.C. En-Temena dug a canal from the Tigris to Lagash, and about 1850 B.C. Sin-Iddinam extended it to his city Larsa on the Euphrates. A century later Hammurabi ordered the cleansing of the Euphrates channel from Larsa to Ur, and farther north he brought the river back to Sippar, from which it had receded. This operation was repeated by Nabopolassar (625–605 B.C.), and at Babylon Nergal-shar-usur (559–556) corrected a westward movement which had taken place since the time of his predecessor Nebuchadrezzar. It is not known when the river finally abandoned its ancient course. Greek and Roman geographers had little detailed knowledge of the area, though Pliny states that by the 1st century A.D. the Euphrates and Tigris had long had a common mouth. Catastrophic floods in the 6th and early 7th centuries created the Great Swamp, which covered most of the country southeast of Al Kufah until the 16th century. The Euphrates entered the swamp by two channels, the main stream by Kufah, the other by the present Hillah branch; it received the Tigris waters by the Shatt al Hai channel, and the combined rivers debouched into the Shatt al Arab. The largest surviving remnant of the Great Swamp is the Hawr al Hammar marsh northwest of

Basra. The Hillah branch later became the main stream, but after 1800 it progressively lost its waters to the Hindiyah, approximately the medieval Kufah channel. Many attempts to establish equilibrium culminated in the building of the Hindiyah barrage.

The Euphrates made possible the great civilizations of southern Mesopotamia from Sumerian to Abbasid times, and its valley provided a highway of trade and invasion linking them with Syria and the Mediterranean. Sumerian influence extended to Tell Brak on the Khabur, and the first Semitic ruler in the south, Sargon of Akkad (about 2350–2300 B.C.), campaigned in Syria and Cilicia. Mari, near Abu Kemal, was a trading city of importance throughout the 3rd millennium, and in the time of Hammurabi Babylonian caravans spread the use of their cuneiform script through Syria and Anatolia. At the beginning of the 1st millennium B.C. the valley was divided between the Babylonians in the south, the Aramaeans on the middle Euphrates and the Hittites in the north, with their fortress city at the important crossing of Carchemish. Aramaean trade brought about the adoption of the first alphabetic script, derived by them from the Phoenicians. Their territory formed part of the Late Assyrian empire, and detailed knowledge of its geography comes largely from Assyrian royal annals. Later evidence is furnished by Xenophon (401 B.C.) and, after the Syrian Euphrates became the frontier between Rome and Parthia in 64 B.C., by Greek and Roman geographers and military historians, notably Strabo, Isidorus of Charax, Ptolemy and Ammianus Marcellinus, and by documents such as the Tabula Peutingeriana. The frontier city of Dura-Europos yielded archaeological evidence of life in the valley under Roman and Parthian dominion.

Prosperity continued, according to the numerous accounts of Arab geographers, as late as the 14th century A.D., and Raqqa was the favourite residence of the Abbasid caliph Harun al-Rashid. Thereafter European trade was diverted by other routes, and Arab caravans usually took the desert road to avoid the lawless Euphrates tribes, although from the 16th century western travelers often preferred the short overland route to the east and from 1573 to 1603 many went by boat from Birecik to Al Fallujah or even Basra. L. Rauwolff (1573) gives the earliest European account of this stretch of the river. The first modern survey was undertaken in 1835 by the Euphrates expedition under Col. F. R. Chesney to explore the possibility of steamship traffic, for which the river proved unsuitable. The project of a Euphrates valley railway, also conceived by Chesney, was abandoned in 1870 and after 1886 the route fell rapidly into disuse. The best modern account is given by Alois Musil, who traveled on the middle Euphrates in 1912 and 1915. (E. E. D. M. O.)

BIBLIOGRAPHY.—Sir William Willcocks, *The Irrigation of Mesopotamia*, 2nd ed. (1917); M. G. Ionides, *The Regime of the Rivers Euphrates and Tigris* (1937); F. F. Haigh, *The Control of the Rivers of Iraq and the Utilization of Their Waters* (1951); G. M. Lees and N. L. Falcon, "The Geographical History of the Mesopotamian Plains," *Geogr. J.*, vol. cxviii (1952); S. Lloyd, *Twin Rivers* (1947); G. Le Strange, *Lands of the Eastern Caliphate* (1905); S. H. Longrigg, *Four Centuries of Modern Iraq* (1925); A. Musil, *The Middle Euphrates* (1927). (E. E. D. M. O.; D. D. Cr.)

EUPHRONIUS, a distinguished Athenian potter and vase painter of the late 6th and early 5th centuries B.C. Several vases with his signature are preserved. The name of the artist who decorated the vases which Euphronius signed as potter is not known, but he has been named the Panaitios painter from the inscriptions "fair Panaitios" which occur on his works.

Among the best-known vases decorated by Euphronius are the crater with Hercules and Antaeus (Louvre) and the Aphrodite cup (British Museum).

EUPHUISM is the name given to an excessively precious prose style made popular in England by the publication of John Lyly's *Euphues, the Anatomy of Wit* (1578). So popular did the style become, especially after *Euphues and His England* (1580), that Edward Blount in his preface to the *Sixe Court Comedies* of Lyly (1632) could declare without much hyperbole: "Our Nation are in his debt for a new English which he taught . . . All our Ladies were then his Schollers; And that Beautie in Court, which could not Parley, *Euphueisme*, was as little regarded; as shee which now there, speakes not French." Euphuism was fashionable, however, only during the 1580s; in the next decade it became the object of ridicule and contempt. Lyly himself soon stripped the style of its worst extravagances in his own comedies, and it maintained its popularity primarily because it was taken up and imitated by such Elizabethan novelists as Robert Greene and Thomas Lodge.

The word "Euphues" (Gr. *Euphues*) Lyly discovered in Roger Ascham's *The Scholemaster* (1570), where it is defined as "he, that is apte by goodnes of witte, and appliable by readiness of will, to learning . . ." But Lyly's adoption of the word as the name of his hero is ironic, and it can be argued that Lyly's euphuistic style was, on one level at least, developed to further the irony. For euphuism was not invented by Lyly. Its essential features had appeared in such works as George Pettie's *A Petite Pallace of Pettie his pleasure* (1576) and in certain sermon literature and Latin tracts. Nevertheless, it was Lyly who perfected the distinctive rhetorical devices on which the style was based, and thereby demonstrated that English could be written with grace and verve and clarity—qualities previously only nascent in English prose literature.

The euphuistic sentence is generally constructed upon the principles of balance and antithesis, upon syntactical schemata common in medieval Latin writing and ultimately traceable to the classical rhetoricians, Gorgias and Isocrates. The three schemata, that typify Lyly's euphuism are: (1) isocolon ("like length"—phrases or clauses of equal length, or nearly so, which appear in succession, balanced against each other); (2) parison ("like form"—the precise balancing against each other of key verbal elements in successive clauses and phrases); and (3) paromoion ("like sound"—the correspondence of sounds between words or syllables, especially between words already balanced against each other in successive clauses and phrases). Examples from *Euphues* are: of isocolon, "What hope can you haue to obtayne my loue, seeing yet I coulde neuer affoord you a good looke?"; of parison, "But certes I will eyther leade a Uirgins lyfe in earth (though I leade Apes in Hell) or else follow thee rather then thy giftes: yet am I neither so precise to refuse thy proffer, neither so peeuish to disdain thy good will." This sentence also illustrates two forms of paromoion which Lyly found irresistible: (1) alliteration; and (2) word likeness or repetition.

Theoretically, nothing can be said against the use of such syntactical and verbal devices. All good writers have relied upon them for point and precision, for clarity and accuracy of thought. But what distinguishes Lyly's euphuism is his almost complete reliance upon them in *Euphues* (less so in *Euphues and His England*) and his initial failure to perceive that an endless string of balanced and antithetical sentences can only result in a monotony of cadence that inevitably erects a barrier between writer and reader. Sound tends to obliterate sense rather than enhance it. The reader's attention begins to wander, and when he becomes aware of Lyly's mannered dependence upon another device, the Lylian similitude, he begins to smile. To cite but one example: to prove that "euery thing will dispose it selfe according to Nature" Lyly exploits his fancy by asserting that "the stone *Abeston* being once made hotte will neuer be made colde" and so on to the limit of his ingenuity. These annoying comparisons Lyly either drew from the *exempla* of medieval sermons, from the bestiary and from the unnatural natural history of Pliny, or invented himself. It was his overuse of the device and of alliteration that led to castigation of his style by critics such as Sir Philip Sidney and Gabriel Harvey, and it was probably Harvey who coined the term "euphuism" as a synonym for bad style.

It was not Lyly's conscious aim to overwrite. In the Epistle Dedicatory to *Euphues* he disclaims all intention of writing a "fine" style. "Though the stile nothing delight the dayntie eare of the curious sifter, yet wil the matter recreate the minde of the courteous Reader. The varietie of the one wil abate the harshnes of the other . . . It is therefore me thinketh a greater show of a pregnant wit, then perfect wisedome, in a thing of sufficiēt excellencie, to vse superfluous eloquence . . . It is a world to see how English men desire to heare finer speach then the language will allow, to eate finer bread then is made of Wheat." But where in English prose before *Euphues* can sentences shaped so artfully be found?

Lyly's fatal Cleopatra consisted, not in his criteria, which make sense, but in his virtuosity, which overwhelmed his judgment. In the long run this was all for the best. Elizabethan writers, most of whom studied Lyly, could learn most quickly from his very extremes how to organize sentences free from the rambling periods of 16th-century Ciceronianism. His vices they could ignore; his virtues they could emulate—and they did. Lyly was, in fact, the first genuine prose stylist to leave an enduring impression upon the language. Together with Sir Francis Bacon (Senecan in style) and Sir Thomas Browne (in style Ciceronian at its controlled best) he prepared the way for the flexible prose instrument John Dryden developed soon after the Restoration.

BIBLIOGRAPHY.—F. Landmann, *Der Euphuismus* (1881); C. G. Child, "John Lyly and Euphuism," *Münchener Beiträge zur romanischen Philologie*, vol. 7 (1894); R. W. Bond, "*Euphues* and Euphuism" in Lyly's *Complete Works*, vol. 1, pp. 118–159 (1902); A. Feuillerat, *John Lyly, Contribution à l'Histoire de la Renaissance en Angleterre*, pp. 411–474 (1910); M. W. Croll, introduction to his and H. Clemons' edition of *Euphues* (1916); J. A. Barish, "The Prose Style of John Lyly," *A Journal of English Literary History*, vol. 23, pp. 14–35 (1956).

(W. N. KG.)

EUPOLIS (*c.* 446–411 B.C.), Athenian poet of the Old Comedy, is ranked by Horace, along with Cratinus and Aristophanes, as one of the greatest writers of his school. He was reputed to equal Aristophanes in the elegance of his style and Cratinus in his command of satire. Born at Athens *c.* 446 B.C. he grew up during the Peloponnesian War. He is said to have been drowned by Alcibiades, whom he had attacked in his *Baptai* (?416 B.C.), but it is more likely that he died fighting for his country. Although he was at first on good terms with Aristophanes, their relations became strained, and they accused each other of plagiarism. Of the 17 plays attributed to Eupolis, with which he won the first prize seven times, only fragments remain. Of these the best known are: the *Kolakes* ("Flatterers"), directed against the spendthrift Callias; and *Marikas*, an attack on Hyperbolus, the demagogue. The political *Demoi* ("Townships"), of which considerable fragments have been found in papyri and *Poleis* ("Cities") portrayed the disastrous effects of the Peloponnesian War.

See J. M. Edmonds, *Fragments of Attic Comedy*, with trans. (1957); G. Norwood, *Greek Comedy* (1931). (M. PR.)

EURE, a *département* of northern France, was formed in 1790 from part of the ancient province of Normandy. Pop. (1954) 332,514. Area 2,331 sq.mi. It lies astride the lower Seine south of the *département* of Seine-Maritime, and is bounded to the south by Eure-et-Loir and to the west by Orne and Calvados. Eure consists of rolling chalk country trenched by the valleys of the rivers Seine, Eure and its tributary the Iton, and the Risle. Le Vexin normand is the district east of the Seine, while that between the lower Seine and the Risle, which flows north to an independent mouth in the Seine estuary, is known as Roumois. The chalk is in part capped by fertile *limon* soil, but elsewhere clay with flints forms the surface and is less fertile and sometimes left as woodland. Wheat and sugar beet, with fodder crops in the rotation, are extensively grown, and in this moist, mild maritime *département* pasture is extensive and orchards, especially apples and pears, are very widespread. Much cider is made. The *département* is largely agricultural, but contains some outlying textile factories of the lower Seine industrial area, which focuses upon Rouen. There are no large towns, Evreux (*q.v.*), the capital and largest centre, had a population (1954) of 20,441.

Eure is the seat of a bishopric in the province of Rouen, and its cathedral is noteworthy for its fine stained glass and its wooden screens.

The *département* is divided into three *arrondissements* centred upon Evreux, Les Andelys and Bernay. The court of appeal is at Rouen, and the *département* comes under the educational division (*académie*) of Caen. (AR. E. S.)

EURE-ET-LOIR, an inland *département* of northern France, was formed in 1790 from portions of Normandy and Orléanais. Pop. (1954) 261,035. Area 2,293 sq.mi. It is bounded north by Eure, west by Orne and Sarthe, south by Loir-et-Cher, southeast by Loiret, and east by Seine-et-Oise, and is drained northward to the Seine by the Eure river and southward to the Loire by the Loir river. In the west is the hilly country of the districts of Perche (*q.v.*) and Thimerais, but a large part of the *département* is formed by the level limestone platform of Beauce (*q.v.*). The fertile *limon* soils of this flat tract have long been rich grainlands and are now hedgeless stretches of continuous arable land, yielding heavy crops of wheat and sugar beet and set with large, clustered villages. The more varied, hilly country in the west of the *département* is less continuously cultivated and cattle pastures are more extensive.

Perche was famous for its draft horses. The market towns have industries such as brewing, milling and sugar refining associated with the agricultural economy. Chartres (*q.v.*), the chief market centre of Beauce and capital of the *département*, is renowned for its magnificent cathedral, the crowning achievement of Gothic architecture. The *département* is divided into the four *arrondissements* of Chartres, Châteaudun, Dreux, and Nogent-le-Rotrou, and comes under the court of appeal and *académie* of Paris.

(AR. E. S.)

EUREKA, a city of California, U.S., and seat of Humboldt county, is located on landlocked Humboldt bay in the northwestern corner of the state, about 280 mi. N. by W. of San Francisco. Founded in 1850, it was incorporated and became the county seat in 1856. It is the leading transportation, industrial and commercial city in northern California and is the chief seaport between San Francisco bay and the Columbia river.

EUREKA SPRINGS, a city of northwestern Arkansas, U.S., in the heart of the Ozark mountains, one of Carroll county's two seats. Within the city limits are 63 springs, held in trust by the city for free public use. Basin spring, near the centre of the city, was first discovered and recognized for its therapeutic value by Alvah Jackson, pioneer hunter and physician, in 1854. A tent settlement sprang up around it and the city was named in 1879 and chartered in 1880.

Eureka Springs, sometimes called "the little Switzerland of America," is a health resort and arts, crafts and tourist centre. Its attractions include magnificent scenery; a mild climate; unique winding streets, terracelike one above another; over 50 mi. of rock retaining walls; bath houses, modern hotels, motels, apartments and cottages. Table Rock lake is nearby. Over 150,000 tourists visit the town each year and it is the site of the Ozark Folk festival, held annually in October. (S. T. K.)

EURE RIVER, in northern France, is a left-bank tributary of the Seine, which it enters above Rouen. It is 225 km. (140 mi.) long. Rising in Normandy in the Collines du Perche about 800 ft. above sea level, it first flows southeast, turns sharply north past Chartres, then skirts the western margin of the Tertiary rocks of the central part of the Paris basin. Its main tributaries, the Blaise, Avre and Iton, enter on the left. The early significance of the position of Chartres (*q.v.*) was that it lay in the northern part of the open *campagne* country of Beauce in a gap between the forests of Perche and those that surround Rambouillet west of Paris. Farther north, however, the *limon*-covered Tertiary limestone promontory of the Mantois district between the Seine and Eure valleys is rich arable country, and other rich grain-growing lands extend to the west on the chalk *plaines* on either side of the Iton valley. Louviers, situated in the Eure valley below the Iton confluence as the river approaches the textile-manufacturing district of Rouen, is an industrial outlier, with old, established woolen mills. Although generally the Eure basin is agricultural and rural, there are some accretions of industry to be found at Évreux and other market towns. Chartres attracts many tourists by its glorious Gothic cathedral, and Évreux is an important road and railway centre. (AR. E. S.)

EURHYTHMICS is the term applied generally to the use of harmonious bodily movement as a form of artistic expression and specifically to the Dalcroze system of musical education, which is founded on a rhythmical discipline of bodily movements. The present article is concerned only with the Dalcroze system, developed early in the 20th century by Emile Jaques-Dalcroze (*q.v.*). His experience as a teacher of harmony at the Geneva conservatory having convinced him that the conventional professional training of musicians was radically wrong, Jaques-Dalcroze attempted

to develop his students' musical faculties generally, as a basis for the later study of a specific instrument and to awaken a sense of the ultimate basis of music and all other arts "rhythm" by translating sound into physical movement.

About 1905 Jaques–Dalcroze worked out a system of pedagogic eurhythmics, applying his method to children as part of their general elementary education; it has been adopted in schools throughout Europe and America. In 1910 a school for instruction in the Dalcroze method was established in Hellerau, Germany; subsequently it was moved to Laxenburg, near Vienna. The method is also taught at the Dalcroze School of Music in New York city.

The Dalcroze method consists of three parts: (1) rhythmic movement; (2) ear training; (3) improvisation (practical harmony). Of these, the first is the essence of the method; the second is closely linked with the first and is necessary to the pupil; the third is necessary to the teacher.

In the system of exercises upon which the method is based, time is shown by movements of the arms, and time-duration, *i.e.*, note values, by movements of the feet and body. In the early stages of the training this principle is clearly observed. Later it may be varied in many ingenious ways. For instance, in what is known as plastic counterpoint, the actual notes may be represented by movements of the arms, while the counterpoint in quarter, eighth or sixteenth notes is given by the feet.

In the series of movements to represent note values, the quarter note is taken as the unit and is represented by a step; higher values from the half note to the note of twelve beats are represented by a step and movements without progression; a note of twelve beats by a step and eleven movements. Thus, for each note in music there is one step or progression in space. Notes of shorter duration, quavers, triplets, etc., are also expressed by simple steps.

The whole training aims at developing the power of concentration and rapid physical reaction. For example, the teacher plays one or two bars, which the student then executes while the music plays the next rhythm; thus the student listens to a new rhythm pattern while executing one already heard, an exercise requiring and at the same time developing concentration. (AN. S.)

EURIPIDES (*c.* 485–407 B.C.) was the youngest of the three great Greek tragedians, following Aeschylus and Sophocles. It is unlikely that either of the traditional dates of his birth, 485 and 480 B.C., is founded on a genuine record, but neither is implausible since he entered his plays for the first time at the Dionysiac festival in 454. He died in Macedonia in the winter of 407–406, and the news reached Athens in time for Sophocles to dress his chorus in mourning at the Proagon of the Dionysia in the spring of 406. The following statements about his life can be made with some confidence. He was the son of Mnesarchus, or Mnesarchides, and Cleito, whose home was at Phlya, east of Hymettus. There was a property also on Salamis where Euripides may have been born, and where he was accustomed to sit in a cave looking out to sea, a habit considered eccentric in an age when solitude was easily achieved and little valued. It is to be inferred that he was well off, since Aristotle mentions that he was challenged to an exchange of properties (*antidosis*) and since he possessed a library. He took no part in public affairs except to write an epitaph for the Athenians who fell at Syracuse, though it is possible that he once went as ambassador to that city. He wrote an ode to celebrate the Olympic victory of Alcibiades in the chariot race, probably in 416 B.C. In 408 he left Athens to go and live at the court of Archelaus, king of Macedonia, who was a noted patron of the arts, and it is likely that he accepted the invitation in spite of his advanced age because unpopularity made life at Athens disagreeable. The fewness of his victories at the Dionysiac festivals, four in his lifetime and one posthumous, suggests some lack of appreciation.

Statements to be received with more suspicion are that he was prosecuted for impiety by Cleon, though this is by no means impossible; that he was twice married and much afflicted with domestic unhappiness; that before taking up tragedy he tried athletics and painting; and that he met his end as the victim of the hounds of his royal patron. While there is nothing unlikely in the tearing to pieces of an old man by Greek dogs, it is hardly credible that it should not have been mentioned in Aristophanes' *Frogs* produced shortly afterward, as Aristophanes took a lively interest in Euripides and he figures largely in this play. The story that his mother was of low birth is hardly compatible with the possession of a family cult of Apollo Zosterius.

The general impression left by tradition may be more reliable than the details. It is of a figure austere and unconvivial. He was passionately interested in ideas—and never were ideas more plentiful than in the Athens of his time—but acquaintance with them brought restlessness rather than conviction. A later age loved to call him the philosopher of the stage, but no philosophic school could claim him as a follower. It is unlikely that he was in any sense a disciple of Anaxagoras, and of the eldest sophists he was the contemporary rather than the pupil. But it is likely enough that he was deeply influenced by them, and it may be true that Protagoras read his agnostic treatise *On the Gods* at Euripides' house. The Greek comic poets associated him also with Socrates, who was alleged to have helped him with his plays.

WORKS

Nineteen plays of Euripides have survived as compared with seven each from Aeschylus and Sophocles. Owing to the greater popularity of Euripides in later antiquity ten of his plays were selected for reading in schools, and that nine others have survived is due to a lucky accident. Their quasi-alphabetical order suggests that they are a sequence from the Alexandrian edition, and therefore a random collection of plays, a point to be remembered if they are compared with the selected plays of the other two dramatists.

Euripides was said to have written 92 plays in all, and this statement can be reconciled with another, that he competed at the festival of Dionysus on 22 occasions, which would imply 88 plays, if it is assumed that the discrepancy is accounted for by the inclusion in the larger total of four plays whose authenticity was doubted, the *Rhesus*, *Peirithous**, *Rhadamanthys** and *Tennes**, the last three of which were attributed by some to Critias. But the question is less simple than it might appear because the *Archelaus**, written in honour of his Macedonian host, was not performed at Athens, and the *Andromache* did not appear in the records under his name.

A number of plays were lost between the end of the 5th century and the publication at Alexandria by Aristophanes of Byzantium (*c.* 257–180 B.C.) of the complete edition of all plays extant in his time. The plays in this edition appear to have numbered between 76 and 78. All but one of the titles are known, and in many cases information is available which, with the help of fragments from the plays themselves, enables us to form some idea of their nature.

Chronology of the Plays.—The dates of production of nine of Euripides' plays are reliably known on evidence which goes back, in the last resort, to the official Athenian record, and as several others can be placed within a few years on internal evidence, it is possible to establish a fairly complete series and to deduce from this a remarkably regular pattern of development of stylistic features and of dramatic technique into which the remaining plays can be fitted. The most important factor, because the most easily expressed statistically, is the occurrence of resolved feet in iambic verse (*i.e.*, feet of three syllables: tribrachs, dactyls or anapaests in place of the normal two-syllable iambi and spondees). On the basis of such data as these a table (below) of the extant plays can be constructed with a probable error of only a few years.

There is no real evidence as to the date of the one remaining play, the *Cyclops*. The cynical views of the Cyclops on moral standards perhaps suggest a not very early date. If the extant *Rhesus* is the play which Euripides wrote, it is an early work. However, the authenticity of the play was challenged in antiquity, it is not known on what grounds. It is a quite spirited dramatization of the night encounter related by Homer in the tenth book of the *Iliad*. It is not inferior to some certainly authentic plays,

*An asterisk denotes a lost play.

Date	Play	Resolved feet per 100 lines
454	*Peliades** (first production)	
442	(First victory)	
?	(*Rhesus*)	8.1
438	*Alcestis, Cressae**, *Telephus**, *Alcmaeon in Psophis**	6.2
431	*Medea*, *Philoctetes**, *Dictys**, *Theristae**	0.6
430	*Heracleidae*	5.7
428	*Hippolytus*	4.3
427	*Andromache*	11.3
424	*Hecuba*	12.7
422	*Supplices*	13.6
418	*Heracles*	21.5
415	*Troades*, *Alexander**, *Palamedes**, *Sisyphus**	21.2
413	*Electra*	16.9
412	*Helen*, *Andromeda**	27.5
411	*Iphigeneia in Tauris*	27.4
410	*Ion*	25.8
411–409	*Phoenissae*, *Oenomaus**, *Chrysippus**	25.8
408	*Orestes*	39.4
405	*Bacchae*, *Iphigeneia in Aulis*, *Alcmaeon in Corinth** (posthumous production)	37.6, 32.7

Note: Dates in *italics* are established. *Lost play.

and there is nothing positively un-Euripidean in style or technique. The main reason for doubting his authorship is that it shows no trace of the intellectual restlessness and love of generalizing which are lacking nowhere else in his work.

THE EXTANT PLAYS

General Characteristics.—The most important single fact about Greek tragedy is that its subject was, with rare exceptions, myth, and the basic difference between the three great tragic poets is to be found in their attitudes to the myth. In myth is included all that the ordinary Greek believed about the gods and about the men of the heroic age, all that had happened between the first beginnings of the universe and those quasi-historical events which were related in epic poems on Thebes and Troy. For Aeschylus myth was that through which reality might be comprehended, for Sophocles it revealed man's potentialities at their highest, but for Euripides it was a mere collection of stories without any particular authority, or, rather, with a false authority used to support primitive and pernicious conceptions. Nonetheless he had perforce to use the myth as the raw material for his plays. His younger contemporary Agathon once produced a play with an invented plot, and Aeschylus and his older contemporary Phrynichus had each written on the war with Persia which was the great event of their time, but in general there seems to have been no possibility of breaking away from the traditional material. All that Euripides could do was to use it in an untraditional way, often displaying the falsity of the values which underlay the natural interpretation of his story. Sometimes he would turn it upside down by transporting it into the contemporary world, as when he made nonsense of the old story of Orestes' vengeance by setting it in a world where courts of law already existed. Or, if he used a story that dealt with forces undeniably real, Aphrodite or Dionysus, he showed the moral inadequacy of a personal conception of such powers. These methods produced interesting and stimulating plays but they were marred by a fundamental disharmony. A writer who is quarreling with his subject matter is not likely to produce perfect works of art. Euripides wrote many plays with great and striking scenes. He can awaken profound pity and display brilliant psychological insight, but he has few memorable characters and his plays rarely satisfy as a whole.

The surviving plays of Euripides form a less homogeneous body of material than those of Aeschylus and Sophocles. It is partly that more of his plays have been preserved, partly that they are to some extent a random collection; but probably he wrote more kinds of plays than they did. It is not possible to divide them into groups in a wholly satisfactory manner but a number of them do fall into definite categories. Further, it is possible to assign some of the lost plays to these same categories and thus to get a more complete picture of the whole body of his work. The great popularity of Euripides in the centuries after his death led to an abundance of quotations in the works of later authors and of extracts in anthologies, and for the same reason he is, along with Homer, the poet most fully represented in the papyri (*i.e.*, the works of ancient writers, discovered mainly in Egypt, written on papyrus). Accordingly it is convenient to consider his works not strictly in chronological order but by the categories to which they can reasonably be assigned.

Plays of Psychology and Passion.—The *Medea* is one of the few Greek plays in which the action is largely determined by the character of the protagonist. Medea was wronged by the man who owed her his glory and his life, who was bound to her by marriage and by the bond of oaths yet more solemn. When Jason cast her off to marry a Greek princess she sought, as many women would, revenge. Others might have tried to kill their rivals and perhaps their husbands. Only a Medea punishes her husband by sparing him and leaving him to grow old without wife or child. The passionate sense of injury and the dread of her enemies' scorn which cause Medea after a dreadful struggle with herself to kill her own children are almost incredible. Yet Euripides with his strange gift for bringing sympathy to those who seem beyond the pale makes Medea credible and at times even allows her to speak for women as a Greek woman. In spite of a romantic and legendary background the play is set in contemporary Greece. Only in the final scene where, as often, there is a subtle change in the assumptions, does the action enter the world of the supernatural and Medea finds herself possessed of a dragon-chariot and a dramatically convenient knowledge of the future.

Like the *Medea*, the *Hippolytus* has for its theme the struggle of the human will to master human passion, but here the position is complicated by the presence of gods who both symbolize real forces and behave like irresponsible individuals. Aphrodite in anger at the haughty indifference of Hippolytus to the call of sex causes his stepmother, Phaedra, to fall in love with him, knowing that both will thereby be destroyed. Phaedra resists her passion and tries to starve herself to death, but her secret is betrayed to Hippolytus. Phaedra, ready to give her life rather than succumb to temptation, is so mortified at her denunciation by the boy who does not know what temptation is that she makes against Hippolytus a charge of having tried to violate her and hangs herself in a last effort to save her own and her children's honour. Her husband not unnaturally believes the charge, especially as Hippolytus keeps his oath not to reveal what he has been told under pledge of secrecy. His father's curse, backed by the promise of Poseidon, leads almost at once to the death of Hippolytus, dragged and mangled by his runaway horses; but before he dies his innocence is revealed by Artemis, the divine expression of the purity that is the opposite of the sexual drive, Aphrodite. Artemis reveals that Phaedra too was Aphrodite's victim, and Hippolytus before he dies forgives his father, freeing him from the dreadful stain of bloodguilt. The graciousness of Artemis gives the close of the play a peacefulness in contrast to the white heat of hatred at the end of the *Medea*. Yet she consoles the dying Hippolytus with the promise that she will make away with Aphrodite's favourite mortal, and leaves one wondering of what sort are these petulant deities who cannot be accepted as gods but at the same time represent something which is more than human. Considering the nature of the plot the play is of remarkable delicacy; Phaedra speaks no word to Hippolytus throughout, a great contrast to the poet's earlier play of the same name, now lost, in which Phaedra both solicited Hippolytus and defended her conduct to Theseus. The *Hippolytus* is the only surviving play of those produced in the poet's lifetime that is known to have won a first prize.

The *Hecuba* is a simpler play without internal conflicts. It shows the double disaster that reduced the Trojan queen by sheer weight of hatred and misery to a merely animal ferocity indicated by the closing prophecy of her transformation into a bitch. She loses first her daughter Polyxena, led off to be sacrificed at the tomb of Achilles, then discovers during the preparations for Polyxena's funeral the corpse of her son Polydorus, who had been murdered by his Thracian host Polymestor. Since Polymestor is a Greek ally her appeal to Agamemnon for justice seems likely to fare no better than her earlier supplication of Odysseus to help her save Polyxena, but a timely reminder that Hecuba is the mother of his favourite concubine elicits from Agamemnon what he would not concede to a prayer for mere justice, the permission to punish Polymestor if she can achieve it by her own devices.

Polymestor is enticed into the Trojan quarters where the women blind him and murder his two sons who have come with him. Such is the power of misery to deprave.

The *Andromache* turns on the jealousy felt by Hermione for her husband Neoptolemus' concubine, the Trojan Andromache, who is the mother of his only son. After an exciting beginning marked by strong anti-Spartan feeling most of the original characters disappear and the interest is dissipated. There is little subtlety in the psychology, but the same erotic element is noticeable which seems to have been a feature of other plays of the period, *Phoenix**, *Stheneboea** and the first *Hippolytus**. The introduction of this theme into tragedy aroused much indignation.

The *Alcestis* may be mentioned here, not as a play of passion but because Admetus, as a study in unconscious selfishness, has something in common with Jason in the *Medea*. The *Alcestis* took the place of the burlesque satyr-play that normally followed the three tragedies. Its tone is lighter and the drunken Heracles is a traditional satyr-play figure. No play is more easily misunderstood by the modern reader. Admetus is allowed a second life if he can find someone willing to die in his place; but the obvious question, whether he ought to accept such a sacrifice, is not really raised. That a man's life is more valuable than a woman's is a basic assumption throughout. The folly of Admetus is that he imagines himself more fortunate than he really is. But after his wife's death he realizes his mistake, and when Heracles wins her back from death and restores her to him he has perhaps learned something. Alcestis' own emotion seems directed more toward her children than her husband. Though it is a happy play, the wrangle between Admetus and his father Pheres, who, having refused to die for his son, is turned away from the funeral, is full of the slightly sour comedy with which Euripides often diversified his tragedy.

Along with the *Alcestis* was produced the *Telephus*, which, judging from the numerous references in Greek comedy, marked a stage in the development of Euripides' art. The hero disguised as a beggar and perhaps reduced to beggary was dressed in rags more realistic than any previously seen on the stage, a sight not compatible with the standard of tragic dignity commonly accepted.

The "Bacchae" and the "Heracles."—Two of the most remarkable plays stand rather in isolation, the *Bacchae* and the *Heracles*. The former has a link with the *Hippolytus* in that Dionysus, like Aphrodite, is a real force that is also an unsatisfactory personal god, while the latter seems to raise the whole question of the government of the world by the sort of god that the myth presents.

The worship of Dionysus, at least in its most vital form of which Euripides may have seen something in Macedonia, was alien to the spirit of normal Greek religion. The congregation of the faithful, consisting mainly of women, raised itself to a state of hysteria by music and ritual and at the culminating moment tore in pieces and devoured hot and bleeding the animal in which their god was temporarily incarnate. The ecstasy of union with the god and the delight of escape from self must be balanced against the dangerous irresponsibility which goes with the loss of self-consciousness. Euripides in his marvelous choral odes and in his description of the reveling bacchanals on Cithaeron does full justice to the beauty and the lure of this type of religious activity. His expression of the relief from the weary burden of reason which is brought by abandoning self to faith has been taken by many to mean that the old rationalist had himself surrendered at last, as many of his countrymen were surrendering, to other foreign faiths that were among the imports of the Piraeus. But the words of Dionysus' victim "gods should not be of like temper with men," the same reproof as is offered to Aphrodite in the *Hippolytus*, hardly suggests an acceptance of this tempestuous faith.

The story of the *Bacchae* probably represents a genuine folk memory of the coming of Dionysus and the resistance that the new faith encountered from those who were very naturally disturbed by its social consequences. The ruthless vengeance of Dionysus on those who refused to recognize him, though not incompatible with the Greek conception of a god, is not so described as to rally the audience to the side of Dionysus.

The *Heracles* is the most baffling of the plays. Euripides has taken great liberties with the traditional story, according to which Heracles killed his wife and children in a fit of madness and undertook the 12 labours to obtain purification. In the play he undertakes them as a dutiful son to win for his father Amphitryon his return to his native Argos, from which he has been exiled for manslaughter. While he is away in Hades performing his last and most perilous labour, there has been a revolution in Thebes, and when the play opens his wife and children are suppliants at the altar in peril of their lives. Heracles returns in the nick of time and destroys the usurper Lycus, comforting his frightened children with almost sentimental tenderness. All seems well and the chorus celebrates the justice of Zeus, who has not abandoned Heracles of whom, with Amphitryon, he is the joint father. Suddenly Iris, and Lyssa, the goddess of madness, appear and deliver what is virtually a new prologue which begins a new action. While Heracles was still performing his labours he was shielded by Zeus and Destiny. Now that they are finished Hera is free to vent her spite on the fruit of her husband's illicit love. She will cause him to kill wife and children in madness. Heracles does so, and is revealed surrounded by their corpses. When he recovers consciousness his first thought is to kill himself, but he is restrained by Theseus, whom he has rescued from Hades. Theseus offers the conventional consolation: the gods have their own troubles to endure and man must put up with his, to which Heracles replies, almost in the words of Xenophanes, that he does not believe miserable tales of poets about gods who commit adultery and put each other in bonds, thus denying the basic assumption about his own birth. He lives because it is cowardly to die, and departs to the refuge that Theseus offers him at Athens. The assumption that the cause of his troubles is Hera's jealousy is maintained except for this one passage. It is natural to expect a symbolic meaning for Hera comparable with those that attached to Aphrodite and Dionysus, but if Hera had a meaning of this kind, it is lost. If Hera has no symbolic meaning, the relation of the two halves of the play is that of mere juxtaposition. Heracles' misfortune follows hard upon the apparent end of his troubles; it happens after but there is no causal connection. Chance would appear to be supreme.

The Patriotic Plays.—Any list of the legendary glories of Athens would contain the stories of the rescue of the children of Heracles from the persecution of Eurystheus and of the enforcement of the right of the Argive assailants of Thebes to the burial that the Thebans barbarously denied them, both examples of disinterested intervention on the side of right. The first of these stories was the theme of the *Heracleidae* produced near the beginning of the Peloponnesian War, a simple glorification of Athens memorable chiefly for the self-sacrifice of Macaria and the miraculous rejuvenation of Iolaus. The play is unusually short and perhaps has not been preserved complete. The *Supplices* dramatizes the second of these stories. It may have been inspired partly by the refusal of the Boeotians to hand over the Athenian dead after the battle of Delium. It is a weightier play than the *Heracleidae* and full of political debate. The patriotism is less unambiguous. The speech of Adrastus over his fallen comrades, which can hardly have failed to suggest the wartime funeral orations of Athens, contains a eulogy of the blackguardly Capaneus that may hint at the insincerity customary on such occasions. A further surprise is the suicide of Evadne, the wife of Capaneus, who despite her father's entreaties throws herself on her husband's blazing pyre. This startling episode may have been some compensation for the lack of drama, but it does not contribute to the harmony of the play considered as a whole.

The *Erechtheus**, probably of the same period, was also a glorification of patriotic sacrifice, the sacrifice by Praxithea of her daughter as the price of victory in the war against Eleusis.

The Trojan Women can be grouped with these plays as being in a sense political propaganda, though intended not to spur the war effort but to discourage war. It was produced in the spring before the Athenians set out on their disastrous expedition against Syracuse, and it is possible that the mood of the play is influenced by the capture of Melos in the previous autumn. As at Melos, so at Troy, the men were put to death, the women sent into slavery.

The play is little more than a series of episodes that reveal the sufferings of the conquered. First Cassandra is carried off to be the concubine of Agamemnon, then Andromache to be the slave of Neoptolemus, the son of her husband's slayer and the slayer of her husband's father Priam. Her child and only comfort Astyanax is taken from her and hurled from the wall of Troy so that he may not grow up to be a trouble to the Greeks. But the glory of the Greek victory is not only brutal, it is empty. For in an artificial yet strangely effective prologue Poseidon the friend of Troy, and Athena, who is angry because the Greeks have desecrated her temple during the sack of the city, agree together to destroy the Greek fleet on its homeward voyage. "Foolish is the man," says Poseidon, "who sacks cities and gives over to destruction temples and tombs which are sacred to the dead, only to perish himself." While the Greeks dispose of their victims they are seen as future victims themselves.

The "Electra."—Though the *Electra* has no political purpose it has in common with the *Trojan Women* a savage rejection of the values of ancient myth. Euripides' treatment of the famous story of Orestes and the murder of a mother to avenge a father is very revealing. Electra has been married to an Argive peasant so that her children may be of low birth and no danger to the usurpers. She lives in a cabin on the frontier of Argos nursing her hatred of those who have deprived her of her birthright, and dreaming of her hero brother Orestes on whose return her hopes depend. He comes, in disguise, a rather ordinary young man who keeps a prudent eye on his line of retreat to safety. He learns what manner of woman his sister is; she does not need the sanction of Apollo's oracle before wanting to kill her mother. Rather to his dismay he is recognized by an old servant; there is no chance of retreat now. A plan is laid. Aegisthus is about to sacrifice to the nymphs at a neighbouring shrine. If Orestes times his passing properly he will be invited to join the celebration and have a chance of striking down his stepfather. Electra gleefully undertakes to deal with her mother. She will pretend that she has had a child; her mother will come to help with the tenth day rites, and once inside the hut can be despatched. The plans work; Aegisthus is struck down as he peers at the victim's entrails. At the sight of his mother, however, Orestes falters and doubts the validity of Apollo's oracle, but he is scornfully overborne by his sister's stronger will. Clytemnestra is made as sympathetic as her record will allow; she is a broken woman anxious to go more than halfway where there is any hope of reconciliation, but Electra only insults her before leading her in to her death. Yet the horror of the actual murder shatters Orestes and Electra as well. The play ends with the appearance of the Dioscuri, the divine brothers of Helen and Clytemnestra. They agree as openly as decorum allows with Orestes' doubts of Apollo's command that a son should murder his mother. This is an effective and well-made play, but all the characters except Electra's peasant husband are necessarily disagreeable and arouse little sympathy.

The Romantic Plays.—Euripides wrote, especially in his later years, a number of plays with highly organized structure leading up to a recognition scene in which the discovery of the identity of a character produced a complete change in the situation (*peripeteia*). The *Cresphontes** was a famous, and perhaps the earliest, instance. Here Merope was on the point of killing a sleeping man whom she believed to be her son's murderer, when it was revealed, while her weapon was already poised to strike, that the man was no murderer but her long-lost son. The *Iphigeneia* is a play of the same type. Iphigeneia after her supposed sacrifice was carried by Artemis to the Crimea and made priestess in her temple. Here it was the practice to offer strangers as sacrifices to Artemis. Orestes was told by the oracle to remove the statue of Artemis to Greece, but he was captured and handed over to his sister to be sacrificed. A cleverly contrived recognition took place and Iphigeneia, declaring that the statue needed purification owing to the presence of a matricide, took it down to the sea. There they would have made good their escape, but a storm drove their ship back, giving occasion for Athena to appear and announce the establishment of a new cult of Artemis in Attica.

The *Iphigeneia* is exciting; much of the *Helen* is frankly comic as well. Euripides adopts the version of the myth according to which Helen never went to Troy but remained in Egypt while Greeks and Trojans fought for a phantasm. Menelaus and the supposed Helen are wrecked on the coast of Egypt on their way home from Troy. Menelaus meets the true Helen but declines to believe it is she until the supposed Helen has evaporated. United at last, they make good their escape by inventing a Greek custom that those who are lost at sea must also be mourned at sea; Helen is thus able to borrow a ship from the Egyptian king in which to perform the funeral rites for the supposedly drowned Menelaus, and they return to Greece while the Dioscuri reveal the divine will to the baffled king.

The *Ion* is the most brilliant and the most complex of these plays, since it contains a false as well as a true recognition. Ion was the child of an Athenian princess, Creusa, and of Apollo. Creusa bore the child in secret and left it in a cave under the Acropolis. Hermes, who speaks the prologue, tells how, to oblige Apollo, he took the babe to Delphi, where his priestess found and reared it. Ion is now a temple slave. Creusa supposed her child to have been the prey of wild beasts. Later she was married to a foreigner, Xuthus, who had helped the Athenians in war. They had no children and now visit Delphi to ask the god what they are to do. Creusa has a secret hope that her own child may be restored to her. Hermes explains that it is Apollo's intention to give Ion to Xuthus as his own child and to reveal the truth to Creusa only after they have all returned to Athens; thus scandal may be avoided. At Delphi mother and son meet and come to the brink of recognition. Xuthus is duly told that Ion is his son but Apollo's plan miscarries. Creusa in the bitterness of her disappointment, and in her anger that a stranger, as she supposes, should be foisted into the royal house of Athens, denounces Apollo in his own holy city as a ravisher and tries to poison the usurper who is really her son. The plan is thwarted by a miraculous mischance, and Ion is on the point of dragging his mother from Apollo's altar to which she has fled for sanctuary when the priestess appears with the cradle in which she had long ago found Ion. Now that he is going out into the world it is time for him to be given the evidence of his birth and the tokens by which he may recognize his unknown mother. Creusa knows the cradle as the one in which she left her child and knows the tokens it contained. The false recognition of the father is followed by the recognition of the mother. Apollo does not care to appear himself but sends Athena to clear up the mess and to foretell the origin from Ion of the Ionian race and its four tribes, a matter of some patriotic interest to the Athenians.

A remarkable number of lost plays led up in the same way to a recognition between parent and child. In addition to the *Cresphontes**, and the *Alexander**, which contained the recognition of Paris by his mother Hecuba, the *Auge**, *Alope**, *Melanippe Desmotis**, *Antiope**, *Alcmaeon at Corinth** and *Hypsipyle**, of which a considerable portion has been preserved on papyrus, all seem to have had a strong family likeness. It is no accident that a recognition is the theatrical device to which Aristotle in the *Poetics* devotes more space than to any other.

Even in antiquity it was realized that Euripides developed tragedy in the direction of New Comedy. No play goes further than the *Ion*. The recognition of a child by the tokens with which it was exposed at birth, and the begetting of a child, as Xuthus supposed that Ion was begotten, at a festival of Dionysus, are both recurrent motives in New Comedy. So also is the theme of young lovers happily united, which was presented, probably for the first time, by Euripides in the *Andromeda**. His interest in love as a personal experience, both in its lighter and more sombre aspects, was a sign of the coming separation of the individual from the community which was a feature of the Hellenistic age and of its literature.

The "Orestes," "Iphigeneia in Aulis" and "Phoenissae."—The *Orestes*, though it contains no recognition, has a climax of melodramatic excitement. Orestes and Electra have been condemned to death by the men of Argos for killing their mother—a remarkable modification of the traditional version. Menelaus, who has just returned from his wanderings, is too cowed by the threats of

Tyndareus, his father-in-law, and too pusillanimous, to take action to help Orestes. The condemned pair hit on a desperate scheme to save themselves in which Orestes' loyal friend Pylades joins them. They will punish Menelaus by killing Helen and will hold his daughter Hermione as surety for their own safety. Helen, being divine, escapes, but Menelaus is storming at the palace doors and Orestes is on the point of setting it on fire with himself and all those in the palace when Apollo appears and restores harmony—an example of the use of a god to bring about the necessary ending. A panic-stricken Phrygian adds a touch of comedy to the excitement. But the first half of the play is more in the manner of the *Electra*, and the scene in which Electra nurses her brother when plunged in delirium by the hauntings of his mother's Furies is an example of Euripides' mastery of a theme that seems to have had a particular appeal for him. It was for his scenes of love and of madness, Longinus says, that he was particularly celebrated.

Had the *Iphigeneia in Aulis* survived complete and uncorrupt it would probably be among his most popular plays. It is remarkable that in his last years Euripides should have broken new ground in such different ways as in this play and in the *Bacchae*. Here he is on the verge of writing bourgeois tragi-comedy. The realistic and contemporary atmosphere is unaffected by the implications of the myth, which are still important in the *Electra* and to some extent in the *Orestes*. When Agamemnon was told he must sacrifice his daughter if he wanted to sail against Troy, he was confronted with the problem of finding an excuse for bringing her to Aulis; he pretended that she was to be married to Achilles. In this play Agamemnon is shown in such an agony of distress that even Menelaus is prepared to call the expedition off. But it is too late, the oracle too widely known. The meeting of Agamemnon with his eager and excited daughter is poignant in a way new for Greek tragedy. Embarrassment spreads to Clytemnestra when she chances on Achilles and warmly greets him as her future son-in-law, and is answered by blank amazement. The truth emerges, and Achilles is furious at the way his name has been used. Iphigeneia begs pathetically for life and Achilles pledges her safety but finds that his own Myrmidons will not obey him. To Iphigeneia comes a sudden realization, which Aristotle found unconvincing, of her position, and she is ready to die for the Panhellenic cause. She goes willingly to her death at the altar, but presumably Artemis saves her. The present ending is a late forgery, and it is possible that Euripides never finished the play.

The *Phoenissae* was a great favourite with the Byzantines, perhaps because of its wealth of mythical content. It is a sprawling play which has been little admired since. The action covers the attack of the Seven against Thebes from before the assault through the self-sacrifice of Menoeceus and the slaughter of Eteocles and Polynices to an ending which, as it has survived, seeks to combine what scarcely can be combined, the burial of Polynices by Antigone and her departure into exile with Oedipus. Probably an interpolator has been at work here. One scene shows something of the poet's force of imagination, that in which Polynices enters Thebes by night under safe conduct, the result of Jocasta's despairing effort to reconcile her sons.

The Structure of the Plays.—Dialogue between two actors remains the staple of drama in Euripides' plays; the third actor is freely used but genuine three-cornered conversation remains extremely uncommon. In the *agon* or set debate which is a feature of most of the plays the two disputants sometimes argue before a third actor who represents the judge in these exhibitions of rhetorical virtuosity, exhibitions in which the forensic accomplishment of the speakers is not always consistent with their characters as presented in the remainder of the play, but only two actors at a time is the general rule.

The chorus tends to grow more detached from the action and in the latest plays it may consist of mere spectators. Its songs become more ornamental, the connection with the action tenuous and at times nonexistent. Even when the chorus is important, as in the *Bacchae*, it may be ignored for long stretches by the actors. In the *Ion*, surprisingly, the chorus by speaking on its own initiative affects profoundly the course of the play. As the chorus grows more remote the sung parts of the actors, especially the monodies or solos, assume greater importance.

The chief structural peculiarities of the plays are the prologues and the use of a god (*deus ex machina*) at the end. Almost all the plays start with a monologue, sometimes with dramatic motivation but more often a bare chronicle supplying information which today would be printed with the list of characters and scenes. The purpose was no doubt to achieve clarity in presenting unfamiliar versions of the myth and to save the time that a more dramatic exposition would have required. When knowledge of the future is needed the speaker is a god. Mainly for the same reason a god is introduced at the end of eleven of the surviving plays and in two others a human being is given prophetic powers. Euripides liked to tidy up his play by revealing the future fortunes of his characters and often to attach his play to some contemporary cult. The god suspended in mid-air may have been a popular spectacle, and some of the audience may have been reassured by return to comparative orthodoxy of treatment in the final scene, which is often at variance with the spirit of the rest of the play. Only in the *Orestes* is the god needed to end the impasse which the action has reached.

Characters.—Euripides had moments of brilliant psychological insight and enlarged the range of man's understanding of himself. This judgment does not mean that he was pre-eminent as a portrayer of character and in fact the men and women who appear in his plays are less memorable than those of Sophocles. This may be partly a consequence of his approach to myth and his impulse to bring heroic figures down to the scale of everyday. But it is partly that in single scenes he is stimulated to enter into the mind of a character with extraordinary completeness, particularly, it was observed by Longinus, in scenes of love and madness. But when the situation changes inspiration fades, and taking the plays as a whole we find few really striking characters though there are many moments of remarkable realization.

Style.—The word habitually used in antiquity to describe the ordinary iambic style of Euripides was *lalia* ("chatter"), alluding probably both to its comparatively light weight and to the volubility of his characters of all classes. Such a description is to be found in the *Frogs* of Aristophanes. Though not more flexible than the style of Sophocles it is less consistently dignified. It is often plain, but when it suited him Euripides was quite ready to use an unmistakably poetic diction. His lyric style has considerable charm and sweetness. In consequence of the introduction of a new style of music associated with the name of Timotheus there was a marked change in the lyrics after about 415 B.C. They became more luxuriant and emotional, sometimes to the verge of hysteria. Words are repeated freely and sentences become structureless. At its worst this style is hardly distinguishable from its parody in the *Frogs*, but in the *Bacchae*, where frenzied emotion is in place and there is something definite for the chorus to express, the songs are unsurpassed in power and beauty.

Views and Attitudes.—It is in the nature of drama that the dramatist cannot give direct expression to his own opinions unless, like Bernard Shaw, he does so in a preface. There is fairly general agreement where Euripides' sympathies lay, that they were for the "enlightenment" against tradition, so that he would not believe in gods whose conduct was, by human standards, defective. One so adept at seeing both sides of a question is unlikely to have been an extremist. But such generalizations are of no great value. Further, his views must have undergone development and change during the third of a century which separates his earliest from his latest surviving work, and of such change nothing can be traced, save that of the rancorous hatred of Sparta there is no trace after the Archidamian war. There is no agreement as to whether the *Bacchae* indicates a change of attitude toward religion.

There can be little doubt that Euripides throughout his life rejected completely the traditional picture of the gods. The objection raised by Xenophanes in the previous century, that the gods as described by Homer and Hesiod were inferior in their moral standards to men, is echoed unmistakably in the *Heracles*, and Zeus himself is the object of some plain speaking by Amphitryon in the same play. Apollo is the deity most often attacked and this attack may have been connected with the Spartan leanings of his

oracle at Delphi. Athena for a similar reason is always treated with respect. That gods are much more often presented on the stage by Euripides than by the older dramatists is due to their convenience as a source of information which could not otherwise be made available, not to any respect for them. Only in the *Hippolytus* and *Bacchae* has the divine element more than superficial significance.

At the time when Euripides was writing there must have been many among his audience who were ready to take a denial of Homeric theology quite calmly. He was on more dangerous ground when he challenged the traditional beliefs with reference to ritual purity, as he did in spectacular fashion in the *Heracles*, where Theseus refuses to concern himself with the taboo arising out of the shedding by Heracles of the blood of his kin, and in the *Auge** where it was suggested that worse things might happen in a temple than the birth of a baby.

Euripides was frequently called the philosopher of the stage, but the suggestions of the word are misleading. During his lifetime the modern conception of philosophy did not yet exist. The philosophers who influenced Euripides were mainly concerned with the nature of the material world, and the conception of an unmaterial form of reality was only then struggling into existence. That Euripides was interested in such speculations is quite clear. Not only does he echo in the *Orestes* Anaxagoras' famous assertion that the sun was an incandescent lump of rock, but his statement in fragment 839 (of the collected fragments of Euripides' lost plays) that what men call birth and death is only a change in the combinations of material particles to form new entities is of doubtful relevance to the context of any conceivable play. In his *Melanippe the Wise** he allowed his heroine, in a speech censured by Aristotle, to try to prove on philosophic principles that her twin sons by Poseidon who were being suckled by a cow could not, as her father supposed, be the offspring of the cow, as such a birth would be in defiance of unalterable laws of nature.

In spite, however, of numerous statements by later writers that Euripides was a pupil of Anaxagoras there is no clear reference to his most celebrated doctrine, that of *Nous*, the Intelligence that shaped the world, and it is difficult to avoid the suspicion that there has been an artificial association of the most famous philosopher of the mid-5th century and its most philosophical poet. There is more reason to suspect the influence of a pupil of Anaxagoras, Diogenes of Apollonia, the last of the school of Ionian philosophers whose characteristic it was to find the source of all things in a single physical substance, in his case air. *Aether* which was sometimes regarded as the purer, more fiery air of the sky remote from earth, does not seem to have been distinguished from plain air by Diogenes, but Euripides preferred the word with the loftier associations. In a number of striking passages in the plays and fragments *aether* is addressed as the supreme power or equated with more familiar forms of divinity. It seems to be the object invoked in Hecuba's famous prayer in the *Troades* "O thou that art the prop of earth and hast thy home upon the earth, whoever thou art, hard to conjecture, Zeus, or necessity of nature, or mind of men, to thee I pray; for treading thy silent path thou bringest all mortal things to pass according to justice." The first part of this prayer can be explained in the light of fragment 941, where *aether* is described as holding the earth in its life-giving embrace, and the last from fragment 1018, "The mind in each of us is god." "Mind of man" suggests something other than the *Nous* of Anaxagoras. At the beginning of his contest with Aeschylus in the *Frogs* Euripides himself makes his prayer to novel deities "*aether* and the hinge of the tongue," or science and rhetoric. His prayer is not heard any more than is the prayer of Hecuba who hopes to see justice done by the punishment of Helen whom she is about to denounce.

Two matters of great importance for a tragic poet are closely connected with this conception of *aether*, the nature of man's soul and the place of justice in the government or the working of the universe. If *aether*, the purest form of matter and the nearest, probably, that Diogenes could get to the conception of the immaterial, is the substance of man's soul, it may be immortal and, if there is room for such a term within the system, divine. And this indeed seems to have been, if not the opinion of Euripides, an idea by which he set some store. In both the *Helen* and the *Supplices* he states that the mind (*nous*) of the dead does not indeed live, but it possesses eternal consciousness, being mingled after death with the eternal *aether*. That this conception of the soul as divine material that is merged with the general mass of the divine substance, thus preserving consciousness but not individuality, was not unfamiliar at this time is shown by its occurrence in an inscription on the monument for those who fell before Potidaea at the beginning of the Peloponnesian War.

This is not the only conception of the soul and its lot after death to be found in Euripides. Besides numerous conventional references to Hades there is also in the *Iphigeneia in Aulis* and in fragment 532 the common Greek notion of death as a lasting sleep, and also in fragments 638 and 833 the Orphic notion that this life is really death and the true life is beyond the grave.

With regard to the larger question of divine justice, it is clear that rewards and punishment beyond the grave are no more than a remote possibility. It is on its record here that life must be judged. In the extant plays are to be found neither such reconciliations as concluded some of the trilogies of Aeschylus nor the serene resignation of Sophocles. The ostensible rulers of the world are denounced in the *Heracles* and the *Electra*, made faintly ridiculous in the *Ion*, and revealed as a blind force in the *Bacchae*. Once they are swept away as an inadequate symbol it is not clear what is left, whether *aether* or chance. In some moods Euripides saw chance as supreme. "Zeus, what am I to say?" says the kindly Talthybius at sight of the fallen Hecuba, "that you have regard for men? Or is this a vain illusion and chance supreme over the affairs of mortals?" And it is doubtful whether the outcome of the *Electra* gives any warrant for a hopeful answer to the challenge "Who will believe in gods if justice is not done?" And of the *Heracles* it can be said that the play as a whole seems to be a denial of divine providence.

If there is no expectation of justice from on high, what of the dealings of men among themselves? Whatever may be the precise meaning of the Greek philosopher Protagoras' saying "Man is the measure of all things," there is no doubt that in the second half of the 5th century the basis of morality was widely questioned and a distinction made between the law of nature whose violation brings its own punishment, and the law of men that can be violated with impunity by anyone so long as he is not found out. There is surprisingly little reference to this idea of ethical relativity; in the lost *Aeolus** the question of the validity of the prohibition of incest seems to have been raised, and in the *Phoenissae* Eteocles defended his unscrupulous violation of the agreement to share the throne of Thebes with his brother by arguing that men use words like "noble" and "wise" but mean different things by them. But the great discovery of Euripides in the realm of behaviour was that man is not the master of his own decisions. He revealed in the *Hippolytus* that man must be regarded as an irrational creature. "Men seem to me to go wrong not because of a defect in their natural understanding, for sound judgment is common; but we should look at it like this: we know what is good and recognize it, but we fail to carry it out, some from laziness, others because they prefer some pleasure to the good." This was Phaedra's analysis. Medea in her agony of indecision expressed the same thought. The fragments of the *Chrysippus** show that a similar situation occurred there. It is a conception that can cause a modification of judgment in many cases where the common Greek view would lead to immediate condemnation, and it is not unconnected with what seemed to many to be Euripides' tenderness toward the wicked. At all events the denial of the supremacy of reason is in direct contradiction to the Socratic paradox of the dependence of moral goodness on knowledge. It has even been supposed that the relevant speeches in the *Medea* and in the *Hippolytus* were written with direct reference to Socratic doctrine.

About Euripides' own political attitude not a great deal can be safely inferred. For him the city was the focus of life, even though his apparent political inactivity and praise for the retired and contemplative life may foreshadow the growth of individualism. Athens was never more nobly praised than in his chorus of

the *Medea*, and if praise there is for Athens as the home of the arts and humanity he showed in the *Supplices* and the *Andromache* that he was not immune to the more chauvinistic emotions which go with patriotism, either in celebration of the glory of his own city or in denunciation of its enemies. Plays from the early years of the Peloponnesian War show unhesitating participation in the war spirit, and so no doubt did the *Erechtheus**. The *Supplices* may conceivably have a connection with Alcibiades and his policy of alliance with Argos. There is more reason to see in the *Troades* a manifesto against the Sicilian adventure, and the *Bellerophon** may have dealt in part with the seamy side of imperialism. Whether Euripides' final departure from Athens had anything to do with his lack of sympathy for the restored democratic regime after 410 B.C. must remain matter for surmise.

Euripides' own political insights emerge mainly in the *Supplices*, and in the *Orestes* written when he was an old man, and are not necessarily evidence for what he thought in his earlier years. In both a preference for the middle class is manifest. It is praised by Theseus, rather incongruously by contrast with the upper class, which is governed by avarice and ambition, and the lower, whose jealousy of the rich makes it a willing tool in the hands of worthless demagogues; these are described with some shrewdness by the Theban herald, while the ambition of the aristocracy is ludicrously illustrated in the *Iphigeneia in Aulis* by the bogus cordiality of Agamemnon when he was canvassing for the office of generalissimo—one of Euripides' most daring anachronisms. In the gallery of political portraits contained in the report of the meeting of the assembly for the trial of Orestes and Electra in the *Orestes* the peasant freeholder, the man who rarely comes to the city, is the only one who emerges with credit. The peasant husband of Electra in the play of that name is of the same class, and the one character in the play who is not more or less disagreeable.

An acuteness of observation which calls to mind Thucydides himself is to be found in the debate in the *Supplices* between Theseus and the Theban herald. Here the herald raises the valid objection to the Athenian form of democratic government that the ordinary citizen, even if he has the intelligence, has not the leisure to put himself into a position to take decisions on matters of high policy, to which Theseus replies not altogether convincingly by pointing out the disadvantages of tyranny, whereas under a democracy every man is free to put his ideas at the service of the city. Even more Thucydidean is the herald's warning not to take the decision of peace or war under the influence of emotion. "Hope is a bane to mortals; often has she embroiled cities by raising spirits to too high a confidence. When it comes to voting on war no one thinks of his own death but each assumes that it will be on someone else that misfortune will fall. If men had had death before their eyes when they cast their votes, Greece would not have been perishing in a frenzy for war." Theseus' war may have been a credit to Athens, but the war was prehistoric and the herald's warnings have an immediate application.

Though Euripides was quick to see the point of view of the downtrodden and to sympathize with the oppressed, whether slaves or women or the victims of war—it is no accident that he wrote so many of his plays from the point of view of the defeated—it should be remembered that there is no sign that he ever considered the possibility of changing the institutions on which slavery, for instance, depended. Such lines as "The name of slave will not destroy a good man, and many slaves are better men than the free" (fragment 511), and similar expressions in the extant plays, the *Helen* and the *Ion*, together with a keen awareness of the hardship of a woman's lot, which is best illustrated by *Medea* 230–51 (a passage without relevance to Medea's own situation), hardly justify seeing Euripides as a standard-bearer of liberal humanism. The possibility of abolishing slavery never seems to have been seriously considered in the ancient world; that it was even questioned as it was in the next century may have been partly due to Euripides' ability to look on human customs and institutions with a mind unprejudiced by convention. More cannot be claimed for him.

With regard to women Euripides' reputation is perplexing. It would seem that the dramatist who numbered among his heroines Alcestis and Macaria and Iphigeneia, and who portrayed Medea and Phaedra and even Clytemnestra in such a way as to arouse sympathy, should count as the champion of women. However, in his own day his reputation was that of a woman-hater, and a whole play of Aristophanes, the *Thesmophoriazusae*, turns on the indignation of the Athenian women at their portrayal by Euripides. It is true that the plays containing the wickedest of his women, the Phaedra of the first *Hippolytus** or his *Stheneboea**, are lost; and it is true that Euripides' taste for generalizations when applied to women, though it works both ways, is usually at their expense. In some cases, as in the *Hippolytus*, the abuse is fully explained by the dramatic context, but this is by no means always so. Whether the stories of his own domestic unhappiness arose from the hostility to women expressed in the plays, or whether the plays are to be explained by his private life, there is no means of judging.

See also references under "Euripides" in the Index volume.

BIBLIOGRAPHY.—*History and Transmission of the Text:* U. von Wilamowitz-Moellendorf, *Analecta Euripidea* (1875) and the introduction to *Herakles* (1889, etc.); D. L. Page, *Actors' Interpellations in Greek Tragedy* (1934); A. Turyn, *The Byzantine Manuscript Tradition of the Tragedies of Euripides* (1957).

Editiones Principes: J. Lascaris, *Medea, Hippolytus, Alcestis, Andromache* (1496); M. Musurus, all plays except *Electra* (1503); P. Victorius, *Electra* (1545).

Critical Editions: A. Kirchoff (1855); Gilbert Murray (1902–9).

Fragments: A. Nauck, *Tragicorum Graecorum Fragmenta* (1926); D. L. Page, *Greek Literary Papyri* (1941). R. A. Pack, *Greek and Latin Literary Texts From Graeco-Roman Egypt* (1952) lists all known fragments.

Scholia: E. Schwartz (ed.) (1887).

Commentaries: F. A. Paley, 2nd ed., 3 vol. (1872–80); H. Weil, *Sept Tragédies d'Euripide*, 3rd ed. (1905). *Individual Plays:* U. von Wilamowitz-Moellendorf, *Herakles* (*see* above), *Hippolytus* (1891), *Ion* (1926); N. Wecklein, *Medea, Iphigeneia in Tauris, Bacchae, Hippolytus, Phoenissae, Electra, Orestes, Helen, Andromache, Ion, Supplices, Iphigeneia in Aulis* (1876–1914); A. C. Pearson, *Helen* (1903), *Heracleidae* (1907); *Phoenissae* (1909); in the Oxford *Euripides* series: D. L. Page, *Medea* (1938), M. Platnauer, *Iphigeneia in Tauris* (1938), A. S. Owen, *Ion* (1939), J. D. Denniston, *Electra* (1939), E. R. Dodds, *Bacchae* (1944), A. M. Dale, *Alcestis* (1954).

Books and Articles: A. Dieterich, in Pauly-Wissowa's *Real-Encyclopädie der classischen Altertumswissenschaft*, vi (1907); A. W. Pickard-Cambridge, in *Oxford Classical Dictionary* (1949); Gilbert Murray, *Euripides and His Age* (1913, etc.); G. M. A. Grube, *The Drama of Euripides* (1941); A. Rivier, *Essai sur le tragique d'Euripide* (1944); R. P. Winnington-Ingram, *Euripides and Dionysus* (1948); L. H. G. Greenwood, *Aspects of Euripidean Tragedy* (1953); G. Zuntz, *The Political Plays of Euripides* (1955). The books of A. W. Verrall are still of interest though few would accept his conclusions: *Euripides the Rationalist* (1895), *Essays on Four Plays of Euripides* (1905), *The Bacchants of Euripides* (1910). On new evidence from papyri *see* A. W. Pickard-Cambridge in *New Chapters in the History of Greek Literature*, 3rd series, ed. by J. U. Powell (1933). Among books in which considerable attention is devoted to Euripides are: H. D. S. Kitto, *Greek Tragedy*, 2nd ed. (1950); W. Schmid and O. Staehlin, *Geschichte der Griechischen Literatur* 1:3 (1940), the most ample account; D. W. Lucas, *The Greek Tragic Poets* (1950); E. R. Dodds, *The Greeks and the Irrational* (1951). A. Lesky, *Die Tragische Dichtung der Hellenen*, 148–213 (1956); T. B. L. Webster, on "Tragedy," in *Fifty Years of Classical Scholarship*, ed. by M. Platnauer (1954).

Translations: Into English prose: E. P. Coleridge (1891; etc.). A French prose version together with the Greek text is printed in the Budé edition (1925). Fifteen plays have so far been published ed. by L. Méridier, L. Parmentier and H. Grégoire. English verse translation by A. S. Way printed with Greek text, Loeb series (1912) and in *The Complete Greek Tragedies*, ed. by David Grene and Richmond Lattimore (1959). Among numerous translations of single plays the renderings of nine plays by Gilbert Murray should be mentioned (1902–54).

Euripides' Influence: F. L. Lucas, *Euripides (Our Debt to Greece and Rome)* (1924); G. Highet, *The Classical Tradition* (1949).

(D. W. Ls.)

EUROMART: *see* ECONOMIC UNION.

EUROPA, in Greek mythology, according to Homer the daughter of Phoenix or, in a later story, of Agenor, king of Phoenicia. The beauty of Europa fired the love of Zeus, who approached her in the form of a white bull and carried her away from Phoenicia to Crete, where she became the mother of King Minos of Crete, of King Rhadamanthys of the Cyclades islands (both eventually judges of the dead) and of Prince Sarpedon of Lycia. She was worshiped under the name of Hellotis in Crete, where the festival Hellotia, at which her bones, wreathed in myrtle, were carried round, was held in her honour. She is apparently a per-

sonification of the continent of Europe.

EUROPE, the smallest (except Australia) of those principal divisions of the land surface of the globe that are usually distinguished by the conventional name of continents. The earliest mention of Europe is in the Homeric *Hymn to Apollo,* but there Europe is not the name of a continent, but is opposed to the Peloponnese and the islands of the Aegean. The distinction between Europe and Asia is found, however, in Aeschylus in the 5th century B.C., but there seems to be little doubt that this opposition was learned by the Greeks from some Asian people. On Assyrian monuments the contrast between *asu* "(the land of) the rising sun" and *ereb* or *irib* "(the land of) darkness" or "the setting sun" is frequent, and these names were probably passed on by the Phoenicians to the Greeks, and gave rise to the names of Asia and Europe. Where the names originated the geographical distinction was clearly marked by the intervention of the sea, and this intervention marked equally clearly the distinction between Europe and Libya (Africa). The difficulty of fixing the boundary between Europe and Asia where there is land connection has always caused uncertainty. Even in the 3rd century B.C. Europe was regarded by Eratosthenes of Alexandria as including all that was then known of northern Asia. But physical features and climate finally determined that what is known as Europe came to be occupied by more or less populous regions in intimate relation with one another, but separated on the east by sparsely peopled areas from the countries of northern central and southwestern Asia. Within the limits thus marked the inhabitants have had a common history in a greater measure than those of any other continent. There has long been doubt as to the exact boundary between Europe and Asia; in general, the watershed of the Ural mountains and that of the Caucasus has been taken, though sometimes the latter is replaced by the official frontier of the U.S.S.R.

A conference of Russian geographers in 1958 made recommendations, subsequently accepted by the Soviet government. The boundary now follows the foot of the eastern slopes of the Urals and their prolongation the Mugodzhar hills, the Emba river, the northern shore of the Caspian sea, the Kumo-Manychskaya Vpadina (depression) and the Kerchenski strait to the Black sea. Thus the Urals are now all included in Europe, the Caucasus in Asia.

Extent.—Europe consists of a peninsula projecting westward into the Atlantic ocean from the Asiatic land mass, known as Eurasia, separated from Africa to the south by the Mediterranean sea, about 1,000 mi. in length. Its most northerly point is the North cape (Nordkapp) of Norway (71° 11′ N.), its most southerly is Cape Tarifa on the Strait of Gibraltar (36° 01′ N.), giving a north–south dimension of about 2,400 mi. Its most westerly point is Cape Roca (Cabo da Roca) in Portugal (9° 30′ W.), excluding Ireland which extends to about 10° 27′ W. in the southwest. If the eastern boundary is taken to be the Urals (57° E.), this gives an over-all width of about 3,300 mi. The area of "peninsular Europe" (*i.e.,* excluding European Russia) is about 1,911,000 sq.mi., or with European Russia more than 3,990,000 sq.mi (3,860,000 sq.mi. according to the Soviet definition), about 7% of the land surface of the globe.

The coast line is long, for not only is the continent flanked on three sides by sea, but numerous indentations alternate with prominent peninsulas. No part of peninsular Europe is more than 500 mi. from the sea. The submergence in relatively recent (Neolithic) times of the low-lying margins has produced on the one hand a continental shelf that extends westward from Land's End in England for 200 mi. and on the other hand the marginal Baltic and North seas, less than 100 fathoms (600 ft.) deep; the passages between the Danish islands into the Baltic are only 11 fathoms deep. From the continental shelf rise islands, notably the British Isles. By contrast, the Mediterranean, with its several basins interrupted by peninsulas and islands, large and small, is the result of the foundering of a complicated structural area (part of the Alpine folded system discussed below). The Mediterranean basin has some depths exceeding 2,000 fathoms, the deepest being 2,533 fathoms between Greece and Crete. The Black sea (deepest sounding 1,227 fathoms) is separated from the Mediterranean by a narrow series of straits (Dardanelles, Sea of Marmara, Bosporus). The Adriatic sea is a shallow, elongated depression lying between the roughly parallel ranges of the Apennines and the Dinaric Alps. Numerous islands occur in the Mediterranean, from the large ones of Sicily, Sardinia and Corsica to the many small ones in the Aegean sea.

This article contains the following sections and subsections:

I. Physical Geography
  1. Geological History
  2. Relief
  3. Drainage
  4. Climate
  5. Vegetation
  6. Animal Life
II. Natural Resources
  1. Soils
  2. Minerals
  3. Water Resources
  4. Land Use
  5. Fisheries
  6. Industries
III. Anthropology
  1. Physical Anthropology
  2. Languages
  3. Ethnology
IV. Archaeology
  A. Western and Central Europe
    1. Early Paleolithic
    2. Later Paleolithic
    3. Mesolithic
    4. Early Neolithic
    5. Middle and Late Neolithic
    6. Bronze Age
    7. Iron Age
  B. Eastern Europe
    1. Paleolithic Period
    2. Mesolithic and Neolithic Periods
    3. Bronze Age
    4. Iron Age
V. History
  A. Prehistory
    1. Western Mediterranean Peoples
    2. Celts
    3. Teutonic Peoples
    4. Eastern Peoples
  B. Greeks, Romans and Barbarians
    1. Greeks
    2. Romans
    3. The Migrations
  C. The Middle Ages
    1. Germanic Kingdoms
    2. Genesis of Latin Christendom
    3. Final Breakdown of Mediterranean Unity
    4. Carolingian Empire
    5. Carolingian Renaissance
    6. Collapse of the Carolingian Empire
    7. Foundation of the Medieval Empire
    8. Feudalism
    9. Manorialism
    10. 12th-Century Revival
    11. Holy Roman Empire and the Apogee of the Medieval Papacy
    12. Rise of the Western Monarchies
    13. Eastern Frontiers of Medieval Europe
    14. The Church in the Later Middle Ages
    15. Fall of Constantinople
    16. Italy and Germany in the Later Middle Ages
  D. The Emergence of Modern Europe, 1492–1648
    1. Modern Europe and the Renaissance
    2. The New World and Economic Development
    3. Nation-States and Dynastic Rivalries
    4. Turkey and Eastern Europe
    5. Reformation and Counter-Reformation
    6. The Wars of Religion
    7. Sweden and the Thirty Years' War
  E. The Great Age of Monarchy, 1648–1789
    1. Control of the State
    2. Machinery of the State
    3. Provincial Government
    4. Government Officials
    5. Internal Functions of the State
    6. The State and Economic Enterprise
    7. The State in Its External Relations
    8. Military Establishment
    9. Balance of Power
    10. Eastern Europe and the Eastern Question
    11. Basic Change and Crisis

F. The French Revolution and Napoleon, 1789–1815
1. The Revolution in France
2. The Terror
3. Napoleon in Power
G. The 19th Century, 1815–1914
1. The Restoration
2. The Economic Revolution
3. Liberalism and Nationalism
4. New National States in the Balkans
5. The Remaking of Central Europe, 1850–71
6. Rise of Socialism
7. Diplomacy and War, 1878–1914
8. European Civilization in 1914
H. The 20th Century
1. World War I and the Russian Revolution
2. Interwar Years
3. World War II
4. Communism and Europe
VI. Population
1. Component States
2. Distribution of Population

I. PHYSICAL GEOGRAPHY

The physical geography of individual countries is more fully discussed in the country articles. There are also separate articles on the major physical features.

1. Geological History.—The structural nucleus of Europe consists of a mass of near-horizontal but complex and highly metamorphosed Archean rocks, known as the Baltic shield. Exposed around the margins of the Baltic sea, they are continued eastward as the Russian platform, buried beneath newer sedimentary rocks. At least four major phases of mountain building have occurred, separated by long periods of quiescence during which the ranges so formed were attacked by the forces of denudation. The materials thus derived were laid down in seas resulting from the fluctuating marine transgressions, in inland lakes, or on the land surface itself, so forming various sedimentary rocks. Thus the Upper Carboniferous was largely a period of swamp forests, and the Jurassic and Cretaceous were periods of fluctuating estuarine and deep-sea deposition.

The earliest orogeny, sometimes known as the Huronian, is of Pre-Cambrian age, and only vestiges can be traced in northern Norway, the Lofoten Islands and the Outer Hebrides, involving very ancient gneissic rocks. These relics may be fragments of a vanished northwestern continent. Between this land mass and the Baltic shield, mountain ranges were formed as a result of the Caledonian orogeny, of post-Silurian age, which mainly affected northwestern Europe, including the British Isles. The strike of these folds was east-northeast to west-southwest; they can be traced across western Scotland and again in western Scandinavia, once probably continuous. Other folds of later stages of the Caledonian can be traced in the Southern Uplands of Scotland, the English Lake district, northwestern Wales and southern Ireland; in these districts the direction of folding was more east–west. The Caledonian ranges have been subject to long-sustained denudation, virtually obliterating their original structural forms. Subsequent uplift and renewed dissection have revealed the ancient structures in the present "grain" of the country, in the patterns of the hills and valleys.

A third orogenic phase, of Carboniferous–Permian age, produced a series of ranges across central Europe, sometimes referred to as the Altaides. This orogeny is generally known as the Hercynian, though the term Armorican is applied to the more westerly folds in the British Isles and Brittany, the term Variscan to the rest. The Armorican folds run at first from west to east, gradually changing in central France until their direction is northwest to southeast. The Variscan trend is more southwest to northeast, and thus a complex structural "apex" occurs in the south of the Massif Central of France where these systems met. The Hercynian mountains extend eastward into southern Poland, and an outlying range of comparable age is the Urals. The mountains were subsequently subjected to denudation, were again uplifted and faulted during a later orogeny, and in some cases were tilted en bloc. They now comprise individual massifs of Pre-Cambrian and Paleozoic rocks—slate, schist, sandstone, quartzite and limestone, together with intrusive granitic masses of various ages and some superficial products of mid-Tertiary volcanic activity.

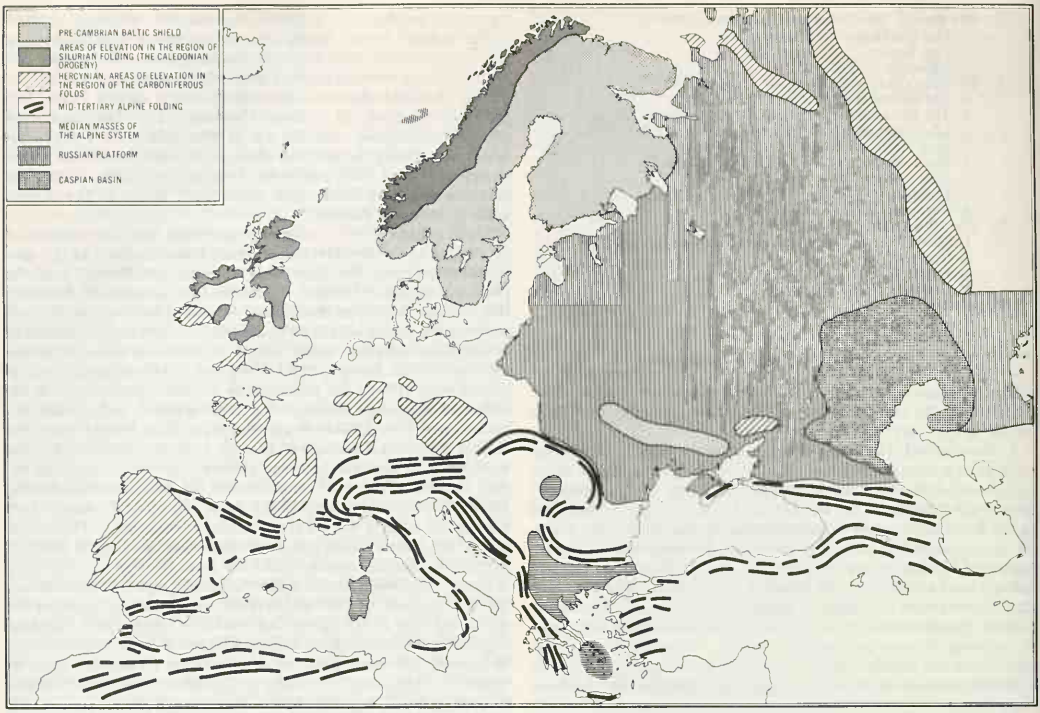
The massifs reveal, in the form of extensive uniform surface levels, the effects of periods of peneplanation. The blocks rise as horsts from the surrounding lowlands, ranging in size from the Massif Central (which occupies a sixth of the area of France) to the Harz mountains of northern Germany. The Hercynian orogeny took place soon after the age of the widespread Upper Carboniferous swamp forests, the remains of which have been metamorphosed into coal, preserved in synclinal basins along the northern flanks of the uplands from south Wales to the Donets basin in southern Russia.

Mid-Tertiary earth movements upfolded vast thicknesses of Secondary and Older Tertiary sediments, accumulated in the geosyncline between the Hercynian foreland of Europe and the plateau-continent of Africa. The direction, nature and degree of this folding (known as the Alpine) depended both on the tectonic compressive forces affecting the geosyncline, and on the position of the rigid outer horsts, which helped to determine their alignment. On the northern margins the folding was relatively gentle, yet its effects were felt as far as southern Britain. It resurrected the elongated west–east dome of the Lake district, and formed numerous periclines in Dorset and the Hampshire basin (where the folding was in places severe enough to invert the Chalk), the Wealden anticline and the broad shallow syncline of the Paris basin. Farther to the south, more intense but still superficial folding formed the arcuate ranges of the Jura. In central and southern Europe the folding was of great intensity, resulting in the formation of recumbent folds, and carrying forward of several series of nappes, overthrust masses forced far from their roots. Involved in these folds, which consist mainly of metamorphosed sedimentary rocks, are blocks of crystalline rocks, comprising portions of the splintered edge of the Hercynian continent caught up in the folding, the deep-seated cores of the folds, and a series of batholithic intrusions. In many parts these crystalline rocks have been revealed by denudation, forming a crystalline axis to the ranges, flanked by limestones and shales.

The Alpine orogeny was responsible not only for these young fold mountains, which form the highest ranges of Europe, but it also uplifted plateaus, such as the Rhodope block in the Balkan peninsula and parts of Corsica and Sardinia, and affected the Hercynian uplands. Conversely, basins and depressions were created, some of which became areas of sea (the western Mediterranean, the Adriatic and the Aegean), while others now infilled with sediment form level plains among the mountain systems, such as northern Italy and central Hungary. A further result was volcanic activity and crustal instability generally. This included the emission of lava to solidify as sheets of basalt in Iceland, northwestern Britain and central France, and the formation of volcanic peaks on a small scale in the Auvergne mountains and on a larger scale in the Mediterranean (Vesuvius, Etna and Stromboli).

The Quaternary Glaciation.—A period of large-scale glaciation, lasting with interglacial intermissions for about 600,000 years and still affecting the higher and more northerly parts, has produced marked results on the landscape of Europe, both in the lowlands and highlands. The continental ice sheets advanced southward from the main centre of accumulation over the Baltic shield, and smaller valley glaciers moved outward from the Alps, Pyrenees, Carpathians and uplands of northern Britain. In the higher parts the work was mainly that of erosion—cirques, arêtes, pyramidal peaks, hanging valleys, truncated spurs and U-shaped main valleys. Over the lowlands the work was chiefly that of deposition—long lines of terminal moraines, sheets of boulder clay of the ground moraine and extensive fluvioglacial sands and gravels. In addition, winds blowing outward from the ice sheets removed vast quantities of fine material, and laid it down as loess (*limon*) in a belt from southern Russia to northern France.

Since the ice sheets, once 6,000 ft. thick, began to disappear, water has been released by the melting of ice. This has produced a rise in sea level (though complicated by isostatic recovery because of the relief of the ice weight), with marked effects on the coast line. The Baltic sea, once a lake, was linked to the North sea by narrow straits between the Danish islands, and Britain was



FROM M. R. SHACKLETON, "EUROPE, A REGIONAL GEOGRAPHY"; REPRODUCED BY PERMISSION OF LONGMANS, GREEN & CO., LTD.

MAIN STRUCTURAL REGIONS OF EUROPE

separated from the continent. The submergence of glacial valleys produced fiords, of river valleys produced rias and funnel-shaped estuaries, and of low-lying coasts produced tidal flats, broad lagoons such as the Zuider Zee, and numerous offshore islands. Since the withdrawal of the ice sheets, the river systems have established themselves as the main agents which modify the landscape. (*See* also GEOLOGY.)

2. Relief.—The net result of the geological history thus outlined has been to produce a continent of marked physical diversity —of ancient upland masses, younger fold mountain ranges and coastal plains, basins and valleys which border the seas or form re-entrants within the uplands. They can be summarized in terms of five major relief regions.

Fennoscandia.—A convenient name used to designate the Caledonian mountains and the Baltic shield, occupying Norway, most of Sweden and Finland. The high plateau of the Scandinavian peninsula (the *fjeld*) drops steeply westward to the fiord-penetrated coast of the North sea. Much of the *fjeld* lies at 3,000–4,000 ft., gently undulating, with low rounded summits and much bare rock. In places some bold peaks rise above the general level, notably Galdhöppigen (8,100 ft.) in the Jotunheimen, the highest summit in northern Europe. Some of the uplands are still covered with ice sheets, notably the Jostedalsbre of about 600 sq.mi. (which, apart from Vatnajökull (3,385 sq.mi.) in Iceland is the largest in Europe), from which small glaciers move down to the west. In the east the plateau descends in a series of steps to the Baltic coastal plain, crossed by many parallel rivers, the upper valleys of which contain narrow "Glint line" lakes, some as much as 100 mi. in length. The plateau of Finland, 250–500 ft. above sea level, has been scoured by the ice sheets and its undulating surface consists mainly of bare rocks, mostly granite, scattered sheets of glacial material and irregularly shaped lakes, estimated to number 60,000. The land rises gently northward toward low ranges of about 1,000 ft. (the Maanselka and the Sualaselka) which form the Baltic–Arctic divide.

North European Lowlands.—This region is far from uniform and extends westward from the vast plains of Russia as far as the Bay of Biscay. In southwestern and northern France occur varied rocks of limestone, chalk, sandstone and clay; the first three stand out as lines of low hills, and the clays form broad vales. Central Belgium consists of low undulating plateaus. Even in Poland and Germany, where the plain is most continuous, the vast sheets of glacial material form an uneven surface; some of the morainic hills (notably the Baltic heights) exceed 1,000 ft. The several large rivers have cut broad valleys across these sheets, leaving blocks of higher country, the heathlands, such as the Lüneburg heath. Again, in Denmark the underlying platform of chalk, which appears only in places on the east coast as high cliffs, is covered with newer deposits: a line of terminal moraines in central Jutland, tracts of sand and gravel in the west, and boulder clay in the east and in the islands. To the south of the heathlands a narrow zone is covered with loess; this is known as the Börde (borderland) in Germany. The lowest and flattest parts of the plain lie near the North sea coasts of Denmark, Germany, the Netherlands and Belgium, where parts are below mean sea level. These now form the drained polderlands, protected on the seaward side by lines of dunes or dikes and bordered with tidal flats and islands.

Central Uplands.—These discontinuous remnants of the Hercynian ranges lie along the southern edge of the North European lowlands, in places projecting into or rising prominently from these plains. They consist of Paleozoic rocks in great diversity—sandstones, quartzites, shales and limestones. The Meseta occupies much of the Iberian peninsula, mostly about 2,000 ft., though with central ridges (*sierras*) rising to 8,000 ft. and a prominent tilted edge on the east and south. In France is the Massif Central, also tilted to the southeast (the Cévennes), with gently undulating

plateaus, down-faulted basins and valleys and volcanic peaks culminating in the Puy de Sancy (6,188 ft.). The peninsula of Brittany consists of granite plateaus, mostly between 600 and 1,200 ft., with basins (*e.g.*, Châteaulin) and valleys between the uplands, and dropping steeply to the coasts. Across the frontiers of France, Germany and Belgium lie a complicated series of uplands. The Vosges and the Black Forest are separated by the rift valley of the Rhine, while farther north the Middle Rhine highlands are crossed and divided into sections by the gorgelike valleys of the Rhine and its tributaries. The western part is known as the Ardennes, rising to 2,277 ft., and other blocks include the Eifel, Taunus and Hunsrück, with a markedly peneplaned surface at about 1,500 ft. Farther east is the diamond-shaped plateau of Bohemia, with its bordering mountain rims: on the north the Sudeten mountains, on the northwest the Erzgebirge and on the south the Bohemian Forest, with bold rounded granite summits at 3,000–4,000 ft. The Harz mountains, also consisting mainly of granite, stand out prominently, culminating in the summit of the Brocken (3,747 ft.). In some parts of this central zone are basins, with deposits of chalk, limestone and sandstone, forming scarplands such as the Fränkische Alb (Franconian Jura) and Schwäbische Alb (Swabian Jura) in southern Germany, separated by river valleys.

Young Fold Mountains.—These form the highest parts of Europe. The Alps can be traced in a sweeping curve from the Mediterranean sea near the Franco-Italian frontier, curving northward and then eastward and becoming gradually lower until they disappear toward the Danube valley near Vienna. Many parallel ranges are separated by deeply cut valleys, occupied by rivers such as the upper Rhône, upper Rhine and Inn. The rocks of these ranges include limestone (forming much of the Bernese Oberland), shales, schists and granite. The last occurs in great masses (Aar-St. Gotthard, Mont Blanc–Aiguilles Rouges, Pelvoux, Belledonne, Mercantour), revealed by denudation as high pinnacled ridges; the Mont Blanc group culminates in Mont Blanc itself (15,771 ft.), a snow hump forming the highest Alpine summit. Glaciers and snowfields are a striking feature, especially in the Oberland and the Pennine Alps; the longest glacier is the Aletsch.

Several other ranges may be traced within southern Europe. From the French Alps, the Apennines curve away southeast through the Italian peninsula, without any complete break for 600 mi. Much consists of limestone, including the Gran Sasso d'Italia mountains culminating in Corno Grande (9,560 ft.), the highest in the system. There are numerous extinct volcanic peaks, with crater lakes, and in the west Vesuvius overlooks the Bay of Naples. The ranges continue beyond the Strait of Messina into northern Sicily, and after a break the Atlas mountains continue this line through north Africa, then northward beyond the Strait of Gibraltar into the Sierra Nevada of southern Spain, ending in the Balearic Islands. The break at Gibraltar is only 14 mi. wide; the depths are about 200 fathoms, sloping steeply on either side to the floors of the Atlantic and the Mediterranean.

Lower ranges, the Provence Alps, run westward from the French (western) Alps. After an interruption by the Gulf of Lions (Golfe du Lion), the Pyrenees form a lofty and continuous barrier more than 250 mi. in length between the Mediterranean and the Bay of Biscay; their highest summit, the Pico de Aneto in the Maladeta, attains 11,168 ft.

From the eastern Alps two main branches can be traced. One trends southeastward into the Balkans as the Dinaric Alps, with limestone ranges and depressions broadly parallel to the Adriatic coast, and on into Greece as the Pindus mountains. In both Yugoslavia and Greece the summits exceed 8,000 ft., Mt. Olympus in the east overlooking the sea rising to 9,570 ft. The other branch, after a break near Vienna by the Danube valley, forms the great curve of the Carpathian mountains, enclosing the plateau of Transylvania. Though there is no permanent snow or ice, some of the summits are rugged pyramids, with deep cirques on their flanks. The highest peak, the Stalinov Stit (8,711 ft.) stands on the frontier ridge between Poland and Czechoslovakia. After another break at the Danube gorge, the Balkan mountains (Stara Planina) curve into Bulgaria and continue eastward toward the Black sea.

Mediterranean Europe, the southern part of the fold region, is an area of peninsulas, islands and basins, with ranges, mostly of limestone, rising steeply from the coast. Level land is limited to the narrow coastal plain, as in southeastern Spain, peninsular Italy and around the shores of Greece. The coastal plain widens only at the mouths of the Rhône, Arno, Tiber and some of the Greek rivers.

Eastern Europe.—Relief consists of vast plains of sedimentary rocks, which deeply cover the ancient rocks of the Russian platform. The surface slowly rises eastward to the Urals. Except for the gently swelling morainic ridge of the Valdai hills to the northwest of Moscow, nowhere does the land rise as much as 1,000 ft. above sea level; most of it is below 500 ft., and parts around the Caspian sea are below sea level. The north and northwest is covered with an irregular mantle of boulder clay and morainic material, and with sheets of sands and gravels. Some of these areas are badly drained and form marshes and peat bogs; the largest are the Pripet marshes near the Russian–Polish frontier. Windblown loess covers much of southern Russia. Across the plains several large rivers flow southward toward the Black and Caspian seas.

3. Drainage.—Most of Europe's drainage finds its way to the sea in a number of master river systems, the watersheds of which interlock in a most complex manner. They may be divided into three groups: one drains west and north to the Atlantic and its marginal seas, the second to the Mediterranean and the third east and southeast to the Black and Caspian seas.

The first group includes in the west the rivers of the Iberian peninsula: the Guadalquivir, the Guadiana, the Tagus and the Douro, which flow westward across the high basins of the Meseta to the Atlantic. The Garonne and the Loire, with their numerous tributaries, drain much of western France before they enter the Bay of Biscay, and the Seine crosses the broad synclinal lowland of the Paris basin on its way to the English channel. The Scheldt and its affluents (Lys, Scarpe, Dender, Demer) drain the Plain of Flanders and the low plateaus of central Belgium. The Meuse pursues a varied course through the scarplands of Lorraine, crosses the Ardennes in a valley cut transversely to the structure, turns at right angles along the "coal furrow" of southern Belgium, and then in a sweeping curve flows across the plain of the southern Netherlands to form a joint flood plain with the lower Rhine. The Rhine is the main river of west-central Europe, 820 mi. in length, crossing the various structural and relief zones from its Alpine sources, across the northern margins of the Swiss foreland, through the rift valley between the Vosges and the Black Forest, through a well-marked gorge across the Hercynian uplands, and so to its plain course in the North sea lowlands. It receives many major tributaries, notably the Moselle which rises in the Vosges and drains much of Lorraine before crossing the Middle Rhine highlands to join the main river at Cohlenz, and also the Main, which zigzags through the scarplands of central Germany. Farther east the several broadly parallel systems include the Weser, the Elbe (723 mi.), the Oder and the Vistula (677 mi.), which rise in the uplands of central Europe and flow in a general northwesterly direction across the lowlands to the North or Baltic sea. Each of these rivers reveals distinct right-angle bends, the result of the Quaternary ice sheets, the margins of which lay along an east–west line so that melt waters escaping to the sea had to flow in a westerly direction, eroding distinct channels (*Urstromtäler*). When the ice sheets withdrew, the rivers occupied some sections of the east–west melt water channels between their northerly courses.

The rivers flowing to the Mediterranean are much shorter than the first group, because mountain ranges usually border or lie not far from the coast; short torrent sections give way abruptly to small flood plains. Only when depressions occurred between the folds and plateaus did a system of any size develop, notably the Ebro in Spain, the Rhône which joins the Saône in a long north–south corridor between the Massif Central and the Fore Alps, and the Po which drains the basin of northern Italy within the curve of the western Alps and the Apennines. A few rivers (Maritsa, Struma, Vardar) manage to reach the Aegean sea through the complicated mountains of the Balkan peninsula.

The longest river of peninsular Europe is the Danube, which

rises on the eastern slopes of the Black Forest and flows for 1,776 mi. to the Black sea. After crossing the plateau of southern Germany, the river enters the plains of the Little and Great Alföld, the latter occupying a large structural basin. The river leaves its southeastern corner in a series of gorges between the Carpathian and Balkan mountains, and flows eastward toward its large delta. As the main drainage artery of southeastern Europe, the Danube receives many tributaries: the Inn and the Drava from the eastern Alps, the Sava and the Morava from the uplands of Yugoslavia, the other Morava from the Bohemian plateau, and the Tisza from the Carpathians.

Several large rivers cross the plains of Russia toward the Black sea, notably the Dniester, the Dnieper (1,420 mi.) and the Don. The Volga, with an over-all length of 2,293 mi. has a basin that occupies nearly half of European Russia; some of its head streams rise to the west and northwest of Moscow, while others come from the flanks of the Urals. The rivers make a number of large bends; as a result, the Don and the Volga are only 40 mi. apart near Volgograd (Stalingrad).

Europe has a considerable number of lakes, many located in the north on the uneven morainic surface of the lowlands, including Ladoga (7,000 sq.mi.), Onega (3,819 sq.mi.), Vänern and Vättern. Behind Stockholm the irregularly shaped Lake Mälaren extends inland for 80 mi. A vast number of lakes covers much of Finland, the largest being Saimaa (1,699 sq.mi.), others lie on the morainic cover behind the Baltic coast in northern Poland and eastern Germany. Other large lakes occur on the forelands flanking the Alps, particularly on the Swiss plateau (Geneva, Constance, Neuchâtel, Zürich), and in the southern Alpine valleys, where Maggiore, Como, Garda and others occupy glacially overdeepened structural troughs, usually with enclosing morainic dams. Lake Balaton in Hungary (232 sq.mi.) averages 5 mi. in width but scarcely over 30 ft. deep, on the flanks of the ridge of the Bakony mountains. Many smaller sheets of water include the cirque lakes in the glaciated mountains, the crater lakes of the Apennines, and numerous coastal lakes (in the Landes, the Rhône delta and the Danube delta).

4. Climate.—The climate of Europe is the result of the interaction of three main factors. The first is latitude, which determines both length of day and night during the various seasons, and also the angle at which the sun's rays fall on the surface. Thus a place on the 50° N. parallel has 16½ hours of daylight on June 21, with a midday sun altitude of 63½°. The second is the position of Europe relative to other land masses and oceans, and therefore its exposure to air masses of markedly different temperature and humidity. The third is relief, the location and alignment of the uplands. While these occur at intervals near the western seaboard of Europe, there is no continuous trend of ranges parallel and close to the west coast (as in the Americas), and air masses from the Atlantic can move into the heart of the continent. In central Europe especially, with its diverse relief, climatic conditions reveal great local variations.

In July the sea level isotherms trend more or less from west to east, though bending slightly northward over the land. The July mean temperatures of northern Scandinavia are just above 50° F., while the lands adjoining the Mediterranean have temperatures of between 70° and 80°. In winter the western coastal lands are affected by relatively warm air masses moving from over the Atlantic, and so experience mild conditions; none of the Atlantic and North sea coasts are icebound, although the North cape lies well within the Arctic circle. Brest, in northwestern France, has a January mean of 45° and frosts are rare. Temperatures in winter become lower with increasing distance east, and the isotherms trend from north to south. Thus Berlin has a January mean of 30°, Warsaw 26° and Kiev 21°. The seasonal range likewise increases; at Brest this is only 20° and the climate is quite equable, at Berlin it is 36° and at Kiev 46°. However, even the last, compared with the extreme conditions of central Asia (with mean ranges of over 110°), is not undue, and most of peninsular Europe merits the terms temperate and equable.

Three major pressure systems exercise control on the climate: (1) the Azores high pressure, (2) the winter Asiatic high pressure and (3) the Icelandic low pressure. The first is best developed in summer, when it is centred around latitude 35° N. It sometimes extends far to the northeast, affecting parts of western Europe and southern Britain, and bringing dry, warm and sunny conditions. In summer an extension occupies the Mediterranean basin, providing an area of light northeasterly winds or calms, with virtually no rainfall.

In winter pressure is very high over the Asiatic land mass, as the result of the intense cold, and from this high-pressure system a "tongue" extends west across central Europe. Its effects may extend still farther over western Europe and the British Isles, bringing a period of intense cold, sometimes with gray overcast conditions and bitter northeast winds, at other times with clear frosty weather.

The Icelandic low-pressure system is centred around that island throughout the year, though pressure is more continuously low in winter. Numerous local low-pressure systems move westward, more frequently and more intensely in winter, when some pass northeastward toward Scandinavia, others eastward through the Mediterranean to the south of the European high pressure. In summer their paths are farther north, not affecting the Mediterranean, and they are generally less frequent and weaker. The rapidly changing conditions of temperature, humidity and winds associated with the passage of various fronts produce sequences of variable weather. When the depressions are intense, gales may be experienced along the western seaboard. Between the depressions, local slow-moving high-pressure systems (anticyclones) bring periods of fine weather of varying duration.

The climates of Europe therefore can be regarded as the sum total of the effects of air masses arriving from different source-regions. There are four main types: (1) tropical maritime air from the Azores (warm, humid), (2) tropical continental from the Sahara (warm, dry), (3) polar maritime from the arctic and North Atlantic, the most common in western Europe (cool, humid), and (4) polar continental from Central Asia (cold, dry).

Precipitation.—The depressions from the Atlantic are associated with moisture-laden air streams from the southwest, and as a result precipitation is heavy on western coasts, particularly where backed by uplands. These upland areas, and also the high mountain ranges farther inland, have more than 60 in. of precipitation each year, some in the form of snow. Places in Switzerland may receive 90 in. though other areas in the rain shadow of these uplands may have little more than 20 in. The drier parts of the Mediterranean lands and much of northern Sweden have 10 to 20 in. Most of the rest of the continent receives 20 to 30 in., and it is the only continent to have no extensive area of marked aridity, except in the neighbourhood of the Caspian sea, and in places along the southeastern coast of Spain.

The seasonal distribution of precipitation is an important characteristic of the various climatic regimes. In western Europe it is fairly well distributed, though usually with a tendency to a winter maximum and a spring minimum, with sometimes a secondary maximum in August. Farther east, as the rainfall total decreases, the summer maximum becomes more marked, the result of an increasing amount of convectional rain from summer thunderstorms. In southern Europe precipitation is concentrated almost entirely during the autumn and winter months, when depressions moving westward through, sometimes forming in, the Mediterranean are most active. The total diminishes eastward, as shown by Gibraltar (36 in.), Palermo (25 in.) and Athens (16 in.), and the summer drought becomes progressively more complete. The mountainous peninsulas and islands introduce much local variation, *e.g.*, Corsica has 60 in. in the central mountains, much in the form of snow.

Climatic Regions.—On these bases, the climate of Europe can be divided into six types, though merging one into another. (1) Arctic, with long cold winter, temperatures 20° F. or less, short summer with means less than 50°, precipitation about 25 in., much in the form of snow. (2) Cold temperate, with cold winters, not so prolonged as in the arctic type, summer means 50° to 60°, still quite short, precipitation about 25 in., well distributed, much in the form of snow. (3) Cool temperate maritime, well-distributed precipitation of about 30 in., higher in the uplands, equable tempera-

tures, much cloud and frequent strong winds. (4) Cool temperate transition, with colder winters than the third type, warmer summers, rainfall decreasing eastward and becoming more concentrated in early summer. (5) Cool temperate continental, with markedly cold winters, warmer summers, rainfall about 20 in., with distinct early summer maximum. (6) Mediterranean, with hot dry sunny summers, mild winters with some rain, sometimes an autumn maximum, with much variation according to position and aspect.

It must be stressed that there is great variability within these generalizations, particularly as a result of position and relief. For example, the Zugspitze observatory at 9,720 ft. in the Bavarian Alps has a January mean of 11° and an August mean of 35°, while the observatory that stood on top of Mont Blanc for a few years once recorded −45°. (*See* also CLIMATE AND CLIMATOLOGY.)

5. Vegetation.—Europe has been settled by man for so long and is so densely populated that the natural vegetation has been almost entirely removed, to be replaced either by cultivated plants or by a plant cover that is wild in the sense that it has not been planted by man, yet owes its existence to his influence. A series of vegetation zones can, however, be distinguished which depend largely on climate, though relief and soils make considerable modifications to any latitudinal zonal pattern.

Tundra.—The "cold desert" of northern Norway, Sweden, Finland and Iceland, and also the high *fjeld*, is the result of long winter cold, a permanently frozen subsoil and a short summer, with a brief though concentrated period of plant activity. Snow covered in winter, in summer the tundra has a close vegetation of low shrubs (crowberry and bilberry) rising from a carpet of sphagnum moss, liverworts and lichens, with an occasional dwarf arctic birch (*Betula nana*) or pygmy willow (*Salix herbacea*). Patches of sheathed cottonsedge (*Eriophorum vaginatum*) grow in badly drained areas.

Coniferous Forests.—These occur extensively in high latitudes, north of 60°, covering Scandinavia (except for the high plateau), much of Finland and extending into Russia. There are not many varieties of trees, and they grow in large uniform stands. The chief types are Scots pine (*Pinus sylvestris*), spruce (*Picea abies*) and European larch (*Larix decidua*, not an evergreen). They have stiff needle-shaped leaves, with a conical tree profile. There are a few hardy deciduous trees, notably European white birch (*Betula pendula*). The undergrowth is thin, partly because the thick layers of pine needles do not rot easily, partly because of the close growth and thick shade of the trees. The forests fade out northward into the tundra, the trees becoming more stunted and isolated, until only an occasional birch can be seen. Coniferous forests also occur in more southerly latitudes, mostly deliberately planted, on the uplands (Alps, Jura, Carpathians, Ardennes), on sandy coasts (the Landes of southwestern France, along the coasts of Belgium, Germany and Denmark) and in the heathlands.

Mixed Deciduous and Coniferous Forests.—This vegetation once covered most of the lowlands of western Europe. Deciduous trees begin to appear in the southern parts of the coniferous forests and ultimately replace them, except in the uplands. The low winter temperatures result in annual leaf shed, but the summer is long enough for leaf growth. In some areas species are mixed, including elm, sycamore, hornbeam and others, elsewhere are continuous stands of oak, beech or an oak–beech mixture. The present woodlands are only the vestiges of the former cover, removed by two millennia of clearance for timber, fuel, cultivation and settlement.

Heathland and Moorland.—Heathland occurs on coarse sandy and gravelly soils. Probably these areas were once under mixed forest, but when cleared it was difficult for the trees to re-establish themselves, and they were replaced by heath plants which flourish in the dry, acid soils developed on the sands and gravels. Sheets of Scotch heather or ling (*Calluna vulgaris*), the dominant species, are widespread, together with bilberry or European whortleberry (*Vaccinium myrtillus*) and dwarf gorse (*Ulex nanus*) and a ground layer of mosses and lichens, notably *Cladonia* species.

On sunny slopes the purple bell or twisted heather (*Erica cinerea*) appears, the crossleaved heath (*Erica tetralix*) grows in damper depressions, and other shrubs with a tolerance of high soil acidity, some degree of adaptation to periods of drought and a low demand on plant nutrients include broom (*Cytisus scoparius*), dwarf whins or needle and silkyleaf woadwaxens (*Genista anglica* and *G. pilosa*), common juniper (*Juniperus communis*) and gorse (*Ulex europaeus*). The bracken fern (*Pteridium aquilinum*) is not a true heath plant but covers parts of the heathlands, and thickets of brambles are common. In places the heath plants are replaced by grasses (notably *Deschampsia flexuosa*, *Agrostis* species, etc.), and by silver birch and dwarf oak growing sporadically, occasionally continuous enough to be called oak–birch heath.

Moorlands and bogs occur on upland areas of hard rocks and thin acid soils where drainage is poor, as in Brittany, the Massif Central of France, the Ardennes and the Rhine highlands. Bogmoss (*Sphagnum*), cotton grass, various moor grasses and reeds (*Carex* and *Phragmites* species) are found and peat accumulates in shallow depressions. Parts of both heathland and moorland have been improved by drainage, the application of lime and fertilizers and the careful sowing of suitable grasses, or by the planting of coniferous trees.

Mediterranean Vegetation.—The lands bordering the Mediterranean sea were once covered with open woodland of holly oaks (*Quercus ilex*), with a few species of pine. The marked summer drought makes it necessary for trees to have protection and adaptation in the form of deep roots to reach the ground water, and also thick bark and glossy or waxy leaves to restrict loss of water by evapotranspiration. Much of the woodland has disappeared, except for carefully fostered plantations of conifers in parts of Yugoslavia, Greece, Italy and Corsica, notably the Corsican pine (*Pinus nigra var. Poiretiana*), the cluster pine (*P. pinaster*) and Italian stone pine (*P. pinea*). Deciduous trees include in favoured places the evergreen oak, the valonia oak (*Quercus aegilops*), the cork oak (*Q. suber*), the European chestnut (*Castanea sativa*) especially in southern France and Corsica, and occasional eucalypts and tamarisks. Aleppo pine (*P. halepensis*) occurs on the lowlands, as round the coasts of Greece. On some of the higher areas in Yugoslavia and Greece firs are dominant (notably *Abies alba* and *A. cephalonica*).

Much of the former woodland has been replaced by low shrub vegetation, known as *maquis* or *macchia*, varying from nearly impenetrable evergreen thickets to arid scrub. This includes myrtle, laurel, oleander, rosemary, juniper, tree heaths and a profusion of creepers, thorns and bulbous plants. In limestone areas with a thin soil an even poorer vegetation, *garrigue*, may develop; this includes lavender, sage, prostrate prickly plants and tuberous perennials, separated by much bare rock. On some of the uplands of southern Spain, Corsica, the Apennines and the Greek mountains, pastures occur, their early growth nurtured by the melting snows.

Steppe Vegetation.—To the east and southeast of the mixed forest precipitation decreases and is concentrated in early summer. From the plain of the middle Danube eastward into the Ukraine extends a vegetation cover in which grasses are characteristic. These grasslands are known as the *pusztas* in Hungary, in Russia as the steppes. The tough grass (*Stipa*) is bluish-green and variegated with flowering plants in spring and early summer, but becomes yellow and strawlike in late summer. Little of this temperate grassland survives, but has been replaced by cereals or pasture.

Mountain Vegetation.—Uplands cover large parts of Europe and interrupt any zonal pattern. Their vegetation varies with situation, altitude, steepness and aspect, and includes conifers on steep valley sides, pastures such as the *alps* on upper slopes and benches, upland heaths, moors and bogs. On the highest peaks, below the permanent snow line, numerous arctic-alpine plants can be found in variety; they include mosses, lichens and small prostrate plants, some of considerable rarity and limited occurrence.

Maritime Vegetation.—A variety of vegetation groups is associated with the coastal margins. These include the salt marsh communities, occurring in zones from mud flats near low spring tides, the sand dune communities (usually dominated by *Ammophila arenaria*, marram or beach grass) and the vegetation of shingle beaches. The development of this vegetation is a major

factor in the natural reclamation of lowland coasts, especially those of western Europe from Flanders to Denmark.

Marshland.—Certain areas of Europe are (or were) permanently or temporarily waterlogged and carry a vegetation of reeds, rushes, tussock grasses and such shrubs as sallow, osier, willow and alder. In many cases peat is still accumulating, forming peat fens, in other cases it has developed under milder moister (Atlantic) conditions in the past. The inland areas of marsh are freshwater, usually along the coast they are saline. The largest areas are in eastern Europe on the waterlogged ground moraine (the Pripet marshes), in the broad alluvium-filled lower valleys of the rivers flowing to the Baltic and North seas, along parts of the North sea coast especially in Flanders and the Netherlands, around some of the large lakes (*e.g.*, the southeastern end of Lake Geneva), in the river deltas (Rhône, Danube) and at intervals along the Mediterranean coasts (Pontine marshes of western Italy, eastern Corsica, Greece). Many of these former marshes have been reclaimed, since they afford level land, with soils derived from fertile silt or from base-rich peats.

Aspects of plant distribution are also discussed in PLANTS AND PLANT SCIENCE and PALEOBOTANY. (F. J. M.)

6. Animal Life.—Zoogeographically, Europe is the western part of the Palearctic region, which extends eastward to Japan and southward to the Himalayas.

Mammals.—Its wild mammals are closely related to those of most of the region, but their numbers have been considerably modified by man's close settlement of much of Europe. In addition, the early isolation of the British Isles from continental Europe resulted in some European forms failing to reach Britain. Over the whole of Europe 7 orders of mammals are represented by 79 genera and about 140 species. Insectivora (hedgehogs, moles and shrews) include 8 genera and 16 species. The Chiroptera contain 13 genera and 28 species of bats. The sole Primate is the Barbary ape of Gibraltar, probably introduced from north Africa. Three hares and the ubiquitous rabbit constitute the Duplicidentata.

The Rodentia are represented by 30 genera and 49 species, including the squirrel, marmot, beaver (now rare, and long extinct in Britain), porcupine and many voles and lemmings, as well as the introduced rat and house mouse from Asia. The 12 genera and 25 species of Carnivora include survivals such as wolf, brown bear, lynx and wolverine, while fox, badger, otter and smaller predators are still plentiful. Among Artiodactyla, with 12 genera and 13 species, the wild boar is still plentiful, red, roe and fallow deer are widespread and elk and reindeer survive in northern regions. In some alpine regions chamois, ibex and mouflon are still found. A few small herds of the European bison are protected in reserves. (RO. W. H.)

Birds.—About 420 species of birds breed in Europe and approximately 360 of these also nest in Asia. By contrast, only 128 of them are found in North America and Greenland. The affinities are thus with Asia and the rest of the Palearctic. Pursuing this comparison further, only 75 species of Europe's birds breed in Africa south of the Sahara, only 20 in South America and only 25 in Australasia.

The avifauna is typical of a north temperate zone. Such cosmopolitan groups as the birds of prey and the large water birds are well represented. There are 4 species of vultures, 10 eagles and 25 other hawks, 13 owls, 22 gulls and terns and 14 herons and heronlike species. The tundra and marshes of the north provide habitats for many ducks, geese and swans, and waders; each of these two groups is represented by nearly 40 species. Turning to the land birds, there are 9 species of woodpeckers, 8 larks, 12 crows, 11 titmice, 21 thrushes and chats, 31 finches and buntings and 35 warblers (the last belong to a different family from the North American warblers). Most of the warblers and a number of other species, more than 100 in all, leave Europe completely during the winter; many are insectivorous and most go to Africa.

Big colonies include flamingos in southern France, pelicans in Rumania, cattle egrets in Spain, spoonbills in the Netherlands, and auks and gannets on the cliffs of Britain. With human population and land development, the more primitive groups—birds of prey, herons, geese—tend to decline, while other species increase because they have adapted themselves to these conditions. Perhaps the most striking example is the collared dove, which in the 20th century has spread northwest from the Balkans to colonize even Scandinavia and Britain. (I. J. F.-L.)

Reptiles.—European reptiles represent 107 species and comprise snakes, tortoises, turtles and lizards, but the number of species of all but the lizards is small. Of the 12 species of venomous snakes only the adder occurs throughout Europe. Low summer temperatures restrict the distribution of reptiles but the adder and the common lizard in bringing forth their young alive have become adapted to withstand severe conditions and they are the only reptiles that extend well within the Arctic circle. The European pond tortoise, extinct in Scandinavia and Great Britain, is confined to more southern areas and the only Chelonians that extend far northward are the marine turtles. (A. G. C. G.)

Fish.—The fisheries of western Europe and the Mediterranean countries are based mainly on marine forms, among which the cod and herring families are outstanding and flatfishes are important. In east and central Europe fresh-water fishes are favoured for food and some, especially the carp, are cultivated. L. S. Berg (*Zoogeographica*, vol. 1, pp. 107–208, 1932) has ably reviewed the zoogeography of the fresh-water fishes. Salmonidae (salmon, trout, char, whitefish and grayling) predominate in the north; Cyprinidae (carp, roach, bream, barbel, etc.) farther south. The salmon migrates to the sea (in its second year in the southern and fourth year in the northern part of its range) and there feeds and grows before returning to spawn. Trout, char and whitefish include both migratory and resident populations, sometimes both in the same river system; the lakes of the northern alpine slopes contain resident populations of all three. Trout are found about the Mediterranean, but few migrate to this sea. The Danube fauna includes the huchen, the wels and other species showing its eastern relationships. Freshwater eels are everywhere abundant although their breeding place is in the western Atlantic. (*See* also *Fisheries* below.) (ET. T.)

Insects and Other Invertebrates.—Europe has peculiar invertebrate fauna which is very different from that of tropical countries. It is a much younger fauna, having been formed toward the end of the Cretaceous period when Europe, previously under marine conditions, became finally continental. At that time great multitudes of insects and other land invertebrates invaded Europe from their old habitation in the valley of the Angara, in central Asia, and bred there. Flowering plants had appeared in Europe in the Cretaceous period and were followed by a huge train of plant-eating, flower-loving and honey-carrying insects which, in their turn, were preyed upon by parasites and predators such as carnivorous beetles, various spiders and a variety of small animals. The fallen leaves of deciduous trees in the forests had produced accumulations of humus where a rich soil fauna developed. The high mountains that had been thrown up all over central Europe offered new conditions of life to the invertebrates that enabled them to multiply in number and form. Thus Europe was populated by swarms of invertebrates belonging to the same stock as those of the Angara valley. Carabidae (ground beetles) are a classic example of insects showing Angaran ancestry.

Invertebrates of other stocks came to Europe from the Americas by way of the North Atlantic lands attracted by the hot climate at the beginning of the Tertiary period. They left fossil traces in the amber of the Baltic. Many of these are now extinct, others were driven to the Mediterranean regions by the increasing harshness of the climate as the glacial periods set in.

These changes of climate during the Tertiary period profoundly influenced the invertebrate fauna of Europe. They also contributed greatly to the populating of the caves and underground habitats. The southern countries of Europe have a cave fauna whose richness is unequalled in any part of the world.

See also ANIMALS, DISTRIBUTION OF; ZOOGEOGRAPHY; ZOOLOGY; PALEONTOLOGY. (R. G. J.)

II. NATURAL RESOURCES

1. Soils.—A broad zonal pattern of soil types may be distin-

guished though many centuries of intensive utilization by man has produced modifications. In the extreme north of Norway, Sweden, Iceland, Finland and the U.S.S.R. occur the tundra soils, the result of low annual temperatures, a permanently frozen subsoil, a slow rate of evaporation and a brief plant growing season. Waterlogging is widespread in summer. There is a surface layer of peat derived from mosses and lichens, then a horizon of mud (frozen in winter) with a bluish tinge and below this the zone of permafrost. South of this, extending east from the Baltic sea, is an extensive zone of podsols, the result of a cool damp climate and a vegetation cover of coniferous trees. There is a thin surface layer of raw acid humus, then a horizon of ash-coloured soils through which water constantly passes, resulting in a heavy leaching of bases and sesquioxides, possibly with an underlying hardpan.

Over the steppelands, from Hungary through the Rumanian coastal lands into the Ukraine, is found the chernozem, or black earth. These soils, with a deep humus layer, a crumbly texture and a natural richness in plant foods, form the great cereal lands. A variation of the chernozem soils, modified by greater aridity, is known as the chestnut-coloured soils, found in the areas of sparse dry steppe with only 8–10 in. of rainfall. They are loose and friable, contain humus formed from the coarse grass vegetation and are little leached. The upper horizon has a distinctive dark-brown colour, becoming paler with depth. They have a wide distribution over the dry steppelands to the south of the chernozem in Russia, and extending westward into Rumania and Hungary. Formed under still more arid conditions are the gray steppe soils, or serozem, with a low organic content because of the scanty vegetation, and which have suffered little leaching. They are rich in plant foods and where irrigated, very productive.

Much of the lowlands of western Europe are covered with brown earths, typical of the area occupied by the former deciduous woodlands of western Europe. They are rich in organic matter derived from the accumulation of leaves. The surface soil is usually slightly acid, since the climate is humid and some leaching of bases, though not of sesquioxides, takes place. Most of these soils have been cultivated for centuries, the original forest cover has long been removed, and constant liming and manuring are now necessary. These brown earths vary much in detail, since they have developed on a wide range of parent materials.

Although this broad pattern of zonal soils is valid, it is obvious that with the varied relief and rock types many intrazonal and azonal (skeletal) soils may be distinguished. On limestone parent materials, *rendzina* soils and terra rossa have developed, the latter under rather more arid conditions. The *rendzina* soils are dark coloured with a surface horizon of friable almost granular loam lying on a subsoil containing limestone fragments, which in turn rests on the solid limestone. Terra rossa is the name applied somewhat loosely to a range of reddish residual soils that have accumulated in limestone depressions, under conditions of prolonged summer drought, as in the Mediterranean lands. In parts of Spain, the Causses of southern France, southern Italy and Sicily, the karst (*q.v.*) of Yugoslavia and northern Greece, the steep slopes, concentrated winter rains and rapid runoff result in the washing down of the indissoluble residue into crevices and depressions. A varied group of peat soils is characterized by the accumulation of partially decayed vegetation. Fen peat soils occur where peat has formed in an area of calcareous ground water that neutralizes the humid acids liberated during plant decay and forms a thick black spongy mass of "mild peat." When drained it affords rich soils as in the English fens and in the *laagveen* of the Netherlands. Bog peat soils are formed where much acid peat has accumulated, as in the moorlands on the uplands of central Europe. Dry peat soils, containing only a little acid humus derived from mosses and lichens, occur on the heathlands of the North European plain, developed on heavily leached sands.

Various azonal soils are slowly developing from new soil-forming material. Some materials have formed a medium ready for cultivation. These include alluvial soils, derived from river-borne sand, silt and clay; when drained they provide valuable farmland, as in the flood plains and lower courses of the rivers of western Europe. Marine soils are reclaimed from the sea, naturally or by man; along the coasts of Belgium, the Netherlands, Germany and Denmark areas of former sea are enclosed by dykes and drained (polders), providing soils developed on marine clays. Marine sands form extensive areas of poor sandy soil, as in the Landes of southwestern France and in northern Denmark, sometimes underlain with a hardpan resulting in waterlogging. In parts the surface is so unstable that moving dunes occur. Glacial soils are widespread on the North European plain, formed on the sheets of ground-morainic boulder clay ("younger drift") in Poland and northern Germany. Fluvioglacial outwash sands and gravels cover much of northern Germany, the eastern Netherlands and northeastern Belgium, on which poor acid heath soils occur. One of the best parent materials is loess or *limon*, probably deposited by the wind during dry interglacial and postglacial periods from the unconsolidated material laid down by continental ice sheets. It is found in the narrow Börde, an area of rolling country of central Germany (to the north of the uplands of central Europe), in the valleys of such rivers as the Oder, Elbe and Rhine, on the low plateaus of central Belgium and in eastern and northeastern France. It develops fine textured, easily worked soils, deep and well drained, thus giving a good tillage bed for wheat and sugar beet. Volcanic material weathers easily and when washed down to lower slopes provides a soil of great fertility, as around the slopes of Vesuvius and Etna. On all the higher mountains scree soils occur, consisting to a large extent of rock fragments. Finer material is moved downhill by soil wash and soil creep, forming better soils on lower slopes. (*See* also SOIL.)

2. Minerals.—Europe is rich in reserves of coal, iron ore, bauxite, common salt and potash, but (with some exceptions in European Russia) deposits of nonferrous mineral ores are quite inadequate and much has to be imported.

Europe's main coal basins are strung out along the northern margins of central Europe. Apart from the British fields, the line begins in northern France in the *départements* of Nord and Pas-de-Calais, continuing along the Sambre–Meuse valley in southern Belgium. Another field farther north lies partly in Belgium (the Kempen field) and partly in the Netherlands (south Limburg). Then after the small Aachen field in Germany near the Belgian frontier comes the Ruhr or Westphalian field, Europe's largest; it produces about 115,000,000 tons per annum, and its reserves are such that it can maintain this rate of production (without mining below 3,280 ft. or 1,000 m.) for six centuries. To the southwest of the Rhine highlands occurs the Saar basin, continued across the French frontier as the Lorraine or Moselle field. Farther east small fields occur on either side of the Ore mountains in Saxony (East Germany) and in Bohemia (Czechoslovakia). In southwestern Poland is the large Upper Silesian field, formerly German, in the 1960s producing about 95,000,000 tons per annum. In the south of the Ukraine is the Donets basin (Donbass), which produces about 100,000,000 tons annually, or about a quarter of the U.S.S.R. output. A few small fields are found in northern Spain, the Massif Central of France, Hungary and Yugoslavia. Lignite is mined in southern France, Austria and elsewhere, to be used chiefly for the production of thermal electricity. Brown coal (an earthy fibrous fuel) is quarried in the Rhineland, Saxony and Czechoslovakia, and is briquetted for use as power station fuel and in the chemical industry. For the first decade after 1945, Europe's chief industrial problem was coal shortage; by the early 1960s increased production, plus the competition of alternative sources of energy, had resulted in overproduction, enforced stockpiling and restriction of output. Numerous uneconomic mines had been closed, notably in southern Belgium.

The main iron ore fields are in the Ukraine at Krivoi Rog, the Crimea, northern and central Sweden, Lorraine (mainly in France, but extending into Luxembourg and just into Belgium), northern Spain, the Siegerland of west Germany and Austria. The ore is worked in smaller quantities in many other parts, often forming the original basis of local steelmaking. The chief European source of bauxite is in southeastern France, and considerable quantities are worked in Yugoslavia and Hungary. The nonferrous minerals are found sporadically, usually on a small scale, in Spain, Yugoslavia, Sweden and Finland. The largest deposits are in Russia,

notably manganese near Nikopol in the Ukraine and a wide variety in the Ural mountains. The bases of the chemical industry (apart from byproducts of oil refining and coke making) are the common salt deposits of Lorraine and the potash of Saxony and Alsace. Petroleum resources are small compared with Europe's needs (90% of which are imported, mostly from the middle east). The chief fields are on the flanks of the Carpathians, in Rumania, in northeastern Austria, in west Germany on the northern edge of the uplands, in the Netherlands and in southwestern France. A large amount of natural gas is obtained in the last area, and is piped as far away as Paris. Another large-scale source of natural gas is the plain of northern Italy. Other raw materials, such as lime for cement and clay for bricks, occur widely. (*See* also articles on individual minerals for occurrence and production.)

3. Water Resources.—As a continent Europe is well supplied with water and only in the south is there ever a serious shortage. The chief uses of water resources are to provide irrigation, for electric power generation and to supply the industrial and domestic needs of the large urban areas.

Irrigation.—In the areas of summer drought irrigation has been practised from early times, and much agriculture in the 20th century is still dependent on it. In Spain it is chiefly confined to areas along the major rivers where, by a system of canals and aqueducts, considerable areas are watered. The Ebro is thus utilized in Aragon (around Saragossa, Plasencia and Lérida) and in Catalonia, while on the Meseta a large area is irrigated around Valladolid. A water tribunal still meets at Valencia to decide water rights.

In Italy by the mid-20th century as much as 4,500,000 ac. had been irrigated, chiefly as part of large Alpine schemes using the waters of the Italian lakes. Smaller-scale well irrigation follows the spring line at the foot of the Alps and the Apennines. Several of the irrigation canals are old, the one between Adda and Ticino (Naviglio Grande) dating from the 12th century. In Sardinia the dammed Tirso river waters approximately 80 sq.mi.

Hydroelectric Power.—By the early 1960s the estimated hydroelectric power potential in Europe was 658,150,000,000 kw.hr. per year. The chief sources are the Alps, the Pyrenees and Scandinavia, where hydroelectric power is a major contribution to the national income. In the Alps there are two types of projects, those on the fast-flowing Alpine streams and those on the major slower rivers. In addition dams were built to form large reservoirs. On the Rhine there are plants at Ryburg-Schwörstadt, Albbruck-Dogern, Lanfenbourg, Schaffhausen, Kembs and Ottmarsheim and on the Rhône at Verbois and Génissiat, Lavey, Jonage, Donzère-Mondragon and Montélimar. Hydroelectric plants are also associated with the irrigation reserves of the Italian lakes. Smaller plants are distributed throughout the Alps, particularly in Valais where Grande Dixence dam was completed in 1961, and in the Bernese Oberland.

The Pyrenees are another potential source of hydroelectric power. Both the Ariège and the Garonne have been harnessed on the French side while on the Spanish side the Ebro and its tributaries provide the greatest potential. Also of importance in the Iberian peninsula are the Tagus and Guadalquivir.

Approximately one-third of the Scandinavian potential is developed, and electricity is exported to Denmark by a cable under The Sound. The chief stations in Sweden are at Harspränget falls on the Lule, and Trollhätten falls on the Göta, in Norway in the Drammen valley and in Finland at Imatra. In these northerly countries freezing prevents the working of the plants for three months of the year. This explains the wide use of thermoelectricity. Other areas of potential or actual production include the Carpathians, the Massif Central of France, the Apennines and on a smaller scale, Scotland and Ireland.

Urban Uses.—A further demand in water resources is that of urban areas. Besides the careful selection of reservoirs for the drinking water supplies of large cities, most industry requires water. The Ruhr industrial region, for example, is supplied by the Ruhr, Emscher and Lippe rivers, of which the supply from the Ruhr is regulated by dams, while the Emscher and Lippe also remove the factory-polluted water. Other examples can be seen in the Pennines of England and in Lorraine.

4. Land Use.—Europe is a most important continent for agriculture. Even excluding the U.S.S.R., it produces a third of the world's wheat, two-thirds of the potatoes, more than half of the dairy produce and three-quarters of the wine. Cultivation is usually intensive, using rotation of crops and heavy fertilizing, and high yields are customary. It is estimated that Europe (excluding the U.S.S.R.) has only 4% of the world's forest area and nearly two-thirds of its timber requirements are imported.

Agriculture.—Within Europe with its varied relief, soils and climates, agriculture is correspondingly diverse. Summer temperatures and the length of the growing season determine the northern economic limits of cultivation for each crop. Even rice is grown in a few special areas in southern Europe, then in turn northward follow the limits of corn or maize (more or less defined by the 70° F. July isotherm), wheat, oats and barley; the last requires a much shorter growing season than the rest and is even grown within the Arctic circle. In southeastern Europe the winters are too cold for autumn sowing and spring wheat is grown. The incidence of frost is important; the northern limit of the olive is defined by a virtually complete frost freedom. The vine can stand cold winters, provided late frosts are not experienced in spring, and it can be grown as far north as the Rhineland and the Paris basin on south-facing slopes. Oats are rarely found in southern Europe because of high temperatures. Most of the continent receives adequate rainfall for cultivation. In the west, where the annual total exceeds 30 in., with much cloud, grasses and fodder crops tend to replace cereals. Irrigation is rarely needed except in parts of the Mediterranean lands.

The pattern of relief and soils is also reflected in the agricultural distributions. The large-scale areas of cultivation are the lowlands: the north German and Polish plains, the loam covered parts of Belgium and the Paris basin, Aquitaine, the north Italian plain, the Hungarian basisn, the Black sea coastlands and the Ukraine. Much of Europe consists of uplands, where cultivation is limited to the valley floors and lower slopes. The best soils are those derived from loess and from alluvium, also those in eastern Europe and the Ukraine, the chernozem, which contains much humus, accumulated while the lowlands were covered with natural grassland. At the other extreme are the stony or thin acid soils of the uplands, the sands and gravels of the heathlands and the thin residual soils overlying limestone. Rye is widely grown on poor soils, since it is less exacting than wheat.

Much has been done to improve the land by man's efforts: by the draining of excess water (in the Netherlands, the Italian Pontine marshes and in Greece), by irrigation (southeastern Spain), by heavy fertilizing (the sands of the north European plain and the heavy boulder clays of eastern Europe), and by the terracing of slopes to check soil erosion (in the Mediterranean lands).

There is a marked contrast between the countries of western Europe and those of eastern Europe. The latter are more thinly populated and less industrialized and mainly produce cereals, much of which are exported to the U.S.S.R. Many farmers, particularly in the western countries, specialize in activities that give a higher cash return instead of growing crops for which other parts of the world are better suited. The acreage under wheat in the 1960s was much less than a century earlier, since grain could be imported from the great areas of temperate grassland in other parts of the world; Europe imports more wheat than the rest of the world combined. Much land is under permanent or rotation grassland or grows fodder crops for dairying to meet the requirements of large urban populations. This is either the basis of the whole agricultural economy (as in Denmark) or an important item in mixed farming (as in the Netherlands and Switzerland). There is considerable specialization in market gardening and in such commodities as vines, fruit and flowers. Farmers may be required to produce crops in accordance with national needs, either encouraged by subsidies (sugar beet in western Europe) or by direction (sunflowers for vegetable oil in eastern Europe). The pastures of peacetime may be plowed up in wartime as a result of restricted import of grain.

The countries of Europe may be grouped into five broad agricul-

tural regions: northern, northwestern, central, eastern and Mediterranean.

In northern Europe agriculture in Norway, Sweden, Finland and Iceland is limited, the result of a high-latitude climate, with a large proportion of upland and poor thin soils. Much is uncultivated and under forest, tundra, icecap or bare rock. Norway, for example, has about 70% of its area unproductive, 24% forested and only 6% under meadow or arable. Barley and potatoes are grown along the narrow coastal plains of Sweden and Finland; cattle kept both there and in Iceland are stall-fed in winter.

Northwestern Europe is an area where dairying based on permanent pasture and fodder crops is dominant, with sheep in the uplands. Oats are grown in the cloudier west and rye in the east; farther south winter wheat, with various root crops (notably potatoes and sugar beet) in rotation, become important on the better loam soils. Orchards are mainly of temperate fruits, such as apples.

Central Europe has a high proportion of upland and agriculture is limited to the valleys and basins, with fodder crops, meadow grasses, small patches of cereals and vegetables, and fruit in favoured areas. South-facing slopes are favoured for orchards and vineyards (Burgundy, the Rhineland) and tobacco and hops are grown (Rhineland). Use is made of the uplands, for animals are moved to higher pastures in summer after the snow-melt (transhumance). The largest area of lowland, the north Italian plain, grows wheat, corn and even rice on the irrigated lands of the Po valley.

Eastern Europe contains the most extensive arable lands, from the middle Danube to the Ukraine. Spring wheat is dominant in the northern parts, corn in the south, with sugar beet, sunflowers and other vegetable-oil plants. Large herds of cattle and sheep are reared on the drier plains and upland slopes.

The Mediterranean lands are characterized by limited areas of lowland, with uplands rising steeply from narrow coastal plains and a climate distinguished by marked summer drought. Much of the cultivated land occurs on carefully terraced slopes. Small patches of hard wheat and barley, peas, beans and lentils (preserved by drying), olive groves, vineyards, groves of citrus and other fruits (peaches, almonds, walnuts, sweet chestnuts) are characteristic. The farmer usually has a small varied holding, but in some places specialization is practised: the vineyards of southern France and the Jerez district of Spain, the currant grapes of southern Greece, the lemon groves of Sicily, the bitter oranges of the Seville plain, and the cultivation of vegetables—the onions of the Valencia district and the *primeurs* (early vegetables) of southern France and the Naples district. The largest areas of cultivation are on the coastal plains, such as the reclaimed Pontine marshes in western Italy, and on the fertile soils around the Bay of Naples. Many schemes of land reclamation and improvement were in progress in the 1960s, notably in central and southern Italy, Sicily, Sardinia and Greece. The heat and aridity of summer limits the pastures to the uplands and few cattle are kept; sheep and goats are reared for wool and milk and mules and asses are bred for transport. Pigs are kept in herds that forage in the chestnut and beech groves of parts of Spain, Italy and Yugoslavia. (*See* also the *Agriculture* subsections in articles on the various countries.)

Forestry.—Europe's chief timber reserves consist of the coniferous forests (see *Vegetation* above), notably those around the Baltic sea and in the western U.S.S.R. A large part has been logged at least once; cutting is carefully regulated and as one area is cleared it is replanted. The importance of timber to the economy is outstanding in the case of Finland; about 65% of its area is under timber (the highest proportion of any country in the world), and sawn timber and timber products account for more than 90% of exports by value. Many other countries in Europe maintain coniferous plantations in areas otherwise unsuitable, sometimes to protect the terrain, or as a contribution to the vast amounts of soft wood, pulp and newsprint consumed. These plantations have been established on most uplands, on the sandy heaths and along the coasts where it is necessary to check dune movement. (*See* also FORESTS AND FORESTRY.)

5. Fisheries.—The peninsular character of Europe, with its long indented coast line, numerous natural harbours and broad continental shelf resulting in extensive shallow seas rich in plankton, is responsible for the importance of fishing, particularly to the countries of northwestern Europe.

The fishing grounds visited by European vessels extend in the north from Greenland to Bear Island, Spitsbergen and the Barents sea and continue as far south as the Moroccan coast. Particularly important are the waters around Iceland, off the Norwegian coast in the neighbourhood of the Lofoten Islands, the North sea (especially over the Dogger and Great Fisher banks), the seas off southwestern Ireland and the Bay of Biscay. About 40% of the catch by weight consists of herring (mainly from the North sea), 15% cod and 10% haddock.

Many other types of fish are caught in smaller quantities, though they represent a value far greater than their weight, *e.g.*, brill, sole, turbot, plaice. A large proportion of fish is caught by the fleets operating from the ports of Bergen, Aberdeen, Grimsby, Hull, Esbjerg, Scheveningen, Boulogne and Lorient; a feature of the industry has been its increasing centralization at a few large ports. The Mediterranean fisheries are much less important, but sardine, tunny and anchovy are caught, particularly in the western basin. There is much small-scale inshore fishing all around the coasts of Europe, both on a commercial basis and for subsistence (*e.g.*, off western Norway and Brittany, where it supplements agriculture).

Shellfish are important in some countries (*e.g.*, they form one-tenth of the value of the total French catch) and include lobsters, shrimps and oysters, the latter cultivated along the silt-filled estuaries of the Biscay coast at Concarneau, Audierne and Marennes. Fresh-water fisheries in rivers, lakes and artificial pools (pisciculture) are important in many countries. The sturgeon is caught in rivers flowing to the Black sea, and trout and carp are reared widely.

Industries ancillary to the fisheries are concentrated at the larger ports, including packing and canning, smoking of herring, extraction of oil and production of glue and fertilizers as by products.

6. Industries.—The countries of western Europe form one of the world's major industrial regions; only the U.S. and the U.S.S.R. can as yet compare with their output. The industrialization of Europe came at the end of the 18th century, when power was first applied on a large scale. Activity centred on the coal fields. The Industrial Revolution began in England and rapidly spread to Belgium and France; it came rather later in Germany but developed rapidly on a large scale following unification during the last decades of the 19th century. Switzerland, Italy, Norway and Sweden, without the coal fields on which industry depended, later developed specialized industries based on hydroelectric power. After 1918 two new industrial countries, Poland and Czechoslovakia, developed industries on the basis of small coal fields and on the legacy of activities inherited from Germany and Austria. Until World War II industry was concentrated in western Europe, while countries in the east and south remained largely agricultural.

After 1945, the states of eastern Europe, under Russian influence and direction, sought to stimulate industry through a series of five-year plans, and similarly Spain and Yugoslavia made much progress by carefully planned and state-sponsored development. Postwar programs of industrial recovery were carried out in western Europe to rebuild the devastated cities, ports and communications. By the early 1960s production for every country was exceeding the prewar level. In addition, the export drive, together with the need for maintaining a high level of armaments, stimulated industry. Europe's industries therefore boomed and exports reached record proportions. Economic groupings such as the European Coal and Steel Community contributed to modernization and increased efficiency. But other extra-European countries also increased manufactures and in some European industries unemployment resulted from overproduction and competition.

The main industrial regions can be located in six groups, according to the factors responsible for their location. More than one factor may operate and different factors may operate at different times:

1. The main heavy industries are on the coal fields—iron and

steel, heavy engineering and chemicals. Both large and small fields show this concentration of industry, notably the Ruhr, the Donbass, Upper Silesia, southern Belgium and northern France.

2. Many industries are located at the source of bulky raw materials, notably iron and steel (Lorraine), timber (Finland and Switzerland), salt (Lorraine), potash (Saxony), lime and clay (widespread). Wool produced on the Vosges, flax grown in Flanders and raw silk near Lyons and in northern Italy have helped make these into textile-manufacturing districts.

3. Europe imports so much of its raw materials and foodstuffs that there is much processing at the ports—grain milling, oilseed crushing, sugar refining. Since the bulk of the crude oil is imported, refineries are situated mostly at or near ports (Marseilles, Bordeaux, Seine estuary, Antwerp, Rotterdam and Hamburg). Shipbuilding, repairing and marine engineering are obvious port industries.

4. Many factories are on inland waterways, where fuel and raw materials can be cheaply brought together (*e.g.,* at Liège, Strasbourg, Rouen and Cologne).

5. Many capitals and other long-established cities have a well-developed industrial life partly because of traditions, partly because of huge local markets and partly because of available labour supplies. The largest industrial centres in Europe are Paris, Berlin, Brussels, Amsterdam and Rome; Copenhagen contains one-third of the industries of Denmark. In eastern and southern Europe industrialization came later, government stimulated and directed, and a high proportion of each country's activities is concentrated in its capital: Belgrade, Soña, Budapest, Madrid, Athens and Bucharest. Other large cities are industrial centres: Birmingham and Manchester in England; Augsburg, Nürnberg and Münich in southern Germany; Milan and Turin in Italy; Basel and Zürich in Switzerland; Lyons in France; and Barcelona in Spain.

6. The availability of hydroelectric power has developed "old" areas (power applied to original small-scale craft industries) or quite new ones; these include the north Italian plain, the Swiss plateau, the Alpine and Pyrenean valleys of France, the Norwegian fjord coasts and many parts of Sweden. Industries include aluminum refining from bauxite, production of special steels and steel alloys in electric furnaces and the manufacture of chemicals.

See also NATURAL RESOURCES and *The Economy* sections in the individual country articles. (F. J. M.)

III. ANTHROPOLOGY

The migrations and culture of the peoples of Europe are discussed in the initial sections of *History* below.

1. Physical Anthropology.—Stature increases from the Mediterranean northward to mid-Norway and Sweden, to diminish again in the arctic, but the changes are less regular, for the Dinaric peoples and the Castilian Spaniards are mostly tall. The Iberian peninsula, Sardinia, Corsica, Sicily, south Italy have longheaded populations and so have various areas in France, the British Isles and Norway and Sweden, apart from the arctic area in which the head runs very broad. The great mass of peoples of central Europe, from central France eastward to central Russia, and from Greece and north Italy to Prussia is broadheaded though less pronouncedly so in the north. Pigmentation of skin, hair and eyes diminishes from the Mediterranean northward to mid-Norway and Sweden, to increase again in the arctic.

Biologically, the upright head, resting on the backbone, has allowed reduction of the neck muscles, and hands freed to use tools have helped brain growth and consequent skull expansion. Jaw muscles still heavy in the Old Stone Age gave weight forward which might be balanced by expansion backward giving very long heads (some with breadth/length ratio, *i.e.,* cephalic index 72 or 73), or the skull might become more rounded balancing median and lateral expansion and often giving indices of 83–88. Jaw muscle reduction eased lateral expansion and many European longheads came to have indices of 77–79 rather than 72–73. These abound near the western Mediterranean, in the British Isles and in Norway and Sweden. Indices of 83–88 are a feature in central and eastern Europe. Head form seems to be associated with several genes, and the cephalic index and details range widely in most groups.

The stature increase among longheads from south to north in Europe is supposedly in response to diet and to later sex maturity in the cool north. Often 1,600 mm. (5 ft. 4 in.) or less in the western Mediterranean, they rise to 1,700 mm. (5 ft. 8 in.) in the British Isles with figures a little lower among Welsh brunets, and appreciably higher among Scottish and Scandinavian blonds. Broadheads are often moderately tall in the Dinaric region, but stocky short men are typical among central and eastern European peasants. French villagers around Limoges and in the interior of Brittany are conspicuously short but the sailors of west Brittany and parts of Normandy are taller, probably because of inheritance from Norse sea rovers. Generally, welfare has increased stature. Part of this increase is thought to be a hastening, and sexual maturity is becoming earlier (as Swedish data show), so growth may be ending earlier in the mid-20th century than it did at the end of the 19th, though later than in the 16th and 17th centuries.

Brown pigment in hair, eyes and skin is in part protective against cytolysis (cell dissolution) caused by high-violet and ultraviolet radiation. In the western Mediterranean there is a preponderance of dark hair and eyes, often with an olive-brown tinge in the skin. The British Isles have many brunets, many with gray, but some with brown eyes, but fairer complexions and blue eyes, often with blond hair, are commoner where Norse or Anglo-Saxon elements have come in. Tall blonds with blue eyes are numerous in eastern Scotland, Scandinavia and Finland. In the arctic, the sallow-skinned Lapps are said to have the skin blood vessels well below the surface and it is well known that snow glare may cause cytolysis. Tall fair longheads (Scandinavia, Scotland) are often called the "Nordic" race. Short dark longheads (western Mediterranean) are analogously called the "Mediterranean" race. Broadheads usually of intermediate colouring and shortish stature are called the "Alpine" race. These distinctions are of course highly artificial and should not be used to imply separate origins. Many British are intermediate between Nordic and Mediterranean but this need not imply that they descend from a mixture of these two groups.

Some blood groups may be associated with single genes, and the genes for A, B and O groups have been studied to some extent in correlation with the characters discussed above. It seems that the O group is in some sense fundamental and has been to various degrees replaced by A or by B. The O type is specially notable in the Basque area, *i.e.,* among a people of very ancient settlement (see *Language* below). Sardinia, also notable for extreme longheads, and Corsica, Ireland, Scotland (especially the western Highlands) and, curiously, Iceland have the O type in more than 70% of the population. It is nearly as abundant in parts of England and Wales, but it falls below 60% in the Soviet Union. The *B* gene characterizes 15%–20% of people in most parts of Russia and 20%–25% from the Ural region eastward. It is an Asian feature and, like broadheadedness, seems specially to belong to central Asia. The *B* gene occurs in 0%–5% of people among the Basques and in less than 10% in peoples west of the Soviet Union. The *A* gene may reach 30%–40% or even, locally, 50% in the west of the Iberian peninsula; in the Burgundian region, in Hungary, Scandinavia and Finland some think there is a link between the *A* gene and the gene for relative tallness. Most Scandinavian samples show 30% or more *A* and this contrasts with the 15%–20% in Iceland which according to history was largely colonized from Norway. The *O* gene shows a figure exceeding 70% in Iceland but under 65% in most of Scandinavia. Toward an interpretation one may mention that the *O* percentage is a little higher in mid-Norway where, also, the *A* percentage is a little lower than elsewhere in the peninsula. It is also known that some Irish settled in Iceland before the Norsemen and some Norse colonists went to Iceland via Ireland and apparently took some Irish wives. (*See* also BLOOD GROUPS.)

Trade and migration with intermarriage have made most European populations polymorphic so that averages rarely throw much light. In the case of the Basques there is an interesting parallelism between blood group and language (*see* BASQUE). In Switzerland the German-speakers are of rather lighter colouring, but otherwise there is little linkage between linguistic groups and physical type.

Polymorphism is illustrated in southern Italy by coastal broadheads contrasting with interior longheads while in southwest coastal Norway a broadheaded type with intermediate colouring occurs in numbers in a country generally characterized by tall blond longheads. In head form and blood groups the modern Greek population ranks mainly with eastern Europe.

2. Languages.—Basque, spoken on the northern and the southwestern sides of the Pyrenees and influencing pronunciations in Spanish and in Aquitanian patois, is thought to be the oldest surviving language in Europe.

Indo-European languages, almost universal in Europe, are believed to have taken shape first on the loess lands of the Ukraine and, perhaps, the Hungarian Danube where the soils were relatively easy to clear and cultivate in early times. Archaeology illustrates the movements of people especially westward from this area, and some of these movements in the 2nd and 3rd millenniums B.C. brought Indo-European forms to be modified by local contacts and intermarriage (*see* INDO-EUROPEANS). Celtic was spoken on the southern German loess lands soon after 1000 B.C. and spread westward to survive as Gaelic, perhaps much modified by contact with some earlier forms, in western Ireland and Scotland and until recently in the Isle of Man.

Ancient Celtic was submerged by Latin in Gaul, but Brythonic Celtic, emphasizing *p*, *b* and soft *f* sounds in contrast with Gaelic *k* and *qu*, is the national language of Wales and was taken to Brittany by refugees fleeing from Anglo-Saxon conquerors. It has died out in Cornwall (Eng.) in the last few centuries. A curious linguistic detail is that the *th* sound is used in Spain for a *z* or for a *c* followed by *e* or *i* (Valencia) and in western Jersey (Channel Islands) for an *r* (*père*). Around the Irish sea there is *th* in England, soft *dd* in Wales, grading from father to fader in western Ireland. Denmark makes much use of the *dd* sound. Spain, the Irish sea and Denmark are all on the track of the early megalith builders.

Celtic-Italic, from Danubia, ousted older languages in Italy, as well as Etruscan which was probably brought from southwest Asia after 1000 B.C. In spite of the cultural pre-eminence of classical Greece, the modern form of Greek is spoken by fewer than 10,000,000 persons. Albanian has pre-Greek elements with contributions from Greek and Latin. The Roman empire spread Latin to give birth to Italian, Portuguese, Spanish and French. German took shape south of the Baltic, and Danish, Norwegian and Swedish northwest of the Baltic. Norse, the ancestral form of these last three languages, spread to Iceland and Greenland in the mild 9th and 10th centuries A.D. but died out in Greenland in the colder 14th and 15th centuries.

North or Low German spread over the Low Countries and, under pressure from westward-moving Slavs, reached Britain as dialects of Old English. Loss of inflections and genders and rich borrowing from Latin and French, gave English great penetrating power. Early Indo-European Lithuanian (*Litus*-shore), near the southeast Baltic, has the related Lettish as its northern neighbour. Spread of iron axes led to post-Roman forest cutting in the relatively mild 7th–12th centuries and promoted Slav expansion from the Ukraine, northeast as Russian, northwest as Polish, Wendish, Czech, etc., and southwest as Serb and Bulgar (with a few non-Indo-European elements). Later growth of German organization, and severer winters, halted Slav advance and the Slavs also faced Mongols spreading from inner Asia. A counterspread of German to Transylvania (note the legend of the Pied Piper) and to Russia survived until ruined by 20th-century wars.

Provincial Latin in Roman Dacia, blended with Slavonic, has formed Rumanian which survived Bulgar and Magyar invasions, thanks to its refuge on the high plateau pastures of the Carpathians. Of non-Indo-European languages, Magyar entered Hungary via the Carpathian passes about the 9th and 10th centuries A.D. Finnish and Estonian are derived from a source akin to Magyar. Lapp is used in the arctic north. Turkish is spoken in and around Istanbul. (*See* LANGUAGE and articles on individual languages.) (H. J. F.)

3. Ethnology.—Social anthropology has been mainly concerned with the nonliterate peoples of the world. Increasingly, however, it has turned its attention not only to the peasant communities of the west and the orient but also to urban and industrial communities, where such aspects of social life as kinship patterns have been intensively studied. Originally ethnology embraced the comparative study of man's racial as well as his cultural characteristics, but the word has come to stand particularly for the study of material culture and folklore.

Ethnologists of the older school were able to see Europe as a cultural unit and they promulgated fruitful if now somewhat discredited theories of cultural evolution. In general the customs, beliefs, social structures and folk cultures of the peoples of Europe were studied and interpreted in the light of the historic linguistic groups and the hypothetical racial divisions of the continent. This approach to the subject persists in terms such as the "Celtic field system," the "Saxon house" or "Nordic culture." Though some customs are markedly local or linked with language groups, what emerges from the extensive research and field surveys of later years is the wide distribution of innumerable culture traits throughout the European continent. Revived national pride, for good or ill, has had much to do with the stimulation of ethnological research. If on the one hand it has fostered extravagant claims of priority and superiority which have at times been exploited for political ends, on the other hand it has led to the establishment of many excellent museums and research institutes. In folk music and folklore in particular, nationalism has been a driving force inspiring research and leading to cultural revivals: Finland and Ireland may be cited as examples. From Greece to Scotland and from Spain to Sweden, efforts have been made not only to record folk tales and to preserve examples of old tools and craftwork, but also to stimulate the retention or revival of traditional skills that are obsolescent.

So far as oral literature is concerned many of the folk tales of Europe, scientifically studied by Scandinavian scholars, have been shown to be international, belonging to types (catalogued by Aarne Thompson) which occur widely in India. Some idea of the wealth of material that has been collected is revealed in the number of international folk tales (*märchen*) in the archives of the Irish Folklore commission, which approaches 25,000, nearly all of them in Irish. On the other hand the shorter prose narratives, sagas and legends frequently relate to particular regions and national events. As an aid to their interpretation, and for folk life studies in general, scholars in many European countries tend to regard the continent as falling into broad ecological zones which, though they have been in contact with each other since Neolithic times, display characteristic forms of social life and customs which can be related to environment and regional evolution.

Mediterranean Zone.—The lands surrounding the Mediterranean sea have a long tradition of urban life associated with the cultivation of fruit (especially the olive) and, in many areas, with a rich development of the lore and crafts of maritime trade and fishing. In Spain, following J. C. Baroja's lead in urban studies, J. A. Pitt-Rivers in *The People of the Sierra* has analyzed the social structure of a small southern town, Alcalá, which will stand in many ways for the typical Mediterranean town. He describes the pueblo as highly centralized and the town is regarded as the natural unit of society, the home of agricultural as well as other workers. The Mediterranean conception of community is firmly based on locality, on urban centres with their local patron saints, and with social values that are as old as the civilizations of the ancient Aegean. A strong sentiment of local patriotism is expressed in hostility to rival towns, to whose inhabitants nicknames are customarily applied.

The festival of the local saint is an annual holiday marked by colourful processions and is often associated with markets and fairs. These festivals combine with religious zeal much that is frankly pagan and finds expression in popular ballads and dances, with dialects and flavours that are strongly localized. There is a similar variety in material culture, in tools and costumes, but the spirit is everywhere essentially Mediterranean. Houses are built of stone, and fields tend to be square in shape and are cultivated with the light wooden plow, but there is not the peasant attach-

ment to the soil as there is north of the Alpine mountain belt. Urban life claims the highest values. In these lands of summer drought sheep and goats are more at home than cattle, and they are characteristically moved to mountain pastures in summer. This practice of transhumance has left its mark on the social life and on the crafts of southern Europe. Spinning, weaving and embroidery are well-developed arts, notably in the Balkan peninsula and the Aegean Islands.

Alpine Zone.—The forested mountain belt which extends from the Pyrenees to the Carpathians is another ecological zone where occupations, social habits, customs, folklore and material culture are strongly influenced by environment. There are found timber houses of various styles often set with one gable embedded in the hillslope and providing, under a single overhanging roof, shelter for men, beasts and harvests—the Basque house is a variant of this style of building. Transhumance, usually involving cattle, is long established and persistent, though in the drier eastern mountains (*e.g.*, the Transylvanian Alps) sheep are involved, with very different social and cultural consequences. Settlement is in small units, typically located in mountain valleys where isolation brings independence but limits social life and where inbreeding causes special problems.

In addition to such pastoral arts and crafts as horn blowing, weaving and cheesemaking, woodcarving is highly developed and from being a winter occupation has been commercialized in various forms. Folk art and music, dancing and costumes tend increasingly to be consciously cultivated, for reasons that are partly cultural and partly economic.

Mid-European Zone.—The great plain and fringing lowlands of mid-Europe from northern France to Poland, with outliers in lowland England, southern Sweden, Czechoslovakia and Austria, have developed their rural life against the background of the deciduous forests. The typical tree of the deep glacial and periglacial loams of this zone is the oak, which has been utilized since prehistoric times in many craft industries and in the half-timbered house which was the ancestor of the North American frame house. One variant, the so-called Saxon house, was formerly widely distributed over the north German plain, and its oldest surviving form is the *loshoes* of the eastern Netherlands, an "open house" shared by the farm animals. Settlement there was normally grouped in fairly large villages practising open-field agriculture. The scattered distribution of plow strips is still found in many parts, but in England and Denmark, for example, consolidation and enclosure have facilitated the building of isolated farmsteads.

In France, partly because of extensive romanization, this rural life was deeply involved in urbanization, and it is characteristic of French ethnological and sociological research that it is much concerned with the problems of contact between urban and rural civilizations. Traditional folkways and crafts become more firmly entrenched eastward into central Europe. The German-speaking lands have promoted intensive research in folk culture and have published atlases in which are mapped distributions of such traits as methods of harnessing animals, plow types, harvest customs, house types, unlucky days and calendar customs.

North European Zone.—The European northlands comprise the zone of the coniferous forests and tundra in Scandinavia. There traditional cultures have been exhaustively studied by Swedish scholars, such as S. Erixon, who have also paid much attention to their country's folk-cultural connections with other parts of Europe. In Sweden too, the folk museum movement had its beginnings in Skansen (*i.e.*, the Redoubt) in Stockholm, aided by the practical vision of A. Hazelius. The folklore and customs of the Lapps have been fully recorded and such cultural features as the birdskin cap are held to point to an extremely ancient circumpolar culture region.

Atlantic Fringe.—The maritime fringe of Norway has many folk-cultural links with western Britain, Ireland and western France, northern Spain and northern Portugal. This Atlantic fringe retains social and cultural features which archaeology has shown to have been more widely distributed in prehistoric Europe. Rural settlements are usually isolated and the cultural landscapes minutely enclosed with hedge banks or stone walls, but extensive areas of open mountain and moorland have been utilized communally; and the former prevailing unit of settlement was the joint farm or hamlet with its cluster of houses (*tun* in Norway, *clachan* in Scotland, *bally* or town in Ireland) and its small open infield. The sea brought subsidiary resources in the form of fish and shore products, and also provided opportunities for trade and adventure. For the most part animal husbandry, involving cattle and frequently transhumance, has been the mainstay, and strong kinship ties, associated historically with some form of clan system, have left their trace in the use of patronymics and in forms of society which long remained patriarchal.

The long house, a type of "open house," survives in modified forms and is believed by some scholars to be the descendant of the Neolithic long house. The cruck type of roof support appears to be another survival. In addition to such prehistoric ties with central Europe, which are well illustrated in the survival of the Celtic tongue, the Atlantic fringe has many ethnological links with the Mediterranean zone: for example, it used the two-wheeled cart to the exclusion of the wagon of England and central Europe. In the peninsular ends and highland glens many calendar customs are maintained, with or without the blessing of organized religion: in their concentration on the beginnings of May and November (All Hallows) they reflect an ancient ecological division of the farming year that has reluctantly given way to the Christian calendar.

See also ANTHROPOLOGY; RACES OF MANKIND; and *The People* sections of the individual country articles. (E. E. E.)

IV. ARCHAEOLOGY

A. Western and Central Europe

1. Early Paleolithic.—Chipped flint hand axes appear to have spread from Africa to Europe during the warmer intervals between the ice ages. Flint blocks were chipped to form the axes (*coup de poing*) and this technique gave rise consecutively to the Abbevillian (or Chellean) and Acheulean industries. Flakes from the original flint core were also made into implements, and industries known as the Clactonian and Levalloisian (or older Mousterian) may be distinguished. The oldest human remains in Europe are the jawbone from the Mauer pit (Heidelberg man), the skull from Steinheim and a fragmentary skull from Swanscombe (Kent) that was found associated with implements of the Acheulean industry. The last two skulls show signs of kinship with the ancestors of *Homo sapiens* but all three belong to an earlier phase in human evolution than the Neanderthal men who populated Europe during the first advance of the last Ice Age. In the Mousterian industry associated with Neanderthal man are found tools of bone as well as stone, and evidence of the first definite burial practices. (*See* also MAN, EVOLUTION OF.)

2. Later Paleolithic.—During the interval between the first and second phases of the last (Würm) Ice Age, Neanderthal man disappeared and *Homo sapiens* spread over Europe. His equipment was far superior to that of Neanderthal man and included not only blades, scrapers and spearheads of flint but also needles, lance points and harpoons made from bone, antler and ivory. Several successive cultures may be distinguished: Aurignacian, Solutrean (chiefly central Europe and Burgundy), Périgordian and Magdalenian. One of the most interesting and distinctive features of the Upper Paleolithic period is the production of the earliest known works of art. Not only was the art of cave painting and engraving developed (in northern Spain, southern France and Sicily) but figurines of women and animals were carved and modeled in clay (particularly notable in central Europe). The cave paintings were most frequently naturalistic and probably magical representations of the large mammals that were hunted; this art reached its peak in Magdalenian times. Throughout the Paleolithic period man lived by hunting, fishing and gathering plant food.

3. Mesolithic.—The retreat of the ice sheet about 11,000 B.C. resulted in radical changes of climate, fauna and flora, giving rise to a tundralike vegetation that was later replaced by temperate forest land. The Mesolithic period stretched over the Subarctic, the Preboreal, the Boreal and a part of the Atlantic periods and during the first of these human cultures appear to have been degeneracies of the Upper Paleolithic cultures; adaptation to the new

environment proved difficult. Consequently the equipment of the Azilians of France, Switzerland and southern Germany (descendants of the Magdalenians) was relatively poor; their antler harpoons were large and clumsy, their flint implements, known as microliths, became extremely small, and the Magdalenian cave art degenerated into conventionalized designs painted on pebbles. Full adaptation to the environment was not achieved until the Boreal period. The many local cultures that flourished at that time are collectively referred to as Tardenoisian, known largely from their carefully trimmed microliths. They lived mainly on unforested sandy ground and have left some of the first traces of habitations—skin or birch-bark tents in the summer and half-subterranean timber-framed shelters in the winter.

At the same time in northern Europe communities living in a forest environment developed elaborate hunting and fishing equipment. These Maglemosians, known from excavations such as that at Star Carr (Yorkshire, Eng.), were among the first people to build canoes and to tame a domestic animal, the dog. Toward the end of the Mesolithic period in western Europe (*c.* 4000 B.C.) there is evidence of the ability to grind stone and make rudimentary pottery, techniques probably derived from the advanced civilizations of the near east.

4. Early Neolithic.—Food production by cultivation, dating from at least as early as the 8th millennium B.C. in the near east, gradually spread to western Europe. The diet was still supplemented by hunting, but the change to food production, a more efficient and reliable economy, resulted in a more settled way of life. In the villages of the first Neolithic farmers are found bones of their domesticated animals (dogs, pigs and cattle), together with the ground stone axes, pottery and female figurines that are such distinctive features of Early Neolithic cultures in southern Europe.

Mediterranean Europe.—Among the first Neolithic settlers in Europe, almost certainly immigrants from Asia Minor, were the colonists of the Starcevo culture and those of the Cardial culture. The Starcevo people, found in Greece, the Balkans and Danubia, cultivated cereals and raised pigs and cattle. Some of their pottery was painted but incised ornament was more general to the west and north. The related Cardial culture that spread farther along the Mediterranean to Malta, Italy, France and Spain, possessed the same general Neolithic characteristics, and is named from its pottery, which was decorated by designs made with the edge of a cardium shell.

Central Europe.—Although the Starcevo culture spread over a part of central Europe, it was the important Danubian culture that brought agriculture and stock breeding to most of this area. The Danubians cleared and cultivated the widespread loess soils bordering the Danube and eventually reached Germany and the Netherlands. They lived in fairly large villages composed of long rectangular timber houses and barns, existed largely by agriculture but raised enough cows, pigs and sheep to be able to neglect hunting. Their pottery was decorated with incised linear patterns of meanders and spirals and from a radiocarbon date made in Groningen (Neth.) during the 1950s they appear to have been widespread before 4000 B.C.

Western Europe.—The Mesolithic way of life lasted longer in western than in central and Mediterranean Europe and the origins of the first Neolithic peoples there are not firmly established. To what extent the early Neolithic cultures in southern England (Windmill Hill culture), in Switzerland (Cortaillod culture) and in France (Chassey culture) were interrelated, and to what degree they are associated with the Danubian culture and the early cultures in Spain and southern France had not, by the early 1960s, been satisfactorily substantiated. These Western Neolithic cultures were based on stock breeding and hunting and they all possessed "leathery" undecorated pottery. The most typical remains of the Windmill Hill culture are causeway camps (possibly used as cattle enclosures), flint mines and long barrows used for collective burial. The Swiss settlements are best known from excavations of villages probably built on the shores of the alpine lakes, where in some cases a subsequent rise in water level has preserved such organic material as wood and bread that would otherwise and in other places have had no chance of preservation. (*See* also LAKE DWELLING.)

5. Middle and Late Neolithic.—The "urban revolution" in the near east brought new influences to Mediterranean Europe, readily detectable in the Sesklo culture that succeeded the Starcevo culture in Greece and Macedonia. The economy of the Sesklo people was sufficiently advanced to enable them to live in permanent villages, and the repeated reconstruction of these on the same site led to the formation of small tells or mounds. This culture gave way to the Late Neolithic Dimini culture, characterized by fortified villages and pottery painted with spirals.

Asiatic influences also affected central Europe. In the Danubian II culture, houses became smaller and some of the villages were fortified. The people of Danubian III used copper axes and daggers (Bodrogkeresztur culture). In the Baden culture of this period the villages were larger and class differentiation was more pronounced than formerly. The equipment was still essentially Neolithic although copper trinkets of Anatolian model have been found.

The origins of the Middle Neolithic Michelsberg culture in Germany, Belgium and Switzerland are problematical. Though some of its pottery strongly suggests a Northern origin, the culture also possesses causeway camps, flint mines and some pottery that resembles features of the Early Western Neolithic—the Windmill Hill, Chassey and Cortaillod cultures. Farther north Danubian influence spread beakers with funnel necks to the German loess lands and the Netherlands sometime during the last half of the 3rd millennium B.C.

In other areas autochthonous Mesolithic populations adopted the Neolithic way of life, giving rise to "secondary" Neolithic cultures that often appear to be a degenerate form of the earlier Neolithic cultures. The Campignian culture in France and Belgium, the Seine–Oise–Marne in France and the Horgen in Switzerland are examples of these.

By the second half of the 4th millennium B.C. traders from the more advanced civilizations in the near east were beginning to spread their ritual practices and craftsmanship westward along the Mediterranean. Graves were often collective and either rock-cut or built of loose stone blocks (megaliths). They spread to Malta, where there was a unique and complex development of this phase (*see* MALTA: *Archaeology*), to Sicily, Sardinia and around the Iberian peninsula as well as across southern France. Thence there was a continued spread to Brittany (*q.v.*), the Channel Islands, England and Ireland and northward to Scotland, Denmark and southwest Sweden. The amber, metal ores and hard stone of the Baltic shores must have accounted, at least partially, for the final stages of this extensive movement. In parts of central and western Europe are found remains of the Beaker people—finely ornamented bell-shaped beakers, tanged copper daggers and bowmen's equipment. These bell beakers occur especially in some Megalithic graves in the Iberian peninsula, Brittany and the Channel Islands. The beaker-making traders and prospectors, makers of both long-necked beakers and bell-shaped beakers, usually buried their dead in single graves, although the Megalithic tombs of the native population were also used. In many regions they were the first to import metal implements. Most authorities believe they originated from the Iberian peninsula and then spread to central Europe, although others propose an origin from central Europe and a movement westward. Whichever is correct, they undoubtedly did much to spread the use of metal and to open up trade routes in Europe. In Britain they contributed to the building of Stonehenge (early phases) and Avebury (*qq.v.*).

Toward the end of the Neolithic phase (*c.* 2000 B.C.) the ground stone battle-ax was spread into Europe by warriors from the Caspian area who established themselves as a ruling class over the local populations. Their pottery was ornamented with horizontal cord impressions with local diversities of style probably connected with their manufacture by indigenous women. Battle-axes are found in Macedonia and Greece but not in the Aegean Islands or Crete. Many believe that the Battle-Ax people and the makers of the long-necked beakers contributed to the spread of Aryan speech, which developed regional diversities as the warriors and

traders mingled with indigenous populations through intermarriage with their women.

6. Bronze Age.—*Mediterranean Europe.*—While Europe was still in the Neolithic phase, the near east and Egypt had reached the stage of the "urban revolution." The quest for raw material brought the near east and Egypt into closer contact with the regions around the Aegean and an extensive trade in copper, tin, silver, obsidian, stone vessels and pottery, oil, wine and textiles flourished and the first European bronze industry developed. The Battle-Ax people invaded continental Greece, Thrace and Macedonia and established themselves as rulers, but Crete and the Aegean Islands were not invaded. The Middle Minoan culture of Crete was the time of the vast palaces of Knossos, Mallia and Phaistos. Power and wealth were concentrated in the hands of a small ruling class and the wealth of this civilization was dependent on the intense commerce that took place with Egypt and the near east, the western Mediterranean and also with central Europe via the new route over the Alps. Many elements of the Minoan civilization were adopted by the people of continental Greece but the strong fortifications of Mycenae and Orchomenus attest a more warlike tradition there, and about 1400 B.C. the Mycenaeans appear to have conquered Crete. About this time their influence seems to have been particularly far-reaching; the discovery in 1953 of engravings of a Mycenaean dagger at Stonehenge confirmed the detailed architectural similarities between Stonehenge and the Lion gate at Mycenae. By about 1100 B.C. Dorian invasions from the north overwhelmed the Mycenaean civilization and started the historical period in Greece. (*See* AEGEAN CIVILIZATION; CRETE.)

Aegean traders in Sicily (Castellucio culture) and the Remedello culture in the Po valley promoted bronze production farther west and, about 1500 B.C., the Terremaricoli settled in the Po basin. Probably of northeastern origin they made pottery of central European style and are believed to have possessed an Aryan language. By the early 1960s most scholars no longer believed that they were ancestral to the Romans.

Like Sicily, Spain developed its bronze industry under Aegean influence (El Argar culture). In France, copper and early bronze instruments are found mainly in areas reached by the early prospectors and traders as if an indigenous industry had not yet developed.

Central and Western Europe.—The discovery of the rich supplies of copper and tin ore in Bohemia and trade with the Aegean civilization gave rise to a series of autochthonous bronze-producing cultures in central Europe, of which the most important was the Unetice (Aunjetitz) culture. The weapons and tools of these bronze-producing cultures were carried all over Europe and Unetician influences were felt in both Scandinavia and northern Italy.

In western Europe bronze-producing cultures appeared all along the old maritime trade routes. Of these the Rhône culture of southeastern France and western Switzerland, the Bronze Age culture of Brittany with its strong Megalithic traditions, and the Wessex culture of Britain, where local chieftains accumulated great wealth, were the most important. The Wessex chieftains were buried in barrows with extremely rich grave goods and had much in common with the chieftains of Britanny, though some authorities believed them both to have originated from farther east. Many of their barrows are clustered around Stonehenge; the Wessex chieftains were responsible for the construction of the third phase of Stonehenge. The Bronze Age culture of Ireland (*q.v.*) produced bronze implements, weapons and gold ornaments such as the famous *lunulae*.

As the bronze-producing centres developed and the technique of bronze casting became more widely known, weapons and tools became technically better and more plentiful and hoards were found all over Europe, indicative of considerable prosperity and the practice of recasting old implements. There appears to have been widespread uniformity north of the Mediterranean and this was obviously a time of great commercial activity. Burial under mounds (Tumulus culture) was widespread, but at the beginning of the Late Bronze Age there was a change of funerary ritual. Urns containing cremated ashes were placed in flat graves in large cemeteries called urnfields. This was a period of great population movements, affecting the whole of Europe, and the Dorian invasions of Greece and the Italic immigration in Italy were probably a result of Urnfield migrations.

7. Iron Age.—*Mediterranean Europe.*—The Dorian invasions mark the beginning of the Iron Age in Greece and about the same time Italic tribes crossed the Alps and settled in Italy, where two main groups may be distinguished: (1) west of a line from Rimini to Rome cremation was practised and the area may be associated with the Latin-Faliscan dialects; (2) east of this line burial was practised and the area may be identified with the Oscan-Umbrian dialects.

The ironwork and pottery styles of the Villanova culture that spread around Felsina (Bologna) affected Tuscany and Latium and also spread by sea. The partly derivative Este (*q.v.*) culture north of the Po, probably with an Illyrian-speaking population, developed an original metallurgy and spread its products across the Alps to central Europe where it influenced the native population to a considerable extent and played a part in originating the La Tène culture.

The expansion of the Etruscan civilization, the establishment of Greek colonies in Sicily and southern Italy in the 8th century B.C. and the conquest of northern Italy by a Celtic-speaking people mark the beginning of the historical period in Italy. Only in Sardinia (*q.v.*) did the protohistoric period last longer, with its mysterious Nuraghic culture, characterized by its impressive fortress towers (*nuraghi*) and a remarkable production of bronze statuettes mostly representing warriors.

In Spain the beginning of the Iron Age is marked by immigrations of Urnfield peoples who settled in Catalonia. By about the middle of the 1st millennium B.C. the Iberian culture, with strong Greek and Carthaginian influences, and the Celtiberian culture, a result of blending of native populations with Celtic peoples from central Europe, may be distinguished.

Central and Western Europe.—In Austria a population that specialized in salt mining and trading with Italy gave rise to the Hallstatt (*q.v.*) Iron Age culture that spread to Switzerland, eastern France and Belgium as well as to southern Germany, Bohemia and western Hungary. To the west of this region the people spoke Celtic and to the east Illyrian, but otherwise the culture presents a remarkable uniformity. Among the most striking remains are the bronze and iron equipment of geometrical design and the polychrome pottery inspired by Urnfield traditions. The warlike character of these people and evidence of their connections with the Mediterranean are attested by the weapons and rich grave goods found with in-urned cremations under barrows, such as that at Vix in France. Outside the Hallstatt culture area, older Urnfield traditions are found blended with newer Hallstatt influences, as in the Lausitz (Lusatia) area, the Netherlands and some parts of France. In Britain the Iron Age A culture was probably the result of an immigration of late Urnfield peoples.

From the 7th to the 4th centuries B.C. the Celts, who originated from eastern France, southwestern Germany and northern Switzerland, spread over France, Britain, northern Spain and Italy, the Low Countries, Czechoslovakia, Danubia and even a part of Asia Minor. In particular the La Tène (*q.v.*) culture, based on the Hallstatt culture but strongly influenced by Greek, Italic, Etruscan and Scythian elements, developed a highly characteristic art style and considerable military power.

Pressure from the Germanic tribes in the north and from the Romans in the south soon put an end to Celtic domination and only in Britain did the La Tène civilization last a hundred years longer until the Roman conquest in A.D. 43.

Scandinavia, unaffected by Roman conquest, long remained "prehistoric" but when the climate warmed about the 6th century A.D. cultivation spread northward and, with iron tools, the Norsemen built clinker boats (*i.e.*, boats with horizontal boards overlapping one another downward). Sea raids and conquests as well as emigration became a feature until the seasons became more severe again after about the 12th century. Toward 1000 A.D. fortresses and churches indicate a fresh spread of urbanization eastward in Europe north of the Alps. *See* also ARCHAEOLOGY and archaeological discussions in ENGLAND; FRANCE; GREECE; ITALY;

IRELAND; NETHERLANDS; SARDINIA; SCOTLAND; SPAIN; MALTA; SCANDINAVIAN CIVILIZATION. (S. J. DE L.)

B. Eastern Europe

The life of prehistoric man in eastern Europe was largely conditioned by his natural surroundings which varied greatly from north to south. Those who lived in the north, because of the harsher conditions there, developed more slowly. This cultural lag lasted until the early middle ages.

As there are no mountains of significance inside the vast area of eastern Europe, the geographical zones stretch quite regularly from west to east. A belt of tundra bordering the Barents sea is followed by a zone of coniferous taiga whose southern limits run from Leningrad to Gorki, Kazan and Ufa. Mixed forests dominate the western part of the area farther south. A border zone between forest and steppe (wooded steppe) follows the mixed forests in the west, while in the east the forests border directly on the taiga. Typical steppe forms a belt on the northern shores of the Black sea then turns to the northeast to occupy the lower and middle basin of the Don river and the eastern shores of the middle Volga. On the lower Volga and the Ural river there is semi-desert and desert. While the forests of the north grow on poor soils conditioned by movements of the glaciers, the steppes are mostly covered by the fertile chernozem formed on loess.

There are, however, no proofs that the geographical zones had the same character in the past. On the contrary, it has been argued that the steppes were formerly wooded, their deforestation having resulted only from later action by man and his herds of domesticated animals. Nevertheless the prehistoric frontiers of the zones must have been similar to those of today: for example the territory of the Neolithic Tripolye culture coincides neatly with the present wooded steppe west of the Dnieper.

During the Ice Ages most of eastern Europe was covered by the northern European glacier. It moved several times southward and then retreated back to the north, leaving numerous morains where its fronts had formerly stood. The maximum glaciation reached the middle Dnieper south of Kiev and the middle Don near Volgograd, but the middle Volga and the lower Kama were left free of ice. It is conceivable that the climate just to the south and east of the ice sheet was rather severe, with the possible exception of the shores of the Black sea.

1. Paleolithic Period.—Lower Paleolithic artifacts are very rare and have not been found in geologically datable layers. Notably more Middle Paleolithic (Mousterian) sites lie throughout the southern part of the area, the best finds coming from the Crimean caves (*e.g.*, Kiik-Koba, with a skeleton of the Neanderthal type). In the Ukraine an open-air station of this period has yielded remains of an underground dwelling complex similar to those well known from the Upper Paleolithic.

At that time groups of mammoth hunters with the Aurignacian and Gravettian cultures populated the Ukraine and southern Russia very densely. A number of well-excavated sites are famous for their elliptical huts which are partly built of mammoth bones—at Kostienki I on the Don the dwelling area measured 36 m. by 16 m. and included nine hearths in a line, at least three subterranean huts and many pits for the storing of meat and bones. Figurines of naked women have been reported as well as many other works of art. Some of the artistic patterns are purely geometrical, as are those on the mammoth-ivory bracelets from Myzin.

2. Mesolithic and Neolithic Periods.—After the last retreat of the glacier the cultures of eastern Europe became more differentiated. Remains of the typical Azilian and Tardenoisian cultures, almost identical with those of western Europe, have been found in the Crimean peninsula. In the belts of the present steppe and wooded steppe existed cultures similar to the Tardenoisian, but farther north areas just freed from the ice cover were populated by culture groups akin to those of the eastern Baltic and the Ural region. These may have developed into the Epi-Mesolithic cultures with the so-called Neolithic pit- and comb-marked pottery.

The Russian terminology differs considerably from the English in that it does not include the food-producing economy among the traits of the Neolithic period. Thus it considers the cultures of the fishermen, hunters and food-gatherers in the northern part of eastern Europe as Neolithic on the basis that they made crude pottery and used ground stone axes. These cultures would be better termed Epi-Mesolithic. Being contemporaneous with the central European Neolithic and Bronze ages (if not even later) they are of independent eastern European origin. Similar cultures have been reported also from the southern zones but their relative chronology and subsequent development are not yet well known.

The first Neolithic cultures in eastern Europe that can be dated with some certainty have been identified as intrusions from central Europe (Linear Bandkeramik) and the northern Balkans (Boian culture), but they occupied only a small area in the southwest (Moldavia). The Boian culture developed into the famous Tripolye culture whose numerous settlements are found on the wooded steppe of the Ukraine. The Tripolye culture, with its long houses and richly painted pottery, dates back to the second half of the 4th and the first half of the 3rd millennium B.C., by which time some of the Epi-Mesolithic cultures of the northern forests must have been in existence.

3. Bronze Age.—In the Early Bronze Age eastern Europe was populated by many different culture groups whose origin had not been fully elucidated by the early 1960s.

On the northern slopes of the Caucasus (*q.v.*) rich burials of chieftains have been uncovered, containing objects that testify to a developed metallurgy (copper axes, silver vases, etc.). From there axes and other copper implements spread farther north and west, even reaching central Europe. The North Caucasian Bronze Age culture can now be followed throughout the Bronze Age up to the Early Iron Age (Koban culture).

The zones of the steppes and wooded steppes were occupied by three successive cultures, each named after its typical form of grave: (1) the pit grave, (2) the catacomb and (3) the chamber tomb. They are often called Ochre Grave or Steppe cultures, although it has been argued that they flourished in wooded areas. Again they are often cited as typical examples of cultures of nomadic herdsmen, but excavations in the Ukraine leave no doubt that they had well-established permanent villages defended by stone walls up to three metres high, and that they practised efficient agriculture. The settlement site at Mykhailivka belongs to the earliest Pit Grave culture, giving little credit to the old theory that this culture originated from the local food-gathering groups.

The central European Globular Amphorae and Corded Ware cultures expanded to the east and reached the middle Dnieper area. From there central European tribes moved north as far as the watershed of the upper Volga and the Oka in central Russia, where their culture, known from extensive cemeteries, has been named Fatyanovo. It developed into the Abashevo culture which moved farther east. The Fatyanovo culture was still surrounded by the Epi-Mesolithic food-gathering groups, that also occupied the vast territory of northern Russia.

These cultures all play an important role in the Indo-European problem. Most scholars believe that at least some of the Bronze Age cultures already mentioned must have been Indo-European-speaking, but opinions differ greatly as to what was the direction of their dispersal. Some archaeologists argue that the central European Corded Ware (Battle-Ax) culture derives from the Ochre Grave cultures of the east. This theory had by the early 1960s been discarded by the majority of specialists, but the opposite theory deriving the pit graves, and consequently also the Indo-Europeans, from the west had not gained much ground. Thus the problem seemed to be more complicated than it had been some decades previously.

4. Iron Age.—The transition between the Bronze and the Iron Age in eastern Europe remains a matter of great controversy. More is known about the Scythian culture which can be traced back to the 7th century B.C. and followed up to the 3rd, when it was superseded by the Sarmatian culture. The Scythians have become well known through their princely burials under huge *kurgans*. These contained, besides the usual grave furniture, skeletons of slain slaves and horses, and above all many beautiful works of art, believed to be products of the goldsmiths living in the

Greek north Pontic colonies. Modern excavations have also uncovered many settlements of the same date as the princely burials; they are mostly well fortified and represent places where developed handicrafts flourished.

Cultures of the same period in the north are not so well known but at least some of the Epi-Mesolithic groups gave origin to such Iron Age cultures as the Dyakovo, known from many fortified settlements. The Ananino culture of southern ancestry is found on the Kama, and because of its large cemeteries and fortifications is better known than the others.

The history of the extreme south of eastern Europe of that time was very different. There were numerous Greek colonies and many have been carefully studied by archaeological methods. These Greek colonists traded with their northern neighbours, who were at first the Indo-Iranian Scythians, but later the closely related Sarmatians with a somewhat different culture which persisted into the first centuries A.D.

At that time most of the Ukraine was inhabited by tribes that buried their dead in extensive cremation cemeteries. Two archaeological cultures have been distinguished—the Zarubynci and the Chernyakhiv. At least one of them is commonly believed to represent the ancient Slavs who came earlier from the eastern part of central Europe.

But the Slavs are not the only ethnic group of the period that can be traced. The Baltic nations must have been obscured by the cultures of the eastern Baltic shores and those farther inland. In the north and the northeast the Finns and related nations were surely under formation and, of course, the steppes of the southeast swarmed with nomads of Iranian and other origin. All have left characteristic remains from which they often can be identified. The Teutonic Goths invaded the Ukraine from the west but were later repulsed by the Huns coming from the east. From the same direction followed a number of nomadic nations, some only stopping in eastern Europe on their way to the west. Such were the Avars in the middle of the millennium, and the Magyars in the 9th century. But some nomadic Bulgars remained on the middle Volga where rich remains of their medieval culture have been uncovered. Only a part of them departed for the former Thracia (later Bulgaria) where they were absorbed by the local Slavonic population.

Early in the 1st millennium the Slavonic tribes began to spread to the north and east, but it was not until shortly after 500 B.C. that large numbers of them left their homeland in the northeastern Ukraine and swiftly colonized the eastern part of central Europe and most of the Balkan peninsula. These movements can also be traced by means of archaeology.

Thus most of eastern Europe had already been taken by the east Slavonic nations by the end of the 1st millennium A.D. (see SLAVS). The archaeological investigations do not, however, stop with the beginning of the 2nd millennium when written records became more numerous and their contents more valuable. Excavations of cities and villages destroyed by the Tatar raid in 1241 have yielded splendid results, and medieval archaeology has substantially supplemented knowledge in many other fields where the written history is inadequately informed. (E. N.)

V. HISTORY

The following is a general account of the development of European civilization from the earliest times to the present day. The historical narrative is here closely related to this theme and refers only obliquely to events outside its scope. In this sense it is a story of continuous expansion until the "Europe" which in the ancient world had originally been limited to the Mediterranean had, by the early 18th century, become identifiable with Europe in its geographical connotation. More detailed information concerning the events sketched in outline below can be found in the *History* sections of the relevant country articles; for the medieval empire in the east, *see* BYZANTINE EMPIRE.

A. Prehistory

It is unnecessary to go back further into the prehistory of transalpine Europe than the opening of the 2nd millennium B.C. to observe the first general disposition of anonymous population groups that conditioned in various ways the eventual emergence of historical peoples and nations. At this chronological horizon all the major groups of Neolithic agriculturalists and pastoralists were in position, from the Dnieper and the Vistula rivers to the Pyrenees and beyond throughout the Iberian peninsula. The beginnings of metallurgy and the long-distance co-operative efforts required to sustain a bronze industry were on the point of emergence. For the understanding of later historical Europeans, however, the appearance at this time, widely throughout transalpine Europe, of pastoralists from or closely connected with the Pontic and Caucasian regions must be noted.

It is convenient (and noncommittal) to refer to these by one of their archaeological labels—the Corded Ware people. Because of their southeastern connections, their evidently warlike nature, their early use of the domestic horse and many characteristics in ritual and material culture, there is a case for believing that the Corded Ware people spoke Indo-European dialects. They provided a continuing dominant element down to and after the florescence of the European Bronze Age in the 15th century B.C.

The inhabitants of the middle Danubian region, shielded by the Carpathians, had not been so extensively overrun by Corded Ware warriors, but they possessed their own eastern Mediterranean and Anatolian connections. These people largely lived in permanent villages. They possessed the more advanced methods of agriculture that such village life implies, and their large associated cremation cemeteries—urnfields—are of special interest for their oriental links no less than for the spread of this rite (with traditional ceramic forms) in many directions in Europe from about the middle of the 13th century B.C. This is the second important chronological horizon, for, with the destruction and abandonment of the villages through causes yet obscure, a new pattern of culture zones came into being everywhere north of the Alps. The leading characteristics of the new Urnfield cultures included not only the cremation rite but also far-reaching changes in farming and in the employment of bronze.

The interrelated cultures thus fully established by about 1000 B.C. had incorporated indigenous traditions and stocks, from region to region, but from this point in time the archaeological evidence in cultural zonation leads on ever more clearly to the appearance in history of the European barbarian peoples. The Iberian peninsula, and with it the islands of the western Mediterranean, formed a barbarian province on the western flank of classical civilization that was both different from and yet intimately linked with transalpine Europe. Its ancient peoples will here be considered first, before proceeding to the Celtic, Teutonic and Eastern peoples.

1. Western Mediterranean Peoples.—Neolithic communities spreading westward from Anatolia along the whole length of the northern shores of the Mediterranean had occupied the coast lands of the Iberian peninsula during the 3rd, if not 4th, millennium B.C. At the end of the 3rd millennium, eastern Mediterranean connections were renewed for a short time through settlers who built small defended townships and worked the copper ores. The influence of this new element on the old population was widespread, especially in the erection of elaborate tombs and the adoption of a funerary art, but the impetus soon waned, and stagnation was prolonged into the opening of the 1st millennium B.C. Only the people of the Bell Beaker culture had emerged from this scene at the opening of the 2nd millennium B.C., as an active, wandering folk penetrating far into northwestern and central Europe, apparently much involved in metallurgy and trading activities. In Sardinia and the Balearic Islands, toward the end of the 2nd millennium B.C., the construction of towers in cyclopean masonry (*nuraghi* or *talayots*) with associated clusters of dry-stone houses suggests other eastern or central Mediterranean newcomers living as lords over the older peasant populations.

An Urnfield culture appearing in Catalonia in the 8th century B.C. and spreading widely throughout the peninsula thereafter indicates the intrusion of a populace of ultimately north Alpine origin that had moved down the Rhône valley and thence to the eastern end of the Pyrenees. These people spoke a Celtic language and were recognized by the Greeks, at least from the 6th century B.C., as Celts. The Iberians who had pushed northward to occupy

Catalonia in the 5th century B.C. must be regarded as descended at least in part from the old indigenous population but with some Celtic mixture, and their culture rested principally on the urbanizing influences of the Phoenician and Carthaginian colonists of the southeastern coasts.

2. Celts.—These people, known to classical authors as the greatest barbarian nation beyond the Alps between the 6th and 1st centuries B.C., are described for this period of their prosperity and decline elsewhere (*see* CELT). In the 6th century B.C. the centre of Celtic wealth and power lay in a region roughly between the upper Danube and the headwaters of the Seine. The general economy and material culture show traditions based on the western group of Urnfield cultures, modified by the adoption of ironworking and the styles of the Hallstatt culture. It was among this western Urnfield population that the Celtic languages seem to have emerged, and there is some reason to believe that this philological province was both more ancient and more extensive than the Celts as a named barbaric nation.

The Urnfield culture in question spread throughout the Rhine valley, eastward to the Seine and northwestward toward the lower Weser. Its extension down the Rhône and thence to the Pyrenees has already been noted. From the lower Rhine, about the middle of the 6th century B.C., an important migration took place to Britain, the bearers being then equipped with iron, and most probably speaking a Celtic language ancestral to the British of Caesar's time and to modern Welsh. Other, smaller migrations associated with the La Tène culture took place during the 3rd and 2nd centuries B.C. For Ireland the sequence of events is yet unclear, but several, if relatively small, groups of Celtic immigrants appear to have been involved, beginning possibly as early as the 8th century B.C. and continuing until the arrival of many refugees from Gaul and Britain during the Roman campaigns in those lands.

3. Teutonic Peoples.—During the 2nd millennium B.C. a remarkably uniform culture province can be distinguished in the west Baltic area, centred on the Danish islands and extending to the lower courses of the Elbe and the Oder. Corded Ware people had been a dominant element, and the rich Bronze Age culture that flourished during the second half of the millennium was supported by a fusion of this and other originally Neolithic stocks. At about the beginning of the 12th century B.C. adoption of the cremation rite together with the appearance of new forms of pottery, weapons and tools suggests something more than mere contact with the Urnfield cultures of middle Europe. The new influences in the north came particularly from the Lausitz Urnfield province between the Oder and the Vistula.

The Lausitz culture itself owed much to middle Danubian sources, and this too may have some bearing on the origin and particular nature of the Teutonic language group. With the disruption of old trade routes across central Europe and with the expansion of the Celts, the west Baltic province became isolated and impoverished. The cultures that came into existence with an iron-using economy, about 500 B.C., embodied local Bronze Age traditions and peripheral Hallstatt influences. However undistinguished in content, Iron Age cultures of the west Baltic type penetrated southward along the Elbe and the Oder, and they undoubtedly represent the early stages of Teutonic expansion. For the immediate prehistoric background of the Teutonic peoples, on the threshold of their great migrations, it may be said that the lands between the Elbe and the Oder furnished the westward-moving tribes, while the lower Vistula was the seat of the eastern group, notably the Goths. But already the whole southern Baltic zone had received additional migrants from Scandinavia. As these peoples possessed no comprehensive name, Teutonic, allied to *Deutsch* (originally meaning "people"), is here preferred to Germanic. The name German is a comparatively modern misnomer revived from the use put to it by Tacitus, who employed it for all Celtic and non-Celtic tribes living east of the Rhine.

4. Eastern Peoples.—The grasslands that extended from the Black sea northwestward outside the Carpathians as far as the Baltic provided at the beginning of the 2nd millennium B.C. not only suitable habitation for pastoralists and mixed farming communities but also means of long-distance intercommunication. Here Corded Ware people had first intruded on older communities. From the middle of the 2nd millennium B.C. various regional cultures have been distinguished, evidently representing continuity in settlement but showing varying degrees of influence from both the Bronze Age cultures of middle Europe and the steppeland. A phase of renewed activity over the steppe resulted in the appearance of the Cimmerians. Some burials and bronze horse trappings of the 9th century B.C. have been ascribed to them. Then, during the 8th and 7th centuries, followed the Scythians, whose warlike qualities and vivid animal art style had a profound influence even beyond the Carpathian shield. By the 3rd century B.C. the Scythians were being threatened by the Sarmatians and were eventually overcome by them.

All these three westward-moving peoples from the steppes are considered to have spoken Indo-European languages, but they may be regarded more as having been transitory overlords rather than substantial settlers. It is impossible to point to any one of them alone as exclusively ancestral to the first historical peoples of eastern Europe. The Slavs, whose historical documentation is later and less informative than that of the Celts and Teutonic peoples, probably represent the ultimate fusion of all the Indo-European elements on the western fringes of the steppeland. The people of the Lausitz culture may also have contributed, but the sequence of events remains very obscure. For the east Baltic Indo-European-speaking peoples, as represented by the Letts and Lithuanians, the archaeological evidence appears to indicate an undisturbed continuity of settlement going back to the time of the intrusion of Corded Ware people into that region. *See* INDO-EUROPEANS. (T. G. E. P.)

B. Greeks, Romans and Barbarians

1. Greeks.—The history of Europe was determined by the diffusion throughout the continent of Indo-European tribes of European origin, of whom the Greeks were foremost both as regards the period at which they developed an advanced culture and their importance in further evolution. The Greeks came into being in the course of the 2nd millennium B.C. through the superimposition of a branch of the Indo-Europeans on the population of the Mediterranean region in the course of the great migrations of nations that started in the region of the lower Danube. From 1800 B.C. onward the first early Greeks reached their later areas of settlement between the Ionian and the Aegean seas. The fusion of these earliest Greek-speaking people with their predecessors produced the civilization known as Mycenaean. They penetrated to the sea into the Aegean region and via Crete (approximately 1400 B.C.) reached Rhodes and even Cyprus and the shores of Asia Minor. From 1200 B.C. onward the Dorians followed from Epirus. They occupied principally parts of the Peloponnese (Sparta and Argolis) and also Crete. Their migration was followed by a dark age—two centuries of chaotic movements of tribes in Greece—at the end of which (*c.* 900 B.C.) the distribution of the Greek mainland among the various tribes was on the whole completed. (*See* GREECE; *History;* ACHAEANS; DORIANS; AEGEAN CIVILIZATION.)

From *c.* 800 B.C. there was a further remarkable Greek expansion through the founding of colonies overseas. The coasts and islands of Asia Minor were occupied, from south to north, by the Dorians, the Ionians and Aeolians respectively. In addition, individual colonies were strung out around the shores of the Black sea in the north and across the eastern Mediterranean to Naukratis on the Nile delta and in Cyrenaica and also in the western Mediterranean in Sicily, lower Italy and Massilia (Marseilles). Thus, the Hellenes, as they called themselves thereafter, came into contact on all sides with the old, advanced cultures of the middle east and ultimately transmitted many features of these cultures to western Europe. By their own genius, however, the Greeks accomplished much more, and their achievements laid the foundations of European civilization.

The position and nature of the country exercised a decisive influence in the evolution of Greek civilization. The proximity of the sea tempted the Greeks to range far and wide exploring it, but the fact of their living on islands or on peninsulas or in valleys separated by mountains on the mainland confined the forma-

tion of states to small areas not easily accessible from other parts. This fateful individualism in political development was also a reflection of the Hellenic temperament. Though it prevented Greece from becoming a single unified nation that could rival the strength of the middle eastern monarchies, it led to the evolution of the city-state. This was not merely a complex social and economic structure and a centre for crafts and for trade with distant regions; above all it was a tightly knit, self-governing political and religious community, whose citizens were prepared to make any sacrifice to maintain their freedom. Colonies, too, started from individual cities and took the form of independent city-states. Fusions of power occurred in the shape of leagues of cities, such as the Peloponnesian league, the Delian league and the Boeotian league. The efficacy of these leagues depended chiefly upon the hegemony of a leading city (Sparta, Athens or Thebes), but the desire for self-determination of the others could never be permanently suppressed and the leagues broke up again and again.

The Hellenes, however, always felt themselves to be one people. In relation to the "barbarians" they were conscious of a common character and a common language, and they practised only one religion. Furthermore, the great athletic contests and artistic competitions had a continually renewed unifying effect. The Hellenes possessed a keen intellect, capable of abstraction, and at the same time a supple imagination. They developed, in the form of the belief in the unity of body and soul, a serene, sensuous conception of the world. Their gods were only loosely connected by a theogony that took shape gradually; in the Greek religion there was neither revelation nor dogma to oppose the spirit of inquiry (*see* GREEK RELIGION). The Hellenes benefited greatly from the knowledge and achievement of other countries as regards astronomy, chronology and mathematics, but it was through their own native abilities that they made their greatest achievements, in becoming the founders of European philosophy and science. Their achievement in representative art and in architecture was no less fundamental. Their striving for an ideal, naturalistic rendering of the third dimension, found its fulfillment in the representation of the human body in sculpture in the round. Another considerable achievement was the development of the pillared temple to a greater degree of harmony. In poetry the genius of the Hellenes created both form and content, which have remained a constant source of inspiration in European literature.

The strong political sense of the Greeks produced a variety of systems of government, from which their theory of political science abstracted types of constitution that are still in use. On the whole, political development in Greece followed a pattern: first, the rule of kings, found as early as the period of Mycenaean civilization; then a feudal period, the oligarchy of noble landowners; and finally, varying degrees of democracy. Frequently there were periods when individuals seized power in the cities and ruled as tyrants. The tendency for ever-wider sections of the community to participate in the life of the state brought into being the free democratic citizens, but the institution of slavery, upon which Greek society and the Greek economy rested, was untouched by this.

In spite of continual internal disputes the Greeks succeeded in warding off the threat of Asian despotism. The advance of the Persians into Europe failed (490 B.C. and 480–79 B.C.) because of the resistance of the Greeks and in particular of the Athenians. The 5th century B.C. saw the highest development of Greek civilization. The classical period of Athens and its great accomplishments left a lasting impression, but the political cleavages, particularly the struggle between Athens and Sparta, increasingly reduced the political strength of the Greeks. Not until they were conquered by the Macedonians did the Greeks attain a new importance as the cultural leaven of the empires of Alexander the Great and his successors. A new system of colonization spread as far as the Indus city-communities fashioned after the Greek prototype, and Greek education and language came to be of consequence in the world at large. (*See* HELLENISTIC AGE.)

Greece again asserted its independence through the formation of the Achaean league (*q.v.*), which was finally defeated by the Romans in 146 B.C. The spirit of Greek civilization subsequently exercised a great influence upon Rome. Greek culture became one of the principal components of Roman imperial culture, and together with it spread throughout Europe. When Christian teaching appeared in the middle east the Greek world of ideas exercised a decisive influence upon its spiritual evolution. From the time of the partition of the Roman empire leadership in the Eastern empire fell to the Greeks. Their language became the language of the state and its usage spread to the Balkans. The Byzantine empire, of which Greece was the core, protected Europe against potential invaders from Asia Minor until the fall of Constantinople in 1453.

2. Romans.—The original Mediterranean population of Italy was completely altered by repeated superimpositions of peoples of Indo-European stock. The first Indo-European migrants, who belonged to the Italic tribes, moved across the eastern Alpine passes into the plain of the Po river about 1800 B.C. Later they crossed the Apennines and eventually occupied the region of Latium, which included Rome. Before 1000 B.C. there followed related tribes, which later divided into various groups and gradually moved to central and southern Italy (*see* ITALY). In Tuscany they were repulsed by the Etruscans (*q.v.*), who may have come originally from Asia Minor. The next to arrive were Illyrians from the Balkans who occupied Venetia and Apulia (*qq.v.*). At the beginning of the historical period Greek colonists arrived in Italy, and after 400 B.C. the Celts, who settled in the plain of the Po.

The city of Rome, increasing gradually in power and influence, created through political rule and the spread of the Latin language something like a nation out of this abundance of nationalities. In this the Romans were favoured by their kinship with the other Italic tribes. Moreover, the Roman and Italic elements in Italy were reinforced in the beginning through the founding of colonies by Rome and by other towns in Latium. The Italic element in Roman towns decreased: a process—less racial than cultural—called the romanization of the provinces. In the 3rd century B.C. central and southern Italy were dotted with Roman colonies, and the system was to be extended to ever more distant regions up to imperial times. As its dominion spread throughout Italy and covered the entire Mediterranean basin, Rome received an influx of people of the most varied origins, including eventually vast numbers from Asia and Africa.

The building of an enormous empire was Rome's greatest achievement. Held together by the military power of one city, in the 2nd century A.D. the Roman empire extended throughout northern Africa and western Asia; in Europe it covered all the Mediterranean countries, Spain, Gaul and southern Britain. This vast region, united under a single authority and a single political and social organization, enjoyed a long period of peaceful development. In Asia, on a narrow front, it bordered the Parthian empire, but elsewhere beyond its perimeter there were only barbarians. Rome brought to the conquered parts of Europe the civilization the Greeks had begun, to which it added its own important contributions in the form of state organization, military institutions and law. Within the framework of the empire and under the protection of its chain of fortifications, extending, uninterrupted, the entire length of its frontiers (marked in Europe by the Rhine and the Danube), there began the assimilation of varying types of culture to the Hellenistic-Roman pattern. The army principally, but also Roman administration, the social order and economic factors, encouraged romanization. Except around the eastern Mediterranean, where Greek remained dominant, Latin became everywhere the language of commerce and eventually almost the universal language.

The empire formed an interconnected area of free trade, which was afforded a thriving existence by the *pax romana*. Products of rural districts found a market throughout the whole empire, and the advanced technical skills of the central region of the Mediterranean spread outward into the provinces. The most decisive step toward romanization was the extension of the city system into these provinces. Rural and tribal institutions were replaced by the *civitas* form of government, according to which the elected city authority shared in the administration of the surrounding country region, and as the old idea of the Greek city-state gained ground, a measure of local autonomy appeared. The romanized

upper classes of the provinces began supplying men to fill the higher offices of the state. Ever-larger numbers of people acquired the status of Roman citizens, until in A.D. 212 the emperor Caracalla bestowed it on all freeborn subjects. The institution of slavery, however, remained.

The enjoyment of equal rights by all Roman citizens did not last. The coercive measures by which alone the state could maintain itself divided the population anew into hereditary classes according to their work; and the barbarians, mainly Germanic, who were admitted into the empire in greater numbers, remained in their own tribal associations either as subjects or as allies. The state created a perfected administrative apparatus, which exercised a strongly unifying effect throughout the empire, but local self-government became less and less effective under pressure from the central authority.

The decline of the late empire was accompanied by a stagnation of spiritual forces, a paralysis of creative power and a retrograde development in the economy. Much of the empire's work of civilization was lost in internal and external wars. Equally, barbarization began with the rise of unchecked pagan ways of life and the settlement of Germanic tribes long before the latter shattered the western empire and took possession of its parts. Though many features of Roman civilization disappeared, others survived in the customs of peoples in various parts of the empire. Moreover, something of the superstructure of the empire was taken over by the Germanic states, and much valuable literature was preserved in manuscript for later times.

It was under the Roman empire that the Christian religion penetrated into Europe. By winning recognition as the religion of the state it added a new basic factor of equality and unification to the imperial civilization and at the same time reintroduced middle-eastern and Hellenistic elements into the west. Organized within the framework of the empire, the church became a complementary body upholding the state. Moreover, during the period of the decline of secular culture, Christianity and the church were the sole forces to arouse fresh creative strength, by assimilating the civilization of the ancient world and transmitting it to the middle ages. At the same time the church in the west showed reserve toward the speculative dogma of the middle-eastern and Hellenic worlds and directed its attention more toward questions of morality and order. When the western empire collapsed and the use of Greek had died there, the division between east and west became still sharper. The name "Romaioi" remained attached to the Greeks of the eastern empire, while in the west the word "Roman" developed a new meaning in connection with the church and the bishop of Rome. Christianity and a church of a Roman character, the most enduring legacy of the ancient world, became one of the most important features in western European civilization.

3. The Migrations.—The wanderings of the Germanic peoples (*see* GERMANIC PEOPLES), which lasted until the early middle ages and destroyed the western Roman empire, were, together with the migrations of the Slavs, formative elements of the distribution of peoples in modern Europe. The Germanic peoples originated about 1800 B.C. from the superimposition, on a population of megalithic culture on the eastern North sea coast, of Battle-Ax people from the Corded Ware culture of middle Germany. During the Bronze Age, the Germanic peoples spread over southern Scandinavia and penetrated more deeply into Germany between the Weser and the Vistula rivers. Contact with the Mediterranean through the amber trade encouraged the development from a purely peasant culture, but during the Iron Age the Germanic peoples were at first cut off from the Mediterranean by the Celts and Illyrians. Their culture declined and an increasing population together with worsening climatic conditions drove them to seek new lands farther south. Thus the central European Celts and Illyrians found themselves under a growing pressure. Even before 200 B.C. the first Germanic tribes had reached the lower Danube, where their path was barred by the Macedonian kingdom. Driven by rising floodwaters, at the end of the 2nd century B.C. migratory hordes of Cimbri, Teutoni and Ambrones from Jutland broke through the Celtic-Illyrian zone and reached the edge of the Roman sphere of influence, appearing first in Carinthia (113 B.C.), then in southern France and finally in upper Italy. With the violent attacks of the Cimbri the Germans stepped onto the stage of history.

These migrations were in no way nomadic; they were the gradual expansions of a land-hungry peasantry. Tribes did not always migrate en masse. Usually, because of the loose political structure, groups remained in the original homelands or settled down at points along the migration route. In course of time, many tribes were depleted and scattered. On the other hand, different tribal groups would sometimes unite before migrating or would take up other wanderers en route. The migrations required skilled leadership, and this promoted the social and political elevation of a noble and kingly class. In 102 B.C. the Teutoni were totally defeated by the Romans, who in the following year destroyed the army of the Cimbri. The Swabian tribes, however, moved steadily through central and southern Germany, and the Celts were compelled to retreat to Gaul. When the Germans under Ariovistus crossed the upper Rhine, Julius Caesar arrested their advance and initiated the Roman countermovement with his victory in the Sundgau (58 B.C.). Under the emperor Augustus, Roman rule was carried as far as the Rhine and the Danube. On the far side of these rivers the Germans were pushed back only in the small area contained within the Germano-Raetian limes (fortified frontier) from about A.D. 70.

The pressure of population was soon evident once more among the German peoples. Tribes that had left Scandinavia earlier (Rugii, Goths, Gepidae, Vandals, Burgundians and others) pressed on from the lower Vistula and Oder rivers (A.D. 150 onward). The unrest spread to other tribes, and the resulting wars between the Romans and the Marcomanni (A.D. 166–180) threatened Italy itself. The successful campaigns of Marcus Aurelius resulted in the acquisition by Rome of the provinces of Marcomannia and Sarmatia, but after his death these had to be abandoned and the movement of the Germanic peoples continued. Soon the Alamanni (*q.v.*), pushing up the Main river, reached the upper German limes.

To the east the Goths (*q.v.*) had reached the Black sea about A.D. 200. Year after year Goths and others, either crossing the lower Danube or going by sea, penetrated into the Balkan peninsula and Asia Minor as far as Cyprus on plundering expeditions. Only with the Roman victory at Nish (269) were they checked. Enriched with booty and constituted imperial mercenaries in return for a yearly tribute, they became a settled population. The Romans, however, surrendered Dacia beyond the Danube.

In 258 the Alamanni and the Franks (*q.v.*) broke through the limes and settled on the right bank of the Rhine, continuously infiltrating thereafter toward Gaul and Italy. Everywhere within the empire towns were fortified, even Rome itself. Franks and Saxons ravaged the coasts of northern Gaul and Britain, and for the next three centuries incursions by Germanic peoples were the scourge of the western empire. Nevertheless, it was only with German help that the empire was able to survive so long. The Roman army received an ever-growing number of recruits from German tribes, which also provided settlers for the land. The Germans soon proved themselves capable of holding the highest ranks in the army. Tribute money to the tribes, pay to individual soldiers and booty all brought wealth to the Germans, which in turn gave warrior lords the means with which to maintain large followings of retainers. In the west, however, among the Alamanni and Franks, the beginnings of political union into larger groups did not go beyond loose associations. Only in the east did the Gothic kingdom gather many tribes under a single leadership. Above all, the development of the East Germans was stimulated by their undisturbed contact with the frontiers of the ancient world. Their economy, however, was still unable to support a growing population, and pressure from overpopulation led to further incursions into the Roman empire. The imperial reforms of Diocletian and Constantine the Great brought a period of improvement. The usurpation of the imperial title by a Frankish general in 356 let loose a new storm along the whole Rhine and subsequently on the Danube, but the frontiers were restored by the emperors Julian and Valentinian I, who repelled attacks by the Franks and the Alamanni.

Now, however, a new force appeared. In 375 occurred the first attack on the Ostrogoths by the Huns (*q.v.*) from central Asia—an event which provoked serious disturbances among the East Germans. The Huns remained in the background, gradually subjugating many Germanic and other tribes. The terrified Goths and related tribes burst through the Danube frontier into the Roman empire, and the Balkans became once again a battlefield for German armies. After the crushing defeat of the Romans at Adrianople (378) the empire was no longer in a position to drive all its enemies from its territories. Tribes that could no longer be expelled were settled within the empire as "allies" (*foederati*). They received subsidies and in return supplied troops. The germanization of the empire progressed, that of the army being nearly completed. However, none of the tribes that had broken into the Balkans settled there. After the division of the empire in 395 the emperors at Constantinople did all in their power to drive the Germanic tribes away from the vicinity of the capital toward the western empire.

From the beginning of the 5th century A.D. the western empire was the scene of numerous further migrations. The Visigoths broke out of the Balkans into Italy and in 410 temporarily occupied Rome. In 406–407 Germanic and other tribes (Vandals, Alani, Suebi and Burgundians) from Silesia and even farther east crossed the Rhine in their flight from the Huns and penetrated as far as Spain. The Vandals (*q.v.*) subsequently crossed to Africa and set up at Carthage the first independent German state on Roman soil. In the battle of the Catalaunian Plains (*q.v.*; 451) the Roman commander Aëtius, with German support, defeated Attila, who had united his Huns with some other Germans in a vigorous westward push. The Balkans suffered a third period of terrible raids from the East Germans; and Jutes, Angles and Saxons from the Jutland peninsula crossed over to Britain. The Franks and the Alamanni finally established themselves on the far side of the Rhine, the Burgundians extended along the Rhône valley, and the Visigoths took possession of nearly all Spain. In 476 the Germanic soldiery proclaimed Odoacer, a barbarian general, as king of Italy, and when Odoacer deposed the emperor Romulus Augustulus at Ravenna the empire in the west was at an end.

The restoration of the imperial power in the west by Justinian I, who reconquered Africa, Italy and parts of Spain, was of short duration. The Visigoths regained their position in Spain; Africa was lost to the Arabs; and in 568 the Lombards (*q.v.*) conquered upper and central Italy, Rome and Ravenna alone excepted. The Lombards had been driven from the plains of the Danube and Tisa by a new wave of mounted Asian nomads, the Avars (*q.v.*). The Germanic migrations on the European mainland were finally at an end. All the Roman territories between the Danube and Rhine and the Atlantic had been penetrated by the Germans. In western Europe a new ethnic pattern was established, which subsequently was scarcely altered.

The Germanic tribes were roughly distributed into three zones. The central part of Germany was composed of peoples of purely Germanic stock. In the second zone a mixture of Germanic peoples ruled in Britain, in the country west of the Rhine as far as the Seine basin and south of the Danube toward the Alps. In the third zone (southern France, Italy and Spain) the German element was much dispersed among the Romance peoples.

The migrations had meant that at first only the Germans' warlike qualities were revealed, and their manifold talents only gradually became apparent. But originating with the Goths, a characteristic decorative art developed which, though eclectic, gave expression to the spirit of the Germanic ruling class. Together with an epic poetry, purely Germanic in origin, it reached its highest point with the growth of independent Germanic states, and its influence was carried on into the middle ages. The Germanic states, with their simpler political systems, saved western society from authoritarian oppression. Beginning with the great translation of the Bible by Ulfilas (*q.v.*), the Goths created a religious literature which, however, was soon destroyed because of its Arianism. Nevertheless, codes of law which, with the help of Roman jurists and the Latin language, were promulgated by the Visigoths (*c.* 450), by the Ostrogoths (508) and by the Franks (*lex salica*, 508–511) survived. To a certain extent the Germans were able to take over constituent parts of the Roman state. Through their conversion to Christianity they became members of the Roman Church, which had most fully inherited the culture of antiquity.

The Slavs (*q.v.*) were the last of the Indo-Europeans to settle in their present home. They originally occupied the area between the Vistula and Dnieper rivers, stretching northward from the Carpathian mountains as far as the Narew river. The Slavs were first mentioned by name in 518, when they broke into the Roman empire by way of the lower Danube. Their social structure was simple, consisting of small tribal units. Their further expansion was caused partly by the Avars, who from 568 were the sole rulers of the Hungarian plain, holding the surrounding Slavs in subjection. In 623–624 the Slavs overthrew the Avar lordship in Bohemia. Gradually they established frontiers with the German tribes in the Eastern Alps at the Bohemian forest and along the Saale and Elbe rivers. By the end of the 8th century at the latest, Slav settlements stretched from the east coast of Holstein as far as the Vistula. The weakness of the Byzantine empire made possible further attacks across the Danube. Even after the heavy defeat of the Avars before Constantinople in 626 the Slavs remained in the Balkan peninsula under Byzantine lordship. The Balkans became a Slav territory, and the Greeks fell back on the eastern and southern coasts and the Romans on the Adriatic or remained in isolation as "Balkan Rumanians."

In 679 the Bulgars, a Turanian people from southern Russia, burst across the lower Danube and conquered the eastern part of the Balkan peninsula. They were quickly assimilated to the Slavs, however, and with them were converted to Christianity by the end of the 9th century (*see* BULGARIA). A new invasion from the 9th century onward of Asian nomads, the Magyars (*see* HUNGARY), divided the Slavs of the south from those of the west and the east. The West Slavs moreover (Czechs, Slovaks, Elbe Slavs, Poles and Pomeranians) received Christianity from the west, and this separated them from the East Slavs (White Russians, Ukrainians and Great Russians), who adhered to the Greek Orthodox Church. After the Mongol invasions from 1223 onward the East Slavs withdrew from the Dnieper territories toward the wooded lands to the northwest, where they drove back the Finns and themselves developed as the Great Russians. (H. A.)

C. The Middle Ages

The ancient world had unified around the Mediterranean—politically by the supremacy of Rome and culturally by a common Greco-Roman or Hellenistic literature, thought and art. This unity was never complete, and signs of disintegration were clearly apparent during the 4th century; within another 250 years the unity of the Mediterranean was permanently broken. Thus antiquity was both more and less than the beginning of western civilization—more because it was also the beginning of the Byzantine and the Islamic civilizations, less because western civilization included elements that were lacking or not fully assimilated in the Roman empire at its height. Medieval Europe expanded far beyond the geographical limits of antiquity and included peoples who were never under Roman rule, and its culture was dominated by Christianity, which originated within the empire as an element alien to pagan and classical antiquity. The heritage of Rome was fused with the Germanic and Christian elements to provide the medieval foundations of western civilization.

1. Germanic Kingdoms.—The decline of Roman authority in the west, culminating in the deposition of Romulus Augustulus (476), and the establishment of Germanic kingdoms in its place is described above. The eastern provinces of the empire fared better during the period of invasions after 378: the emperors in Constantinople ensured their safety not only by arms but also by a diplomacy aimed at sending troublesome tribes westward to gain their fortunes. Thus in 489 Zeno granted the title of patrician to Theodoric, king of the Ostrogoths, and gave him permission to conquer Italy and rule there in the emperor's name. The kingdom of the Ostrogoths under Theodoric (493–526) was the most advanced of the barbarian kingdoms. Like the others, but far more

successfully, the Ostrogothic kingdom strove to preserve as much of Roman institutions and culture as its more limited resources allowed.

The last great Germanic conquest and the most lasting was that of Clovis (481 or 482–511), who led his Franks into northern Gaul and by 507 had pushed the Visigoths south into Spain. Only the kingdom of Burgundy and the Mediterranean coastal strip (Septimania and Provence) lay outside the Frankish Merovingian kingdom. Perhaps the most important reason for Clovis' success, apart from his ability as a warrior, was his policy toward the church. Unlike most of the other Germanic tribes, who were Christians of the heretical Arian sect, the Franks were still heathen when they crossed into Gaul. The conversion of Clovis to the Catholic faith was a decisive moment in the history of western Europe. He was now the only Catholic ruler among the Germanic kings and thus gained what no other ruler could: the support of the church and the Gallo-Roman subject population. His conquest of Gaul became a war of liberation from the yoke of the hated Arian heretics.

In the 6th century none of the Germanic kingdoms achieved political stability, except for brief periods under exceptionally strong rulers. In Gaul the Merovingian kingdom was divided among the sons after the death of the king, a provision that almost guaranteed continual war for supremacy. In Anglo-Saxon England, where partible inheritance was not the rule, there were yet so many original petty kingdoms (Bernicia, Deira, Lindsey, Mercia, East Anglia, Essex, Wessex, Kent and several other regions whose tribal leaders briefly claimed royal status) that warfare there was also endemic as each ruler strove to maintain and expand the political boundaries of his kingdom. On the continent, except in Merovingian Gaul, the barbarian kingdoms declined. Burgundy was conquered by the Merovingians; the Byzantine emperor, Justinian, destroyed the Vandal kingdom, reconquered Ostrogothic Italy and temporarily reoccupied the southern part of Visigothic Spain, in campaigns lasting from 533 to 556. But Italy was so ravaged that the real beneficiaries of Justinian's policy were the Lombards, the last of the Germanic peoples to settle within the limits of the Roman empire. Beginning in 568, the Lombard conquest was rapid, and by about 600 only a few districts, such as Ravenna, Sicily and Rome, remained nominally subject to the Byzantine empire, organized as the exarchate of Ravenna (*q.v.*).

Except in England, where the vestiges of late Roman civilization were almost wiped out, the barbarian conquerors did not purposefully destroy the Roman system they took over; rather, they were unable to maintain it. The *civitas* (city with surrounding countryside) continued to be the unit of local government, but the local count (*q.v.*) had only loose ties to the central royal government and it was so difficult to collect the old Roman taxes that they eventually disappeared. In the stagnant economy, however, this was not serious, for the barbarian government performed almost no service other than to maintain some semblance of law and order. The law was twofold, in accordance with the Germanic concept of "personality of law." The provincial Roman population lived under codes derived from the already simplified version of Roman law in the Theodosian code (*c.* 438), while the various Germanic tribes lived under their *leges barbarorum* ("laws of the barbarians"). The latter, written in Latin, were also influenced by Roman law concepts.

2. Genesis of Latin Christendom.—Christianity originated as an offshoot from Judaism, in an eastern province of the Roman empire. It began as a religion profoundly hostile to many of the values and vested interests of the ancient world. At first merely suspect and despised, later actively repressed by the Roman government, within three centuries it had spread throughout the empire, and the emperor was a Christian. Christianity had absorbed many of the best cultural and institutional elements of antiquity, which it was the historic role of the church to transmit to the middle ages. By the end of the 4th century, not only was Christianity the official religion of the empire but all other religions were proscribed.

The gradual division of Christianity into a Latin and a Greek communion was part of the breakdown of the cultural unity of the Mediterranean. By the end of the 3rd century there were few educated men who could read and speak both Greek and Latin. Equally important in the growth of a distinctly Latin church was the rise of the papacy (*q.v.*) between the 4th and 7th centuries. The popes were recognized as specially pre-eminent in the west by the emperors of the 4th century, and papal supremacy was recognized by the Council of Constantinople (381), though such recognition was not accepted in the Greek east. The absence of the imperial government, now in Constantinople, and the consistent record of the bishops of Rome for conservatism and orthodoxy during the early heresies, also contributed to the rise of the papacy. At an early date the Petrine theory of the supremacy of the pope as St. Peter's successor was widely accepted, though the first pope to enunciate the doctrine formally was Leo I the Great, in the 5th century.

With the decline of imperial authority in Italy the pope became practically an independent secular ruler of Rome, and with the disintegration of the empire into barbarian kingdoms, the church, with the pope at its head, became the most important element of unity in the west. Latin Christendom grew with the conversion of the barbarians to the Catholic faith (either from Arianism or from heathen religious beliefs). Conversion was actively promoted by missionaries of the monastic order founded by St. Benedict, about 520, and their work was often sponsored by the papacy. Especially notable was the conversion of the Anglo-Saxons in the 7th century and of the Germans in the 8th. In each case the papacy either initiated or supported a movement undertaken by monks.

The art, literature and thought of the early middle ages were almost completely dominated by the interests of the church. The age of the Latin Fathers, whose greatest achievement in theology was reached in the works of Ambrose, Jerome and Augustine, was followed by a long cultural decline (*see* FATHERS OF THE CHURCH). The three centuries from about 450 to around 750 were the "dark ages" of western Europe, but even in this period not all was darkness. The monks not only spread the faith but also, in their monasteries, preserved the literary remains of classical antiquity as well as the writings of the Fathers. Among the writers who had the greatest influence upon or interest for the future of medieval culture were Boëthius, Gregory of Tours, Pope Gregory the Great, Isidore of Seville and the best scholar of the whole period from the decline of Rome to the high middle ages, the Northumbrian historian and theologian, Bede.

3. Final Breakdown of Mediterranean Unity.—The Greek half of the Roman empire did not suffer, relatively, much economic or political loss from the barbarian invasions that inaugurated the early middle ages in the west. In fact, the Byzantine empire (*q.v.*) preserved much of the civilization of antiquity and added elements from the Hellenistic and oriental world that further differentiate the Byzantine east from the Latin west. Under Justinian I (527–565) the Byzantine empire enjoyed a prosperity and a flourishing of the arts that made later generations think of that period as a golden age. But Justinian failed in his greatest ambition, to reconquer the lost western provinces and reunite the Mediterranean. His greatest monuments were not military and political, as he hoped, but cultural: the great church of Hagia Sophia and the codification of the Roman law in the *Corpus juris civilis*. Actually, the Byzantine empire had weaknesses that neither Justinian nor his successors were able to overcome. Most notable was the bitter and sometimes violent religious dissension between the Greek and Orthodox people of Constantinople and the nearby provinces and the various Christian sects in the provinces of Egypt, Syria and parts of Asia Minor. Also, for centuries, danger from Persia had drained Byzantine resources; finally, Heraclius (610–641) was triumphant in a great war against Persia, but the empire emerged almost wholly exhausted.

In the struggle between Persia and the Byzantine empire the ultimate victor was Islam. The prophet Mohammed in the 6th century united the Arabs under his religious and military leadership. In less than a hundred years his successors, the caliphs (*see* CALIPHATE), conquered lands that stretched all the way from southern Gaul and Spain to western India. The Byzantine prov-

inces of north Africa, Egypt, Syria and Palestine (where religious antagonism toward Constantinople undermined resistance against the relatively tolerant Arabs) fell quickly; all of Persia was overrun in a few years. Christendom—both Greek Orthodox and Latin—was confronted with a successful and hostile Arab empire whose unity and strength derived from a common religion, Islam, and a common Arab culture and language.

The breakdown of Mediterranean unity was completed by the middle of the 8th century. The Merovingian kings had degenerated into *rois fainéants*, and their power had passed to the Carolingian mayors of the palace (*see* CAROLINGIANS). The popes, threatened by the Lombards who sought to conquer all territories still under Byzantine sovereignty, appealed in vain to Constantinople for help. The Byzantine emperors, hard pressed to defend their frontiers against the Muslims, could spare no troops to rescue Rome from Lombard conquest. Further, the emperor Leo III the Isaurian (717–741) inaugurated a religious policy, iconoclasm, which Latin Christians considered heretical; Leo and his successors attempted to force iconoclasm on the pope. Confronted with this situation the pope sought help from another quarter—the Carolingian mayor in Merovingian Gaul, Pepin the Short, who was eager to win the pope's support for his desire to supplant the Merovingians and become king of the Franks. The alliance between papacy and Frankish monarchy was completed when Pepin was consecrated king in return for intervention in Italy to save Rome from the Lombards. By transferring to the pope the former Byzantine territory which he reconquered from the Lombards, Pepin established the papal states under the pope as temporal ruler. The pope then conferred upon Pepin the title of *patricius Romanorum*. To justify these transactions (which, to be legal, should have been accomplished or at least approved by the Byzantine emperor), the papal chancery or someone familiar with its practice produced the Donation of Constantine (*q.v.*), the most famous forgery of all time. It purported to be a grant of supreme authority in the west from the emperor Constantine to Pope Silvester I and his successors.

4. Carolingian Empire.—As king, Pepin extended his power over all of Gaul by reducing counts and dukes to obedience and pushing the Muslims back across the Pyrenees. His successor was Charlemagne (*q.v.*), who subdued all the hitherto independent German tribes, reduced the Slavonic peoples on the eastern border to a tributary status, conquered the Lombard kingdom in Italy (taking the title of king of the Lombards for himself) and established a Frankish county south of the Pyrenees in Spain. His kingdom now stretched from the Elbe-Saale line in the northeast to the Ebro in the southwest and from the North sea to the marches of Benevento in southern Italy. Except in the British Isles, there was no territory inhabited by Christians in western Europe that did not owe allegiance to Charlemagne. He was acclaimed the "leader of the Christian people" and "defender of Christ's churches." Except for purely spiritual matters, Charlemagne considered himself ruler of both church and state. He used bishops and abbots as officials of the government; and his laws, the capitularies, regulated the affairs of both clergy and laity. His officials and advisers were already beginning to refer to his dominions as a "Christian empire" and to him as successor of Constantine when in the year 800 he went to Rome to put down a local uprising against Pope Leo III. On Christmas day, unexpectedly, Leo crowned him and the assembled clergy and people acclaimed him Augustus and emperor. The actual power and nature of his authority were not changed by the act, which merely recognized and formalized the position he had attained. (*See* HOLY ROMAN EMPIRE.)

5. Carolingian Renaissance.—The revival of cultural activities under Charlemagne and his successors was so impressive that the period from about 775 to about 875 has been called a renaissance. Charlemagne's immediate purpose was to provide a better-educated clergy, as part of his concern for the general welfare of the church. He encouraged monasteries and cathedrals to maintain schools, and he made his court a centre of learning by inviting scholars and writers from Ireland, England, Spain and Italy, as well as Gaul. The most important of these men was Alcuin of York, who later became abbot of St. Martin's of Tours, where he reformed the degenerated handwriting of the times. The new style, called Carolingian minuscule, was employed for the copying of most of the manuscripts of the classical works that have survived (*see* PALEOGRAPHY). It was also in this period that the seven liberal arts were organized into the formal curricula that remained until the 16th century. These were the *trivium* (grammar, rhetoric and dialectic) and the *quadrivium* (arithmetic, geometry, astronomy and music). Very little that was original or profound was produced in the Carolingian renaissance; in the cultural history of Europe it marks a stage in which preservation of the heritage of both pagan and Christian antiquity was the main achievement.

6. Collapse of the Carolingian Empire.—Imposing as it was, Charlemagne's great empire had real weaknesses: geographically it was too large for the rudimentary governmental system; economically it was backward; in communications, transport and military organization it was inferior to the later Roman empire. Also, among the nobility there were deep-rooted desires to control local affairs independently of the central government and the rest of the empire. It was a fatal flaw in the system that it needed a Charlemagne to make it work. These were the underlying causes of the disintegration of the empire after Charlemagne's death. The direct causes were civil wars among his successors and a new wave of barbarian invasions.

The unity of the empire was formally preserved when Charlemagne was succeeded by his only surviving son, Louis the Pious, but long before Louis died his sons were making war over their respective shares of the Frankish inheritance. The principle of unity of the empire, symbolized by the title and supremacy of the emperor and supported by the church, was in conflict with the traditional Frankish principle of equal division of the inheritance of the father among the surviving sons. Louis's eldest son was crowned emperor, but his brothers refused to be deprived of their equal share of territory and they resisted any effort by their older brother to intervene in their kingdoms. On this basis the Wars of the Three Brothers were brought to an end by the treaty of Verdun in 843. Three kingdoms, roughly equal in size and resources, were recognized and their boundaries defined. Louis the German, as king of the eastern Franks, received the area generally east of the Rhine and north of the Alps; Charles the Bald, as king of the western Franks, received the western two-thirds of Gaul; and Lothair, the eldest son, was recognized as emperor and assigned the narrow strip of territory in between, stretching all the way from the Low Countries through Burgundy and Provence and including the Carolingian kingdom of Italy. In 855 this middle kingdom was partitioned equally among the three sons of Lothair on the latter's death. His eldest son, Louis, received the title of emperor and ruled Italy; Charles ruled in the kingdom of Provence; and Lothair II inherited the northern third, called Lotharingia or "Lothair's kingdom" (*see* LORRAINE). On the death of Lothair II in 870 his uncles, Louis the German and Charles the Bald, partitioned Lorraine by the treaty of Mersen. This partition, however, pleased neither side, and the successors of Louis and of Charles waged many wars in their efforts to reunite Lorraine to one or the other kingdom. The Carolingian kingdom of Provence was partitioned between two non-Carolingian claimants: Boso assumed the title of king in the southern part, while Rudolf was recognized as king of Trans-Jurane Burgundy in the north. The two kingdoms were later united as the kingdom of Burgundy under a descendant of Rudolf (*see* BURGUNDY).

The later Carolingians were incapable of ruling over large territories. The medieval habit of dubbing rulers with brutally accurate nicknames suggests the level to which the line had fallen: whereas Charlemagne was "the Great," his descendants of the fourth and fifth generations were "the Simple," "the Stammerer," "the Fat" or "the Child." One cause of their downfall was the necessity of granting crown lands and privileges to nobles in return for their support in the endless wars of the period. The local nobility of a particular region were more obedient to their count or duke than to the ineffective ruler whose Carolingian descent gave him the title of king or emperor.

Scene in the Bernese Oberland, Switzerland, above the valley of Kandersteg near the entrance to the Lötschberg tunnel under the mountains. This area contains some of the highest peaks in the Alpine chain

Flat hinterland of the Netherlands' North Sea coastal belt. This region, called polderland, lies below mean sea level, is crossed by drainage canals and is protected from the sea by dikes and dunes

High striated cliffs near St. Malo typical of the indented rocky Atlantic coast line of Brittany, France

VIEWS OF THE EUROPEAN SEACOAST, UPLANDS AND LOWLANDS

BY COURTESY OF (BOTTOM LEFT) CAS OORTHUYS, PHOTOGRAPHS (TOP) BRADFORD WASHBURN—RAPHO GUILLUMETTE, (BOTTOM RIGHT) KLAUS ESSIKEN

EUROPE

Cima Pordoi, altitude 7,346 ft., one of the peaks of the Dolomites in northern Italy, as seen from Sella pass

Formation of volcanic basalt which forms the Inner Hebrides islands off the western coast of Scotland

Boom of logs on Lake Vänern, one of the large lakes in the central Swedish lowlands between Göteborg and Stockholm

Gently rolling landscape of north central France at Vézelay, part of the Seine river basin

GEOLOGIC FORMATIONS AND RELIEF FEATURES IN EUROPE

PHOTOGRAPHS, (TOP LEFT) J. ALLAN CASH—RAPHO GUILLUMETTE, (TOP RIGHT) KARL-INGVAR NORGREN, (CENTRE LEFT) EOWIN SMITH, (BOTTOM) POUGNET—RAPHO GUILLUMETTE

View of the Danube river from a hilltop sanctuary near Ulm, Germany, where the Iller river enters the main stream. The Danube is the longest river of peninsular Europe, flowing for 1,776 mi. from the Black Forest to the Black sea

Chalk cliffs of Beachy Head, Sussex, facing the English channel. Their height may be judged from the two figures at the top right of the picture

The Rhine river as it bends through Duisburg in the centre of the Ruhr industrial district of Germany. The Rhine is the principal waterway of west central Europe, more than 800 mi. long

TWO PRINCIPAL RIVERS AND A COASTAL SCENE IN EUROPE

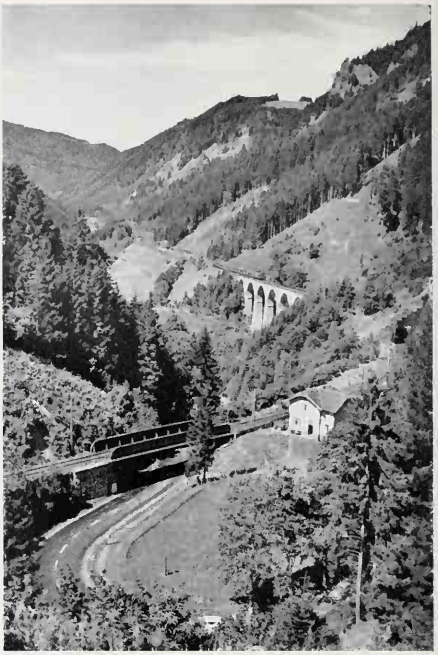
PHOTOGRAPHS, [illegible]

EUROPE

Windermere, largest body of water in the Lake district of England, a domelike area of relief heavily glaciated about 10,000 years ago

View along the Tagus river near Toledo, Spain. The Tagus is one of the principal rivers of the Iberian peninsula, flowing through the tablelands of Portugal and Spain for 625 mi.

Scene in the central Pyrenees, a high range which forms a barrier between the Iberian peninsula and southwestern France. In the background is the Pic de Ger, rising to more than 8,500 ft.



Höllental ("Hell valley") in the Black Forest of southwest Germany, which is a granitic highland covering an area of about 2,320 sq.mi.

Gorge of the Bicaz river, Rumania, in the Ceahlău mountains, part of the Moldavian Carpathians which are characterized by steep forested slopes

RELIEF AND DRAINAGE SYSTEMS OF EUROPE

BY COURTESY OF (BOTTOM RIGHT) THE INSTITUTUL ROMIN, BUCHAREST; PHOTOGRAPHS, (TOP LEFT) AUTHENTICATED NEWS, (TOP RIGHT) FRITZ ESCHEN, (CENTRE LEFT) J. ALLAN CASH—RAPHO GUILLUMETTE, (BOTTOM LEFT) YAN—RAPHO GUILLUMETTE

Meanwhile new barbarian invasions were ravaging western Europe, beginning in the second quarter of the 9th century. From Scandinavia came the heathen Vikings (*q.v.*) or Norsemen. At first contenting themselves with hit-and-run raids on coastal towns, later making more extensive marauding expeditions into the surrounding countryside, they harried the British Isles, northern Germany, northern and western France, Muslim Spain and even the Mediterranean coasts. Then, first in Ireland and later in England and northern France, they made permanent settlements. In these areas (the Danelaw in eastern England and Normandy in France) it took a century and a half before the Norsemen were assimilated to the Christian society they overran. From north Africa came other invaders, the Saracens, whose tactics were similar to those of the Vikings. They established bases on the coast from which to prey upon Christian merchants and pillage neighbouring towns. Sicily was slowly conquered after the first landing in 827. A third wave of invasions came from the east, toward the end of the 9th century. The Magyars swept up the Danube valley and pillaged Lombardy and southern Germany until they were forced back into Hungary and contained.

7. Foundation of the Medieval Empire.—Political recovery from the disintegration that characterized western Europe after the mid-9th century came slowly and took different forms. In England Alfred the Great organized a successful defense against further invasions and won the allegiance of all Anglo-Saxons outside the Danelaw. His successors slowly reconquered the Danelaw and unified all of England under the West Saxon royal house. The strength of the English monarchy was enhanced by the prestige of military conquest; it effectively controlled the local government of the shires under royally appointed sheriffs; and it enjoyed the loyal co-operation of the nobility. In France the monarchy after Charles the Bald (d. 877) declined in prestige as it proved incapable of defending the kingdom or maintaining order. In the 10th and early 11th centuries political stability was rebuilt on the local level by counts and dukes who ignored the king. The crown itself was contested between the later Carolingians and the rising house of Neustria, whose representative, Hugh Capet, finally, late in the 10th century, brought the crown permanently to his family and thus founded the Capetian dynasty. In contrast with the West Saxon kings of England, the Capetian kings in France had almost no influence over most of their kingdom beyond the small royal domain centred on Paris. They were so weak that they aroused no jealousy or resistance from their nominal vassals, the counts and dukes who willingly gave their allegiance.

The most impressive recovery of political stability occurred in Germany, where the last Carolingian, Louis the Child, died in 911 and the greater lords chose one of the powerful dukes, Conrad of Franconia, to be king. When he died, the nobles again elected their new king, Henry I, duke of Saxony, who repulsed the Magyars and laid the foundations of the German monarchy. His son, Otto I the Great, succeeded by what was in fact a hereditary claim though the formalities of an election were observed. In alliance with the church, Otto I stamped out rebellions and enforced his authority over the duchies and principalities that constituted Germany. In 955 he won a great victory over the Magyars, which permanently ended their raids, and in 962 he made good his claim to be king of Italy and was crowned emperor by the pope. The imperial title had been vacant since 924, and this coronation marks the real beginning of the medieval empire that in a later age was called the Holy Roman empire (*q.v.*).

8. Feudalism.—The social and political organization of medieval society was dominated by feudal institutions which developed during the period about 750–1200 and which helped produce the political and social stability that replaced the chaos and disintegration attending the collapse of the Carolingian empire. There was no feudal "system" and nothing systematic about feudal society; each region developed roughly similar feudal institutions, but there was also great variety. Feudal institutions concerned only the greater and lesser nobility. Peasants, merchants and the lower clergy had nothing to do with feudalism except in the sense that their labour supported the nobility. (*See* FEUDALISM.)

Feudal institutions concerned the personal, proprietary and governmental or jurisdictional relations between nobles—in feudal terminology these are vassalage, the fief and "justice" or the jurisdiction of lord over vassal. In both Germanic and Roman society the personal subordination of one person to another was not uncommon; the superior was "lord" while the inferior could be the lord's man or retainer or servant. In the Carolingian period the inferior began to be called a vassal, and the Carolingian kings encouraged warriors to become their vassals to ensure their loyalty and to provide mounted cavalry for the army. The dukes, counts and other nobles imitated the kings, and the hierarchy of lords and vassals began to emerge: a noble, for instance, might become the vassal of a count (who was the king's vassal) and then, in order to provide military service to the count, might have his own men as vassals. To support the vassal, however, the lord would either maintain him as a member of his household or else grant him the temporary use, or usufruct, of a landed estate. The latter was called a benefice. In the later Carolingian period vassalage and the holding of benefices tended to become hereditary, and the term fief began to replace benefice in reference to the hereditary estate of the vassal.

The vassalitic relationship was created by the ceremonial act of homage and a solemn oath of fealty by the vassal to his lord. In return for this homage, fealty and military service, the lord was obliged to provide his vassal with protection and justice. To provide justice each lord held a court for his vassals, where disputes were settled, accusations adjudicated and claims settled. In this court the lord presided but the vassals pronounced judgment; since the vassals of the same lord were all equal, at least in their feudal relation to the lord, this was called judgment by peers.

By the end of the 11th century, feudal institutions dominated and strengthened the social and political organization of northern France and were spreading to other regions. Where these institutions were interconnected or interrelated, a truly feudal society flourished; where they existed but were not interrelated, the society was in a prefeudal and weaker condition.

9. Manorialism.—The great majority of the population were peasants who lived in villages or smaller hamlets or scattered farmsteads. The village was the physical unit of agrarian economy. The administrative or proprietary unit was the manor: the human and other economic resources under the jurisdiction and control of a lord. Village and manor often coincided, the village being entirely under one lord. Often, however, one village was divided into two or more manors and, perhaps less frequently, a single manor might consist of two or more villages or several hamlets or farmsteads. As commerce declined at the end of antiquity and remained stagnant until the mid-11th century, the medieval village tended to be a self-sufficient economic unit dependent on outside supply only for such indispensables as salt or iron (*see* MANOR).

The basis of agriculture was cultivation of the open fields, "open" in that they were not fenced off into separate plots. Husbandry was necessarily co-operative because few peasants owned plows and even fewer possessed a whole plow team of eight oxen. By combining their plows and oxen the villagers could cultivate large fields more efficiently than they could farm small plots with their individual resources. The typical village had two open fields so that fertility could be preserved by rotation. During the 10th and 11th centuries agricultural production was increased by the introduction of three major improvements. The heavy plow, mounted on wheels, sheathed with an iron plowshare and equipped with a moldboard, made cultivation of the heavy, rich soils of the northern European river valleys more efficient. Division of the arable land into three fields for rotation (one lying fallow each year) produced more grain with less plowing, but this was possible only in the most fertile areas because it tended to exhaust the soil. The use of horses instead of oxen in plow teams also increased efficiency where the soil was fertile enough to raise sufficient grain to feed them, so plowing by horses was closely connected with three-field rotation. It was also during this period that the horse collar, horseshoes and tandem harnessing were adopted, three innovations that greatly increased the efficiency of the horse.

The medieval manor belonged to the lord, who possessed not

only proprietary but also jurisdictional rights over it. To his peasants the lord was ruler as well as landlord. In the great hall of his manor house was held the manorial court, where peasants accused of petty crimes were judged and sentenced and where peasants' disputes were settled. The lord's authority, however, was not absolute: the custom of the manor was a force so powerful that it was to the advantage of a lord to observe and enforce it. Though the lord (or his bailiff or steward) presided and pronounced judgment in the manorial court, the peasants participated as suitors who vouched for the custom of the manor and gave testimony in particular cases. Though he had no legal right to his tenement that was recognized or could be enforced outside the manor, the peasant had in fact a high degree of security for his worldly goods and for the land he tilled. The manor was divided into two main parts, the lord's own demesne and the tenements of the peasants. In the open fields demesne strips were intermixed with the peasants' strips of plowed land; the peasants performed labour services regularly on the demesne (week work) throughout the year, and in harvest time they gave their labour to the demesne (boon work) before harvesting crops on their own strips. In addition to these labour services there were many other services and payments, both in kind and in money, varying from region to region and from manor to manor.

10. 12th-Century Revival.—The transition from the early middle ages (*c.* 400–500) to the high middle ages (*c.* 1050–1300) was inaugurated by a widespread political, economic and cultural revival beginning in the middle of the 11th and culminating in the 12th century. Almost all of the institutions and ideas that are characteristically medieval are present in the 12th century, and many of them reached their fullest development in the 13th.

Political Recovery: the Empire and the Papacy.—The medieval empire reached its height under Henry III (1039–56), the greatest emperor of the Salian dynasty. Bohemia, Poland and Hungary acknowledged his suzerainty, and he made his direct rule effective in the kingdom of Burgundy, which his father had annexed to the empire. In Germany he based his power on extensive crown lands; an imperial civil service recruited from the low-born instead of nobles; and the church, which he firmly controlled while he promoted ecclesiastical reform. Within the church, reform was sponsored primarily by the monks of the monastery of Cluny and its daughter houses. The Cluniacs welcomed the co-operation of secular rulers, even when such help also entailed secular influence over the election of bishops and abbots. The papacy was untouched by the Cluniac reform and had fallen on evil days, when the throne of St. Peter was the prize of local Roman noble factions. In 1046 no fewer than three men claimed to be pope, against each of whom various charges ranging from simony to heresy had been brought. Henry III intervened: at the synod of Sutri (1046) he deposed all three and later had his own nominee elected. Under Henry's patronage succeeding popes supported the Cluniacs both in Italy and northern Europe, thus making papal influence effective north of the Alps for the first time since the days of Nicholas I (858–867). (*See* PAPACY.)

A new phase of ecclesiastical reform was inaugurated by Hildebrand, who became Pope Gregory VII (*see* GREGORY). The Gregorian reform party accepted the goals of the Cluniacs but added two more: independence from secular interference in the church and papal leadership of reform. Both issues crystallized in the Investiture struggle carried on by Gregory and his successors against the new emperor, Henry IV (1056–1106), and his successors. A compromise settlement was finally worked out in the concordat of Worms (1122) which limited the influence of lay rulers over episcopal elections and prohibited lay investiture of ecclesiastical office (*see* INVESTITURE CONTROVERSY).

Political Recovery: the Feudal Monarchies.—In England the Norman conquest of 1066 swept away the upper ranks of the Anglo-Saxon nobility and introduced a new feudal nobility from Normandy and northern France. Uniting the strength of the old English monarchy—the king's peace, an advanced system of local government and the power to tax—with the powers that he held as feudal lord of the whole realm, William the Conqueror was the most powerful ruler in western Europe in the most thoroughly feudalized kingdom. Although Normandy and England were divided between his two elder sons, his third son, Henry I, reunited the duchy and kingdom. Because constant warfare on the continent, to protect his duchy, necessitated his frequent absence from England, Henry developed an advanced government, including a specialized department devoted to finances (the exchequer), which could operate under a viceroy, the chief justiciar.

In France the earliest sign of an increase of royal authority appeared under Louis VI (1108–37), the first Capetian to exercise full control over the royal domain and to win the respect of his great vassals. Though unsuccessful in his attempt to wrest Normandy from Henry I of England, he achieved a greater triumph by marrying his heir to the heiress of the duchy of Aquitaine. When Louis VII succeeded to the throne he was the first Capetian king to rule directly a greater territory than any of his vassals.

Expansion of Feudal Europe: the Crusades.—In the two centuries after Charlemagne, when western Europe was disintegrating because of civil strife and barbarian invasions, civilization in the Muslim and Byzantine lands flourished. But at just the time when feudal Europe began to regain social and political stability, the once-brilliant caliphate of Córdoba, the Abbasid caliphate in the east and the Byzantine empire under the Macedonian dynasty were all declining (*see* CÓRDOBA, CALIPHATE OF; ABBASIDS; BYZANTINE EMPIRE).

Under such circumstances the petty Christian rulers in northern Spain began to push back the Moors in a centuries-long struggle, the *Reconquista*, from which the kingdoms of Castile and Aragon emerged as the dominant powers. In southern Italy—an unstable patchwork of principalities under Muslim, Byzantine and Lombard princes—adventurous and unruly knights from the duchy of Normandy sought their fortune and under the lead of Robert Guiscard won control of both the southern mainland and the island of Sicily, which became the kingdom of Sicily in the 12th century. In the east Muslim weakness made possible the conquests of the Seljuks (*q.v.*), who first replaced the Abbasids and then crushed the Byzantine forces in the great battle of Manzikert (1071). After this victory the Seljuks swept over Asia Minor and reduced the Byzantine empire to its European provinces.

To regain his lost provinces in Asia Minor the Byzantine emperor appealed to the west for help, first to secular rulers without success and then to the pope—promising the reunion of the Greek and Latin churches under papal supremacy as the *quid pro quo* by which the pope, Urban II, undertook to raise mercenary troops to reinforce the Byzantine army. At the Council of Clermont (1095) Urban II preached the crusade. The emperor and the kings held back, but the greater and lesser nobility of France and Germany enthusiastically responded to the appeal, some feeling a religious dedication to the liberation of the Holy Land from Muslim domination, others welcoming the opportunity to conquer lands for themselves. This first crusade was successful: Jerusalem was delivered from the infidel, and the crusaders carved out feudal principalities from Muslim Syria and Palestine. (*See* CRUSADES.) But the Byzantine emperor's hopes for reconquest of his lost provinces were dashed and the compact between the emperor and the pope was not fulfilled.

Economic Revival: Towns and Trade.—Although trade never entirely disappeared in the early middle ages, it declined to such a point that the once-thriving cities of the Roman empire became mere shells with only a fraction of their earlier population. The area within the walls was often used for raising crops; the city, except for being the seat of a bishopric or the centre of local administration in the secular government, was little different from a large village.

Urban life began again with the revival of trade in the second half of the 11th century. The cause of this revival was a combination of increasing agricultural production and population growth—two phenomena that were obviously interdependent. But agriculture was vulnerable. If a good season usually produced a surplus, drought or pestilence almost always brought famine. Thus the earliest large-scale, long-distance trade to revive was the sale at high prices in famine-struck areas of food brought from re-

gions where plentiful harvests had lowered prices. The merchants who conducted this trade settled down outside monasteries, castles of friendly lords or other fortified places such as the royal *Burgen* in Germany or the *burghs* (boroughs) in England. These settlements attracted more people whose livelihood depended on the trade of the merchants—artisans, shopkeepers and drivers. As the population grew the town enveloped the original *Burg* and the urban population replaced the original military garrison as the burgers of Germany, the burgesses of England or the *bourgeoisie* of France. Towns grew most rapidly where these mercantile settlements were favourably located with respect to transport and markets, where they were protected by a strong local ruler and where they obtained freedom from the restrictions to which the manorial peasantry was subject. Since neither the clergy nor the nobility engaged in trade, the urban population had to be recruited from the peasants. (*See* COMMUNE.)

The growth of towns provided a new market for surplus food, and manorial lords expanded the area under cultivation. Peasants who provided the labour for this expansion were rewarded with less onerous services and dues, and they also took advantage of the growing circulation of money fostered by trade to commute their obligations into cash payments. Thus, while the lords increased their revenues by selling surplus food in town markets, the peasants benefited even more from the improving economic conditions of the 12th century. In this period technological improvements also played a role in the increase of agricultural production. More new water mills were built between about 1050 and 1200 than in any previous period of comparable duration; windmills were introduced in areas where slight waterfall or absence of streams did not allow water mills; and the introduction of gears made possible new uses for water and wind power (fulling mills, for example, which require an up-and-down motion).

Cultural Revival: the 12th-Century Renaissance.—Though monastic schools were important in the 11th century, cathedral schools, located in the growing towns, soon became the intellectual centres of western Europe. The earliest universities grew out of such schools, first in Italy (where Bologna was the leading centre of the study of law) and then in France (where the University of Paris was pre-eminent by the end of the 12th century). Education was in Latin, the language of religion and learning, common to all Christendom. Most of the students and many of their masters were in minor orders while at the university, from which they might proceed to ordination and a career in the church, or else pursue a career in law, medicine, teaching or the service of a secular government, beginning as a clerk. Both the faculty and students were organized associations. At Paris this association was a "university" of masters; at Bologna the students' association constituted the university. In the north, colleges were organized to provide for students' room and board, and from the tutoring provided for the students there grew up the regular teaching later associated with colleges such as those that constituted the universities of Oxford and Cambridge. (*See* UNIVERSITIES.)

The main achievement of the 12th-century renaissance was the recovery and mastery of the intellectual heritage of both classical and Christian antiquity. The study of Roman law, canon law, patristic literature, logic, philosophy and theology resulted in a new corpus of organized knowledge that the student could learn in the works of such men as Irnerius, Gratian, Anselm, Abelard, John of Salisbury and Peter Lombard. Their work, based mainly on Latin sources, was supplemented by important new works made available by translation from both Arabic and Greek texts of the ancient Greek philosophers, scientists and medical writers, especially the works of Aristotle, which became available in Latin about 1150–1220. It was the achievement of the 13th century to synthesize this corpus of pagan knowledge with Christian doctrine, most notably in the great treatises of St. Thomas Aquinas. In medieval philosophy and theology the distinctive method and the characteristic interests of the greatest writers were those that were taught in the schools. Hence the term scholasticism (*q.v.*) refers to the typical treatises which emphasize formal logic and deductive reasoning. There were schoolmen, however, who were interested in inductive reasoning from the data of sense experience or empirically derived scientific knowledge (the earliest examples of what later became modern scientific method are found in the work of Robert Grosseteste at Oxford).

The period about 1180–1250 witnessed the flowering of medieval art. The noblest achievement was the development of Romanesque architecture (*q.v.*) and, after about 1175, the even more spectacular Gothic (*q.v.*) style. Each was paralleled by a distinctive sculptural style, since medieval sculpture was subordinated to architecture. In the graphic arts, manuscript illumination and illustration were more highly developed than wall painting, but neither reached the perfection of the stained glass "window pictures" which, both aesthetically and in technique, have not been surpassed.

Literature in Latin was not restricted to learned works. Chronicles took the place of the simpler annals of the preceding period; students indulged in anonymous poetry, the goliardic verses which lampooned the clergy and did not spare religion from satire (*see* GOLIARD); and in a more serious vein popular "lives of the saints" and hymns were produced in great number. These works had an international audience because all educated men read Latin. For the uneducated nobility literature was written in the vernacular and was intended for a national audience. Early feudal interests are reflected in the *chansons de geste* (*q.v.*) that immortalize the brave deeds of great heroes, while more sophisticated attitudes and values began to appear in the 12th-century lyric poetry of the troubadours of southern France and in the poetic or prose romances written by the *trouvères* of the north (*see* TROUBADOURS; TROUVÈRE). These portray not only the old chivalric virtues of prowess and honour but also the more recent refinements of courtesy—the courtly ideal or code of behaviour—in which service to the poor and weak, Christian ethics and the notions of gallantry, gentility and generosity replace the simpler ideals of a more crude and primitive age. Most prominent in this revolutionary new courtly literature is the idealization of woman, the feudal lady to whom her dedicated knight proved his homage by the service of love. (*See* COURTLY LOVE.)

11. Holy Roman Empire and the Apogee of the Medieval Papacy.—After the death of Henry V in 1125, Germany suffered from a disputed succession and civil strife that weakened the monarchy. Peace was re-established when Frederick I Barbarossa resolutely set out to restore his regalian rights over the lay nobility and the church in Germany and his imperial authority in Italy. Successful in Germany, his efforts south of the Alps led to stout resistance from the Lombard league of northern communes and a revival of the struggle against the papacy (*see* LOMBARD LEAGUE).

After their victory at the battle of Legnano (1176) the Lombard communes secured almost complete rights of self-government in return for recognition of the emperor's suzerainty. With the pope, Barbarossa came to a compromise that settled none of their conflicting claims but that on balance was a defeat of the emperor's effort to control the papacy and church in Italy as he did in Germany. Papal independence of secular interference was further assured by a decree of the third Lateran council (1179) that a two-thirds majority of the college of cardinals was valid and final in the election of a new pope, without confirmation by the emperor. Barbarossa's only success in Italy was gained by the marriage of his son with the heiress of the kingdom of Sicily.

Under Innocent III (1198–1216) the papacy reached the height of its prestige, leadership and power, and the ideal of the unity of Christendom came nearest to realization. To the papal *curia* appeal lay from all the ecclesiastical courts of Europe, and the courts of archdeacons, bishops and archbishops, steadily enlarging their jurisdiction since the days of Gregory VII, entertained accusations or disputes involving not only all the clergy but all the laity who were *miserabiles personae* (widows, orphans, paupers and others incapable of protecting themselves). The ecclesiastical courts also claimed jurisdiction over all disputes concerning the discipline or administration of the church, property claimed by clergy or ecclesiastical corporate bodies, questions touching marriages, wills, vows and oaths, the sacraments and heresy. Over all these matters the pope was supreme judge (*judex ordinarius*) and even claimed the right to interfere in temporal affairs where

sin might be committed and the salvation of Christian souls endangered—as, for example, when war threatened between the kings of England and France, Innocent intervened as peacemaker in a wholly feudal quarrel. The fourth Lateran council, summoned by Innocent III in 1215, legislated on matters of church reform; condemned the use of superstitious ordeals in all courts, both ecclesiastical and lay; proclaimed a crusade; deposed the emperor Otto IV in favour of Frederick II; condemned the heresy of the Albigensians and excommunicated their lay protector, the count of Toulouse; required all Christians to partake of the sacrament of the Eucharist at least once each year; and pronounced transubstantiation to be a dogma.

Heresy flourished mostly in towns, where ideas as well as commodities from far-off places were traded and where the growth of the population far outstripped the increase of churches and clergy. To minister to these needs and to meet the threat of heresy two new monastic orders were established: the Franciscans and the Dominicans (*qq.v.*), who, instead of retiring from the world, preached in the towns. Where heresy was so strongly entrenched that it had to be repressed, the special papal ecclesiastical court, the Inquisition (*q.v.*), was employed. Dominicans staffed the court; lay rulers were obliged under pain of excommunication to carry out its sentences.

Frederick II (1215–50), who inherited the Kingdom of Sicily from his mother and the Holy Roman empire from his father, renewed the struggle with the papacy. From the conflict both sides emerged exhausted. In Germany imperial authority almost lapsed; in Italy the Lombard towns staved off Frederick's efforts to unite all Italy under his autocratic rule, and the resources of Sicily were drained off by war. The papacy lost prestige by proclaiming a crusade against a Christian monarch and employing spiritual weapons (excommunication or interdict) for worldly ends. The 13th-century popes were able administrators and dedicated reformers, but they failed to preserve the spiritual leadership. After Frederick II's son Conrad there was a period of disputed succession and vacancy in the empire, the Great Interregnum (1254–73) that witnessed the collapse of central authority. Yet the greater prelates and princes of the empire maintained order within their archbishoprics and bishoprics or their duchies, counties and other principalities.

12. Rise of the Western Monarchies.—In England, France and Spain the medieval monarchs founded national states that were to survive, in contrast with the more universal and supreme claims of both papacy and empire, neither of which could preserve the spiritual and temporal authority exercised by earlier popes and emperors. In England the state-builders who accomplished most were Henry II and Edward I. In France, Philip II Augustus (1180–1223) by his conquest of Normandy from John of England not only initiated the growth of the royal domain, which finally included all the kingdom under the king's direct rule, but also precipitated the quarrels between John and his barons that led to the issue of Magna Carta (*q.v.;* 1215). If Philip II won the respect, Louis IX (1226–70) won the love of Frenchmen. St. Louis made the monarchy the focus of national aspiration and pride in France, and his successors made good use of the loyalties he had awakened.

In the 14th and 15th centuries England and France fought the Hundred Years' War (*q.v.*), a conflict in which the English lost all their French possessions (except the port of Calais) and France was ravaged by the pillaging and economic dislocations of war. Yet each monarchy emerged stronger at the end. Emergency conditions in France allowed the kings to assume the power to tax all subjects without consent and to maintain a standing army of professional soldiers. The continual need for money to support the military effort had the opposite effect in England, where parliament assumed more power because the kings had to gain its consent to taxation. The aftereffects of the war disrupted England and led to the Wars of the Roses, the only solution of which was acceptance of a strong monarchy. In Spain, the *Reconquista* was brought to a successful climax under the lead of the kings of Castile and Aragon. After 1270 only the little kingdom of Granada remained unconquered. The marriage of Isabella of Castile and Ferdinand of Aragon in 1469 was followed by the union of Castile and Aragon and the expulsion of the Moors from Granada (1492). Ferdinand and Isabella employed the Inquisition to drive out Moors and Jews and to eliminate political opponents by convicting them of heresy.

13. Eastern Frontiers of Medieval Europe.—At the times of the great migrations (*Völkerwanderung*) the Slavs—referred to by Tacitus, Pliny and Ptolemy under the name of Veneti or Venedi—lived between the middle Vistula to the west, the Carpathians to the south and the Dnieper to the east. In the north their neighbours were the Balts, ancestors of the Old Prussians, Lithuanians and Letts or Latvians. After the departure of the Goths and other Germanic tribes (who moved southward, entering the territories of the Roman empire), the Slavs advanced westward, crossing the Elbe-Saale line in the 5th century, and southward, penetrating into the Danubian basin and the Balkan peninsula. Three centuries later, when Charlemagne conquered and evangelized the lands of the heathen Saxons, the western Slavs found themselves in danger. On the borders of a great empire with a superior culture, they led a pastoral life with a rudimentary social and political organization. A German *Drang nach Osten* (drive to the east) began. The idea of a universal Christian empire inspired it, demographic and economic conditions gave it force. The density of population was greater west of the Rhine than east of the Elbe, and by the end of the 8th century there was little unsettled agricultural land within the Frankish state. The Slavonic tribes living between the Elbe and the Oder—the Polabs (the Obodriti or Bodrici and the Veleti or Lutici) as well as the Lusatian Sorbs—were partly conquered during the 9th century, their conquerors being mainly the Saxons, who had become the outpost of German eastward expansion.

Poland.—To protect them against the fate that had befallen the Polabs and to enable them to defend themselves against the German advance, the Piast dynasty during the 9th century grouped the Polish tribes—Polans, Mazovians, Vislans and Silesians—into one state. The fifth prince of this house, Mieczyslaw or Mieszko I, received Christianity direct from Rome in 966, thus depriving the German emperors of a pretext for a missionary drive into Polish lands east of the Oder. His son, Boleslaw I the Brave, in agreement with the emperor Otto III, secured the independence of the Polish church (A.D. 1000) with an archbishop at Gniezno; in 1024 Boleslaw was crowned king of Poland. But this did not stop the German pressure. More than a century later, in 1107, Archbishop Adelgot of Magdeburg appealed to his compatriots to carry the Christian faith to the Poles in these words:

> They are the worst of pagans, but their land so abounds in the best of meat, honey, corn and all products of the earth that no other land may be compared with it. Wherefore, O Saxons, Franks, Lotharingians and Flemings, you can there both save your souls and gain the best of land in which to live.

For two centuries the Oder was the western frontier of Poland, but in 1135 King Boleslaw III the Wrymouth had to acknowledge the emperor's suzerainty over western Pomerania. Boleslaw died three years later. He divided his kingdom among his four sons, Wladyslaw, the eldest, receiving Silesia. His position as suzerain was enhanced by his tenure of Cracow, the capital, and of Pomerania (Pomorze). By the beginning of the 13th century the Old Prussians and Lithuanians were the only heathen neighbours of Poland. The former resisted conversion by Polish missionaries and often raided Poland instead. Prince Conrad of Mazovia, whose lands suffered most, asked the Teutonic Order (*q.v.*) in 1225 to help in evangelizing Prussia. The offer was accepted and a few decades later Poland had a powerful German state on its northern border. In 1308 the Teutonic Order conquered Pomorze and Danzig. Wladyslaw I the Short reunited the kingdom in 1320 and was crowned king in Cracow, but was unable to dislodge the order from eastern Pomorze. As the western part of this province was transferred in 1225 from the Polish archbishopric of Gniezno to that of Magdeburg, a German corridor separated Poland from the Baltic. This sea became a German lake for two main reasons: first, because by 1227 the present-day Latvia and Estonia had been conquered by the Order of the Knights of the Sword (which in 1237 joined the Teutonic Order); second, because the Hanseatic league (*q.v.*), founded in the mid-13th century by the German

maritime towns to control trade between eastern and western Europe, had become a power able to wage war against Denmark and dictate to King Valdemar IV the peace of Stralsund (1370). At that time more than 50 towns—from Bruges to Reval (Tallinn)—belonged to the league, and the Hanseatic settlements in the east helped the process of germanization.

The Mongol–Tatar invaders who devastated southern Poland in 1241 were stopped and defeated at Legnica by the army of Prince Henry the Pious of Silesia (of the house of Piast), who perished in the battle. To rehabilitate the country, his successors encouraged the immigration of German peasants and artisans who helped the country's economic and cultural advance but also contributed to its germanization. The growing German menace led to a Polish–Lithuanian union under the Jagiello dynasty. Their combined forces defeated the Teutonic Order at Grunwald and secured the return of Pomorze and Danzig to Poland (1466). Poland became a great power in eastern Europe and the *Drang nach Osten* was brought to a halt. (*See* JAGIELLO DYNASTY.)

Bohemia.—Great Moravia, the first western Slavonic state, was founded at the beginning of the 9th century. It included not only the present Czech and Slovak lands but also Lusatia, the southern part of present-day Poland and the western part of present-day Hungary. Great Moravia was weakened by German invasions from the west and, at the beginning of the 10th century, it was destroyed by the Magyars.

In Bohemia proper the Czech Premyslide dynasty founded in the first half of the 10th century a state the first historically known ruler of which was Vaclav (St. Wenceslas), assassinated by his brother in 929. In 1212 Bohemia became part of the Holy Roman empire and German colonization followed. The Premyslide dynasty, extinguished in 1306, was succeeded by the house of Luxembourg. The second ruler of this dynasty—king of Bohemia as Charles I and emperor as Charles IV—resided in Prague, raised its bishops to the rank of archbishops (1344) and founded Prague university (1348). In 1356 he also fixed the attributions and the composition of the imperial college of electors, including the king of Bohemia as one of the four lay members. (The three others were the count palatine of the Rhine, the duke of Saxony and the margrave of Brandenburg; the bishops of Mainz, Trier and Cologne were the ecclesiastical electors.)

Before Charles I's reign the Catholic hierarchy of Bohemia and its *bourgeoisie* were German. A Czech *bourgeoisie* now appeared on the scene, and that marked the beginning of a struggle against German preponderance in the country's cultural and economic life. The movement for church reform headed by John Huss (*q.v.*) was to a great extent a national one. The burning of Huss (1415) was in the eyes of the Czech people an act of collusion between the Germans and the pope. A civil war started in Bohemia (*see* HUSSITES) and it ended with the election of George of Podebrady (1457), a Czech nobleman, as the Hussite king of Bohemia.

Hungary.—The Magyars, after the destruction of Great Moravia, subjugated Slovakia and settled the Danubian plain. Their raids against Germany ended after the defeat they suffered at Lechfeld (955). Géza, a ruler of the Arpád dynasty, was converted to Christianity in 975, but it was under the reign of his son, Stephen I, that Hungary, a multinational state, entered the European community, as a kingdom outside the Holy Roman empire. In 1102 Hungary and Croatia were united under St. Stephen's crown. The Mongol invasion of 1241–42 left the country ravaged and depopulated. When the house of Arpád became extinct in 1301 it was succeeded by the Angevins (1308–82)—a period of power and prosperity for Hungary. In 1345, however, the Ottoman Turks had secured a bridgehead in Europe. In 1363 an allied army lead by King Louis I of Hungary, was defeated by the Turks on the Maritsa river in Thrace. In 1389 the Serbian empire founded by the Nemanya dynasty was destroyed by the Turks at Kosovo. Hungary itself was in danger.

After Louis I's death Hungary had as kings a Luxembourg (Sigismund) and a Habsburg (Albert), who, being elected Roman emperors in succession, were more concerned with the interests of the empire than with those of Hungary. In 1440 Wladyslaw III Jagiello, king of Poland, was crowned as King Uászlo I of Hungary, but in 1444 he fell at Varna, leading a Hungarian-Polish army against the Turks. John Hunyadi, the Hungarian commander who survived the Varna disaster, became regent of Hungary for Albert Habsburg's posthumous son, King László V. When the latter died in 1457 at age 18, Matthias Corvinus became the only Hungarian national king after the Arpáds.

Scandinavian Kingdoms.—The pagan Norsemen discovered toward the 8th century that plundering rich foreign lands was more profitable than extracting a living from the sea or the unfertile soil. While the Swedes turned their attention to the east of the Baltic and gave Russia its first dynasty, the Danes and Norwegians moved westward to Scotland, England and Ireland, and also southward to France, Spain and Sicily. Under Canute the Great (*c.* 995–1035) Britain was part of a Scandinavian empire comprising Denmark, Norway and part of Sweden. Canute became a Christian and supported the evangelization of Scandinavia. After his death his empire fell to pieces. Denmark, Norway and Sweden became independent kingdoms warring among themselves. Between the 12th and 14th centuries Denmark was the leading Scandinavian power. For a time it annexed Estonia, the then Slavonic island of Rügen (Rugia) and part of Pomerania. Sweden undertook in the 12th century the evangelization and conquest of Finland and in 1323 concluded a peace treaty with the Russian duchy of Novgorod that fixed the first eastern frontier of Finland.

In 1397 a union between Denmark, Sweden and Norway was concluded at Kalmar. It was ruled by Margaret (*q.v.*), daughter of a king of Denmark and widow of a king of Norway. The Swedish nobles banished their unpopular king, Albert of Mecklenburg, and elected Margaret as "sovereign lady and ruler." She died in 1412 and was succeeded by Eric of Pomerania. His plan to create a great Baltic empire did not materialize, and the Hanseatic league continued to control the Baltic trade. The Kalmar union continued formally until 1523, although in Sweden there was strong opposition to it.

Russian Duchies Under Mongol Domination.—The invasion of Russia in 862 by Rurik (Hrörekr) and his Norsemen marks the beginning of Russian recorded history. The Norsemen followed the rivers and lakes from the Baltic to the Black sea and established their capital in Kiev. Vladimir Svyatoslavich, grand duke of Kiev (978–1015), probably a direct descendant of Rurik in the third or fourth generation, accepted the Christian faith from Byzantium in 988. There were many Russian duchies and endless family wars waged with the purpose of deciding which among the large number of Rurikovichi was at the time the senior and thus entitled to rule in Kiev, the richest duchy. The Mongol-Tatar invasion of 1237–40 destroyed the Kievan "Rus" and subjected all Russian lands (except Novgorod) to the domination of the Golden Horde (*q.v.*). In spite of Dimitri Donskoi's victory at Kulikovo (1380) it lasted until 1480. (*See* also MONGOL EMPIRES.)

14. The Church in the Later Middle Ages.—The most extreme claims to papal power and ecclesiastical exemption from royal jurisdiction were made by Boniface VIII (1294–1303), who forbade taxation of the clergy without consent of the pope. This precipitated a contest between Boniface and the kings of England and France, each of whom claimed that he had to tax his clergy to defend his realm against attack from the other king. In a brief struggle, replete with vituperative propaganda against Boniface, the monarchs were triumphant and the pope reduced to claims he had no means of enforcing. His successors, French popes, resided at Avignon from 1309 until 1377, when Pope Gregory XI returned to Rome where he died. The cardinals then elected an Italian as the new pope, Urban VI, but then—claiming that election to be invalid because it was made under duress from the Roman mob—elected a new French pope, Clement VII, who promptly returned to Avignon. Thus began the Great Schism (1378–1417) that divided the allegiance of Europe about equally, while each pope excommunicated the other and the other's supporters. This scandalous situation led to demands not only for reunion but for reform of the papacy as well. The great councils held at Pisa (1409), Constance (1414–18) and Basel (1431–37) restored unity by electing Martin V (1417–31) and deposing all other claimants, of whom there were then three. The conciliar movement failed

to achieve significant reforms or to limit papal authority. The price paid for papal victory was high: to woo away the support of secular rulers the popes made concessions that amounted to sharing papal authority with them. But even more costly was the loss of prestige, which the Renaissance popes did nothing to make good. (*See* PAPACY: *The Papacy in Avignon* and *The Great Schism, 1378–1417*; COUNCIL.)

15. Fall of Constantinople.—By the end of the 13th century the last of the crusaders' principalities had fallen to Muslim reconquest, despite the efforts of such leaders as Frederick Barbarossa, Richard the Lion-Hearted, Frederick II and St. Louis, each of whom led a crusade without lasting result. In 1291 Acre, the last city of the kingdom of Jerusalem, fell and only the island of Cyprus remained in Christian hands. In the 14th century the Ottoman Turks swept westward, attacking Muslim, Byzantine and independent territories in both Asia and Europe. They crushed the greatest Balkan power, Serbia, and overran the rest of the peninsula. They were then free to complete the conquest of the rest of Asia Minor and to lay siege to Constantinople. The great fortress city fell in 1453. This event hardly changed the situation in the Levant—Italian traders continued to enter the ports and markets, though Venetian trade suffered.

16. Italy and Germany in the Later Middle Ages.—Both north and south of the Alps political disintegration of the Holy Roman empire was accompanied by the growth of smaller states. In Italy these were the city-states, dominated by the wealthy merchants. Venice, Milan, Florence and Rome were the most important, while the kingdom of Naples was still a power. Though many of the city-states had republican forms of government, almost all were in fact governed either by a small oligarchy or by a single despot. The Renaissance popes, in their policies and their plots and intrigues, were hardly distinguishable from other despots. In Germany the situation is summed up in the term particularism—*i.e.*, the division of the country into autonomous principalities and self-governing free cities. The emperor was elected by seven electoral princes and, after 1356, papal claims to adjudicate a disputed election were explicitly denied. After 1438 the Habsburg dynasty once more came to the imperial throne, this time to retain the title permanently. Socially and culturally Italy and Germany present a striking contrast at the end of the middle ages. In northern Europe society was still dominated by the rural nobility, in Italy by the urban *bourgeoisie*. In arts and letters, as in education, medieval styles, interests and methods continued in the north, while in Italy the Renaissance brought a revival of the literature, artistic styles and interests of classical antiquity. (R. S. Ho.)

D. The Emergence of Modern Europe, 1492–1648

1. Modern Europe and the Renaissance.—In the 15th century changes in the structure of European polity, accompanied by a new intellectual temper, suggested to such men as the philosopher and clerical statesman Nicholas of Cusa that the "middle age" had attained its conclusion and a new era had begun. The papacy, the symbol of the spiritual unity of Christendom, lost much of its prestige in the schism and the conciliar movement, and became infected with the lay ideals prevailing in the Italian peninsula. In the 16th century the Reformation reacted against the worldliness and corruption of the Holy See, and the Catholic Church responded in its turn by a revival of piety known as the Counter-Reformation. While the forces that were to erupt in the Protestant movement were gathering strength, the narrow horizons of the old world were widened by the expansion of Europe to America and the east.

In western Europe nation-states emerged under the aegis of strong monarchical governments, breaking down local immunities and destroying the unity of the European *Respublica Christiana*. Centralized bureaucracy came to replace medieval government. Underlying economic changes affected social stability. Secular values prevailed in politics, and the concept of a balance of power came to dominate international relations. Diplomacy and warfare were conducted by new methods. Permanent embassies were accredited between sovereigns, and on the battlefield standing armies of professional and mercenary soldiers took the place of the feudal array which had reflected the social structure of the past. At the same time scientific discoveries cast doubt on the traditional cosmology. The systems of Aristotle and Ptolemy, which had for long been sanctified by clerical approval, were undermined by Copernicus, Mercator, Galileo and Kepler.

The movement termed the Renaissance (*q.v.*) substituted new standards of culture and methods of thought for the Gothic styles and scholastic exercises of the past. The harmony of classical models was preferred to the disordered grandeur of medieval architecture, while in art pagan realism and the worship of the human form replaced religious symbolism, even in the representation of Biblical subjects. The revival of interest in ancient Greece and Rome was stimulated by an influx of Byzantine learning after the conquest of Constantinople by the Ottomans in 1453. The subsequent invention of the printing press popularized the rebirth of classical ideals and heightened the effect of scholarly criticism of religious and pseudo-historical dogmas. In both its rational, humanist aspect and its mystic, neoplatonic guise, the Renaissance expanded from its centre in the petty despotisms of Italy to all the confines of Europe.

2. The New World and Economic Development.—In the Iberian peninsula the impetus of the counteroffensive against the Moors carried the Portuguese to probe the west African coast line and the Spanish to attempt the expulsion of Islam from the western Mediterranean. In the last years of the 15th century Portuguese navigators established the sea route to India, and within a decade had secured control of the trade routes in the Indian ocean and its approaches. Mercantile interests, crusading and missionary zeal, and scientific curiosity were intermingled as the motives for this epic achievement. Similar hopes inspired Spanish exploitation of the discovery by Christopher Columbus of the Caribbean outposts of the American continent in 1492. The treaties of Tordesillas and Saragossa in 1494 and 1529 defined the limits of westward Spanish exploration and the eastern ventures of Portugal. The two states acting as the vanguard of the expansion of Europe had thus divided the newly discovered sea lanes of the world between them. (*See* also GEOGRAPHY: *Opening of the Oceans*.)

By the time of the treaty of Saragossa, when Portugal secured the exclusion of Spain from the East Indies, Spain had begun the conquest of Central and South America. In 1519, the year in which Ferdinand Magellan embarked on the westward circumnavigation of the globe, Hernando Cortes launched his expedition against Mexico. The seizure of Peru by Francisco Pizarro and the enforcement of Portuguese claims to Brazil completed the major steps in the Iberian occupation of the continent. By the middle of the century the age of the *conquistadores* was replaced by an era of colonization, based both on the procurement of precious metal by Indian labour and on pastoral and plantation economies using imported Negro slaves. The influx of bullion into Europe became significant in the late 1520s, and from about 1550 it began to produce a profound effect upon the economy of the old world.

Economic nationalism in the 16th and 17th centuries expressed itself in the theory and practice of mercantilism (*see* MERCANTILE SYSTEM). There was widespread belief that value was intrinsic in gold and silver, and the acquisition of bullion by the maximum export and minimum import of goods was therefore regarded as the key to national prosperity. Even those who apprehended its function as a means of exchange sought to restrict the credit of rival states by accumulating as much bullion as possible. In their control of the silver supply, the Habsburg rulers of Spain and the Holy Roman empire appeared to be in a commanding position, and certainly the importation of bullion from the new world gave the Habsburgs a political and military flexibility which contributed to their predominance in Europe in the period under consideration. One-fifth of the precious metal mined was payable to the Spanish monarchy, although in practice this was greatly diminished by administrative costs. However, mercantilism was never effective in the age of the great merchant-bankers, upon whom the most powerful sovereigns were often dependent. The Fuggers (*q.v.*) of Augsburg did not hesitate to remind Charles V that his election as emperor resulted from their support. Much of the specie from the Indies went directly into the keeping of international financiers

to repay loans. Despite attempts to retain bullion in Spain it continued to pass from Seville to Antwerp, the free centre of European exchange. Thence it affected the economies of other states. Where it was not possible to tap the supply by legitimate trade and financial loans, piracy was frequently employed by the French, English and Dutch against Spanish treasure ships.

The influx of American silver was the principal factor in the accelerated rise in prices during the century after 1550. Spain, as the direct recipient of the new wealth, was the European state most severely affected, and by 1600 prices were approximately five times their level a century before. Silver tended to find its way into the possession of the landed aristocracy and was not a stimulant to the commercial activity of the merchant class. In consequence Spain suffered a general increase in wages and a shortage of goods. This was not the effect produced by the price revolution elsewhere. In France prices rose steeply, but not to the same degree, while in England the full social effect was not experienced until the 17th century. Portugal possessed the most advanced means of controlling exchange and commerce in the interests of the state. In western Europe, outside the Iberian peninsula, the price rise often favoured private merchant interests, and worked to the detriment of those who were dependent upon relatively stable rents from landed property. This "profit inflation" helped to create conditions for the embryonic capitalist growth of the period. It did not assist governments, which, like the country nobility, were largely dependent upon fixed incomes. Moreover, it invited governmental debasing of coinage—an expedient often mistakenly assumed to be in itself the cause of the price rise. Further disruption was caused by the general alteration in the ratio of gold to silver values, the latter being in far more plentiful supply.

The expansion of Europe was accompanied by a sweeping change in the pattern of trade. It was not so much the appearance of the Turk in the eastern Mediterranean as the Portuguese control of the Indian ocean which interrupted the Levantine spice trade in the first half of the 16th century. Spices found their way to Lisbon during the period of Portuguese monopoly, and thence were carried to Antwerp. Although the Ottoman Turks continued to grant commercial privileges to Venice when they were not at war with that republic, the new Islamic offensive in the Mediterranean hastened the shift of the European economic balance toward the Atlantic and the north. Toward the close of the 16th century the commercial expansion of the English and Dutch brought new forces into play. Antwerp declined during the Wars of Religion, and Portugal fell under Spanish domination from 1580 to 1640. Dutch trading wealth from the herring industry and the produce of the Baltic was augmented by establishments in the East and West Indies and the wresting of the slave market from the Lisbon monopolists.

The power of international financiers began to decline in the last decades of the 16th century. Earlier bankruptcies and repudiations on the part of European monarchies had revealed the instability of the international money market. The operations of the Florentine and Luccan merchant-bankers, who had succeeded to the role of the Fuggers, were increasingly restricted. National monarchies began to appreciate the need to control national credit, and Paris became more important as the centre of government finance than the international mart of Lyons. The new 17th-century Bank of Amsterdam was directed in the Dutch national interest. The great financiers had been internationalists in an age of nationalism, yet their operations, particularly the development of such expedients as the bill of exchange, had been indispensable alike to local merchant interests and the governments they served.

3. Nation-States and Dynastic Rivalries.—The organization of expansion overseas reflected in economic terms the political nationalism of the European states. This political development took place through processes of internal unification and the abolition of local privileges by the centralizing force of dynastic monarchies. In Spain the union of Aragon, Valencia and Catalonia under John II of Aragon was extended to association with Castile through the marriage of his son Ferdinand with the Castilian heiress Isabella. The alliance grew toward union after the accession of the two sovereigns to their thrones in 1479 and 1474 respectively and with joint action against the Moors of Granada, the French in Italy and the independent kingdom of Navarre. Yet at the same time provincial institutions long survived the dynastic union, and the representative assembly (*Cortes*) of Aragon continued to cling to its privileges when its Castilian counterpart had ceased to play any effective part. Moreover, Castilian interest in the new world and Aragonese ties in Italy resulted in the ambivalent nature of Spanish 16th-century policy, with its uneasy alternation between the Mediterranean and the Atlantic. The monarchy increased the central power by the absorption of military orders and the adaptation of the *Hermandad*, or police organization, and the Inquisition for political purposes. During the reign of Charles I (the emperor Charles V) centralization was quickened by the importation of Burgundian conciliar methods of government and in the reign of his son Philip II Spain was in practice an autocracy.

Other European monarchies imitated the system devised by Roman law jurists and administrators in the Burgundian dominions along the eastern borders of France. In England and France the Hundred Years' War had reduced the strength of the aristocracies, the principal opponents of monarchical authority. The pursuit of strong, efficient government by the Tudors in England, following the example of their Yorkist predecessors, found a parallel in France under Louis XI and Francis I. In both countries revision of the administrative and judicial system proceeded through conciliar institutions, although in neither case did it result in the unification of different systems of law. A rising class of professional administrators came to fulfill the role of the king's executive. The creation of a central treasury under Francis I brought an order into French finances already achieved in England through Henry VII's adaptation of the machinery of the royal household. Henry VIII's minister, Thomas Cromwell, introduced an aspect of modernity into English fiscal administration by the creation of courts of revenue on bureaucratic lines. In both countries the monarchy extended its influence over the government of the church. The unrestricted ability to make law was established by the English crown in partnership with parliament; in France the representative estates-general lost its authority and sovereignty reposed in the king in council. Supreme courts (*parlements*) possessing the right to register royal edicts, imposed a slight and ineffective limitation on the absolutism of the Valois kings. The most able exponent of the centralizing and autocratic tendencies of the French monarchy was Charles IX's chancellor, Michel de L'Hôpital, but his reforms in the 1560s were frustrated by the anarchy of the religious wars. In France the middle class aspired to ennoblement in the royal administration, and mortgaged their future to the monarchy by investment in office and the royal finances. In England, on the other hand, a greater flexibility in social relationships was preserved, and the middle class engaged in bolder commercial and industrial ventures.

Territorial unity under the French crown was attained through the recovery of feudal appanages alienated to cadet branches of the royal dynasty. It was also accomplished, as in Spain, through marriage alliances. Brittany was regained in this way, although the first of the three Valois marriages with Breton heiresses also set in train the dynastic rivalry of Valois and Habsburg. Deprived of his bride by the marriage of Charles VIII of France and Anne of Brittany, the Austrian archduke and subsequent emperor, Maximilian, married the heiress of Charles the Bold of Burgundy. In turn their son Philip married Joan, the heiress of Ferdinand and Isabella, so that the elder son of the union inherited the Habsburg domains in Spain, Italy, America, Germany, eastern France and the Netherlands. The dominions of Charles V thus encircled France and incorporated the wealth of Spain overseas. Even after the division of this vast inheritance between his son, Philip II of Spain, and his brother, the emperor Ferdinand I, the conflict between the Habsburgs and the French crown dominated the diplomacy of Europe for more than a century.

The principal dynastic conflict of the age was less unequal than it seemed, for the greater resources of Charles V were offset by their cumbrous disunity and by local independence. In the Low Countries he was able to complete the Seventeen Provinces by new acquisitions, but, although the co-ordinating machinery of the

Burgundian dukes remained in formal existence, Charles's regents were obliged to respect local privileges and to act through constitutional forms. In Germany, where his grandfather Maximilian had unsuccessfully tried to reform the constitution of the Holy Roman empire, Charles V could do little to overcome the independence of the lay and ecclesiastical princes, the imperial knights and the free cities. The revolts of the knights (1522) and the peasantry (1525), together with the political disaggregation imposed by the Reformation, rendered the empire a source of weakness. Even in Spain, where the rebellion of the *comuneros* took place in 1520–21, his authority was sometimes flouted. His allies, England and the papacy, at times supported France to procure their own profit. France, for its part, possessed the advantages of internal lines of communication and a relatively compact territory, while its alliance with the Turk maintained pressure on the Habsburg defenses in southeast Europe and the Mediterranean. However, Francis I, like his predecessors Charles VIII and Louis XII, made the strategic error of wasting his strength in Italy, where the major campaigns were fought in the first half of the century. Only under Henry II was it appreciated that the most suitable area for French expansion lay toward the Rhine.

4. Turkey and Eastern Europe.—A contemporary who rivaled the power and prestige of Francis I and Charles V was the ruler of the Ottoman empire, the sultan Suleiman the Magnificent (1520–66). With their infantry *corps d'élite* (the janizaries) their artillery and their cavalry or *sipahis*, the Ottomans were the foremost military power in Europe, and it was fortunate for their Christian adversaries that eastern preoccupations prevented them from taking full advantage of western disunity. A counterpoise was provided by the rise of the powerful military order of the Safawids in Persia—hostile to the orthodox Ottomans through their acceptance of the heretical Islamic cult of the Shi'a. Ottoman strength was further dissipated by the need to enforce the allegiance of Turkoman begs in Anatolia and of the chieftains of the Caucasus and Kurdistan, and to maintain the conquest of the Mameluke sultanate of Syria and Egypt by Suleiman's predecessor, Selim I. Suleiman himself overran Iraq, and even challenged Portuguese dominion of the Indian ocean from his bases in Suez and Basra. The Crimean Tatars acknowledged his suzerainty, as did the corsair powers of Algiers, Tunis and Tripoli. His armies conquered Hungary in 1526 and threatened Vienna in 1529. With the expansion of his authority along the north African coast and the Adriatic littoral, it seemed for a time as if the Mediterranean, like the Black sea and the Aegean, might become an Ottoman lake.

Though it observed the forms of an Islamic legal code, Turkish rule was an unlimited despotism, suffering from none of the financial and constitutional weaknesses of western states. With its disciplined standing army and its tributary populations, the Ottoman empire feared no internal threat except the periods of disputed succession, which continued to occur despite a law empowering the reigning sultan to put to death collateral heirs. It was not unusual for the sultan to content himself with the overlordship of frontier provinces. Moldavia and Wallachia were for a time held in this fashion, and in Transylvania the *voivode* John Zápolya gladly accepted Suleiman as his master in return for support against Ferdinand of Austria.

Despite the expeditions of Charles V against Algiers and Tunis, and the inspired resistance of Venice and Genoa in the war of 1537–40, the Ottomans retained the initiative in the Mediterranean until several years after the death of Suleiman. The Knights of St. John were driven from Rhodes and Tripoli and barely succeeded in retaining Malta. Even after the combined fleet of Spain, the papacy, Venice and Genoa had crushed the Turkish armament in 1571 in the battle of Lepanto (*q.v.*), the Ottomans took Cyprus and recovered Tunis from the garrison installed by the allied commander, Don John of Austria. North Africa remained an outpost of Islam and its corsairs continued to harry Christian shipping, but the Ottoman empire did not again threaten Europe by land and sea until late in the 17th century.

In eastern Europe the states of Poland, Lithuania, Bohemia and Hungary were all loosely associated at the close of the 15th century under rulers of the house of Jagiello. In 1569, three years before the death of the last Jagiello king of Lithuania–Poland, the two latter countries merged their separate institutions by the union of Lublin. Thenceforth the Polish nobility and the Roman Catholic faith dominated the Orthodox lands of Lithuania and held the frontiers against Muscovy, the Cossacks and the Tatars. Bohemia and the vestiges of independent Hungary were regained by the Habsburgs as a result of dynastic marriages, which the emperor Maximilian planned as successfully in the east as he did in the west. When Louis II Jagiello of Hungary died fighting the Ottomans at Mohács in 1526, the archduke Ferdinand of Austria obtained both crowns and endeavoured to affirm the hereditary authority of his dynasty against aristocratic insistence on the principle of election. In 1619 Habsburg claims in Bohemia became the ostensible cause of the Thirty Years' War (*q.v.*), when the diet of Prague momentarily succeeded in deposing Ferdinand II.

In the 16th century eastern Europe displayed the opposite tendency to the advance of princely absolutism in the west. West of the Carpathians, and in the lands drained by the Vistula and the Dniester, the landowning class achieved a political independence that weakened the power of monarchy. The towns entered a period of decline, and the propertied class, though divided by rivalry between the magnates and the lesser gentry, everywhere reduced their peasantry to servitude. In Poland and Bohemia the peasants' were reduced to serfdom in 1493 and 1497 respectively, and in free Hungary the last peasant rights were suppressed after the rising of 1514. The gentry or *szlachta* controlled Polish policy in the *sejm*, and when the first Vasa king, Sigismund III, tried to reassert the authority of the crown after his election in 1587, the opportunity had passed. Yet despite the anarchic quality of Polish politics, the aristocracy maintained and even extended the boundaries of the state. In 1525 they compelled the submission of the secularized Teutonic Order in East Prussia, resisted the pressure of Muscovy, and pressed to the southeast, where communications with the Black sea had been closed by the Ottomans and their tributaries.

Farther to the east the grand duchy of Muscovy emerged as a new and powerful despotism. Muscovy, and not Poland, became the heir to Kiev during the reign of Ivan the Great in the second half of the 15th century. By his marriage with the Byzantine princess Zoe Palaeologue, Ivan also laid claim to the traditions of Constantinople. His capture of Novgorod and repudiation of Tatar overlordship began a movement of Muscovite expansion, which was continued by the seizure of Smolensk by his son Basil III and by the campaigns of his grandson Ivan the Terrible (1533–84). The latter destroyed the khanates of Kazan and Astrakhan, and reached the Baltic by his conquest of Livonia from Poland and the Knights of the Sword. He was the first to use the title of tsar, and his arbitrary exercise of power was more ruthless and less predictable than that of the Ottoman sultan. After his death Muscovy was engulfed in the "time of troubles," when Polish, Swedish and Cossack armies devastated the land. The accession of the Romanov dynasty in 1613 heralded a period of gradual recovery. Except for occasional embassies, the importation of a few western artisans and the reception of Tudor trading missions, Muscovy remained isolated from the west. Despite its relationship with Greek civilization, it knew nothing of the Renaissance. Though it experienced a schism within its own Orthodox faith, it was equally untouched by Reformation and Counter-Reformation, the political consequences of which convulsed western Europe in the second half of the 16th century.

5. Reformation and Counter-Reformation.—In a sense the Reformation (*q.v.*) was a protest against the secular values of the Renaissance. No Italian despots better represented the profligacy, the materialism and the intellectual hedonism that accompanied these values than did the three Renaissance popes, Alexander VI, Julius II and Leo X. Among those precursors of the reformers who were conscious of the betrayal of Christian ideals were figures so diverse as the fiery Ferraran monk Savonarola, the Spanish statesman Cardinal Jiménez and the humanist scholar Erasmus (*qq.v.*). The corruption of the religious orders and the cynical abuse of the fiscal machinery of the church provoked a movement that at first demanded reform from within, and ultimately chose

the path of separation. When the Augustinian monk Martin Luther (*q.v.*) protested against the sale of indulgences in 1517, he found himself obliged to extend his doctrinal arguments until his stand led him to deny the authority of the pope. In the past, as in the controversies between pope and emperor, such challenges had resulted in mere temporary disunity. In the age of nation-states the political implications of the dispute resulted in the irreparable fragmentation of clerical authority.

Luther had chosen to attack a lucrative source of papal revenue, and his intractable spirit obliged Leo X to excommunicate him. The problem became of as much concern to the emperor as it was to the pope, for Luther's eloquent writings evoked a wave of enthusiasm throughout Germany. The reformer was by instinct a social conservative and supported existing secular authority against the upthrust of the lower orders. Although the diet of Worms accepted the excommunication in 1521, Luther found protection among the princes. In 1529 the rulers of electoral Saxony, Brandenburg, Hesse, Luneberg and Anhalt signed the "protest" against an attempt to enforce obedience. By this time Charles V had resolved to suppress Protestantism and to abandon conciliation. In 1527 his mutinous troops had sacked Rome and secured the person of Pope Clement VII, who had deserted the imperial cause in favour of Francis I after the latter's defeat at the battle of Pavia. The sack of Rome proved a turning point both for the emperor and the humanist movement he had patronized (*see* HUMANISM). The humanist scholars were dispersed and the initiative for reform thenceforth lay in the hands of the more violent and uncompromising party. Charles V himself experienced a revulsion of conscience which placed him at the head of the Catholic reaction. The empire he ruled in name was now divided into hostile camps. The Catholic princes of Germany had discussed measures for joint action at Ratisbon in 1524; in 1530 the Protestants formed a defensive league at Schmalkalden (*see* SCHMALKALDEN, LEAGUE OF). Reconciliation was attempted in 1541 and 1548, but the German rift could no longer be healed.

Lutheranism (*q.v.*) laid its emphasis doctrinally on justification by faith and politically on the God-given powers of the secular ruler. Other Protestants reached different conclusions and diverged widely from one another in their interpretation of the sacraments. In Geneva, Calvinism (*q.v.*) enforced a stern moral code and preached the mystery of grace with predestinarian conviction. It proclaimed the separation of church and state, but in practice its organization tended to produce a type of theocracy. Huldreich Zwingli and Heinrich Bullinger in Zürich taught a theology not unlike Calvin's, but preferred to see government in terms of the godly magistrate. On the left wing of these movements were the Anabaptists (*q.v.*), whose pacifism and mystic detachment were paradoxically associated with violent upheavals.

Lutheranism established itself in north Germany and Scandinavia, and for a time exercised a wide influence both in eastern Europe and in the west. Where it was not officially adopted by the ruling prince, however, the more militant Calvinist faith tended to take its place. Calvinism spread northward from the upper Rhine and established itself firmly in Scotland and in southern and western France. Friction between Rome and nationalist tendencies within the Catholic Church facilitated the spread of Protestantism. In France the Gallican Church was traditionally nationalist and antipapal in outlook (*see* GALLICANISM), while in England the Reformation in its early stages took the form of the preservation of Catholic doctrine and the denial of papal jurisdiction. After periods of Calvinist and then of Catholic reaction, the Church of England achieved a measure of stability with the Elizabethan religious settlement.

In the years between the papal confirmation of the Jesuit order in 1540 and the formal dissolution of the Council of Trent in 1563, the Catholic Church responded to the Protestant challenge by purging itself of the abuses and ambiguities that had opened the way to revolt. Thus prepared, the Counter-Reformation embarked upon the recovery of the schismatic branches of western Christianity. Foremost in this crusade were the Jesuits, established as a well-educated and disciplined arm of the papacy by Ignatius Loyola (*see* JESUS, SOCIETY OF). Their work was made easier by the Council of Trent (*q.v.*), which did not, like earlier councils, result in the diminution of papal authority. The council condemned such abuses as pluralism, affirmed the traditional practice in questions of clerical marriage and the use of the Bible, and clarified doctrine on controversial issues such as the nature of the Eucharist, divine grace and justification by faith. The church in this way made it clear that it was not prepared to compromise, and with the aid of the Inquisition and the material resources of the Habsburgs it set out to re-establish its universal authority. It was of vital importance to this task that the popes of the Counter-Reformation were men of sincere conviction and initiative, who skilfully employed diplomacy, persuasion and force in the offensive against heresy. In Italy, Spain, Bavaria, Austria, Bohemia, Poland and the southern Netherlands, Protestant influence was destroyed. As the movement gathered strength in the age of religious warfare, its course became confused with the national ambitions of Philip II of Spain. (*See* also ROMAN CATHOLIC CHURCH: *Reform and Definition.*)

6. The Wars of Religion.—Germany, France and the Netherlands each achieved a settlement of the religious problem by means of war, and in each case the solution contained original aspects. In Germany the territorial formula of *cuius regio, eius religio* applied—that is to say, in each petty state the population had to conform to the religion of the ruler. In France, the Edict of Nantes in 1598 embraced the provisions of previous treaties and accorded the Huguenots toleration within the state, together with the political and military means of defending the privileges they had exacted. The southern Netherlands remained Catholic and Spanish, but the Dutch provinces formed an independent Protestant federation in which republican and dynastic influences were nicely balanced. Nowhere was toleration accepted as a positive moral principle, and seldom was it granted except through political necessity.

There were occasions when the Wars of Religion assumed the guise of a supranational conflict between Reformation and Counter-Reformation. Spanish, Savoyard and papal troops supported the Catholic cause in France against Huguenots aided by Protestant princes in England and Germany. In the Low Countries, English, French and German armies intervened, and at sea Dutch, Huguenot and English corsairs fought the battle of the Atlantic and the Narrow seas against the Spanish champion of the Counter-Reformation. In 1588 the fate of the Spanish Armada (*q.v.*) against England was intimately connected with the progress of the struggles in France and the Netherlands.

Behind this ideological grouping of the powers, national, dynastic and mercenary interests generally prevailed. The Lutheran duke Maurice of Saxony assisted Charles V in the first Schmalkaldic war in 1547 in order to win the Saxon electoral dignity from his Protestant cousin, John Frederick; while the Catholic Henry II of France supported the Lutheran cause in the second Schmalkaldic war in 1552 to secure French bases in Alsace. John Casimir of the Palatinate, the Calvinist champion of Protestantism in France and the Low Countries, maintained an understanding with the neighbouring princes of Lorraine, who led the Ultramontane party, the Holy League, in France. In the French conflicts Lutheran German princes served against the Huguenots, and mercenary armies on either side often fought against the defenders of their own religion. On the one hand deep divisions separated Calvinist from Lutheran, and on the other political considerations persuaded the Catholic *politique* faction to oppose the Holy League. The national and religious aspects of the foreign policy of Philip II of Spain were not always in accord. Mutual distrust existed between him and his French allies, the family of Guise, because of their ambitions for their niece Mary Stuart. His desire to perpetuate French weakness through civil war led him at one point to negotiate with the Huguenot leader, Henry of Navarre. His policy of religious uniformity in the Netherlands alienated the most wealthy and prosperous part of his dominions. Finally, his ambition to make England and France the satellites of Spain weakened his ability to suppress Protestantism in both countries.

In 1562, seven years after the peace of Augsburg (*see* AUGSBURG, PEACE OF) had established a truce in Germany on the basis of

territorialism, France became the centre of religious wars which endured, with brief intermissions, for 36 years. The political interests of the aristocracy and the vacillating policy of balance pursued by Henry II's widow, Catherine de' Medicis, prolonged these conflicts. After a period of warfare and massacre, in which the atrocities of St. Bartholomew's day (1572) were symptomatic of the fanaticism of the age, Huguenot resistance to the crown was replaced by Catholic opposition to the monarchy's policy of conciliation to Protestants at home and anti-Spanish alliances abroad. The revolt of the Holy League against the prospect of a Protestant king in the person of Henry of Navarre released new forces among the Catholic lower classes, which the aristocratic leadership was unable to control. Eventually Henry won his way to the throne after the extinction of the Valois line, overcame separatist tendencies in the provinces and secured peace by accepting Catholicism. The policy of the Bourbon dynasty resumed the tradition of Francis I, and under the guidance of Cardinal Richelieu the potential authority of the monarchy was realized.

In the Netherlands the wise Burgundian policies of Charles V were largely abandoned by Philip II and his lieutenants. Taxation, the Inquisition and the suppression of privileges for a time provoked the combined resistance of Catholic and Protestant. The house of Orange, represented by William the Silent and Louis of Nassau, acted as the focus of the revolt, and in the undogmatic and flexible personality of the former the rebels found leadership in many ways similar to that of Henry of Navarre. The sack of the city of Antwerp by mutinous Spanish soldiery in 1576, three years after the dismissal of Philip II's autocratic and capable governor the duque de Alba, completed the commercial decline of Spain's greatest economic asset. In 1579 Alessandro Farnese, duke of Parma, succeeded in recovering the allegiance of the Catholic provinces, while the Protestant north declared its independence. French and English intervention failed to secure the defeat of Spain, but the dispersal of the Armada and the diversion of Parma's resources to aid the Holy League in France enabled the United Provinces to survive. A 12-year truce was negotiated in 1609, and when the campaign began again it merged into the general conflict of the Thirty Years' War. It was during this last phase of the struggle that Amsterdam came to replace Antwerp as the trading centre of Europe.

7. Sweden and the Thirty Years' War.—In both its national and supranational aspects the Thirty Years' War (*q.v.*) in Germany resembled the French Wars of Religion on a wider and more destructive scale. In the first phase of the conflict the emperor Ferdinand II, who became as determined a champion of the Counter-Reformation as his cousin Philip II of Spain, allied himself with Maximilian of Bavaria to recover the throne of Bohemia from Frederick V, the Calvinist elector palatine. All Europe, directly or indirectly, became involved in the developments that followed. The opposing mercenary armies of Mansfeld and Tilly were supported by Protestant and Catholic and equally by anti-Habsburg and Habsburg interests extending from England to Savoy and from Spain to the remote lands of Transylvania. After the triumph of the emperor, Denmark intervened in the next phase of the conflict (1625–29) in an attempt to redress the balance. The defeat of the Danes was followed by the march of the greatest of the *condottieri*, Albrecht von Wallenstein (*q.v.*), to the shores of the Baltic, and this in its turn moved the Swedish king, Gustavus Adolphus, to defend the Protestant cause.

The rise of Sweden had been prepared by Gustavus Vasa (1523–60), who had won independence from Denmark and established a variant of the Lutheran branch of the Reformation. In 1611 his grandson, Gustavus II (*q.v.*) Adolphus, had inherited a kingdom which, though having a population of barely 1,000,000, was ready to expand southward to procure control of the Baltic. Before his intervention in Germany Gustavus Adolphus had fought three wars in which he had contained Denmark and gained possession of the Baltic coast line held by Poland and Russia. He combined the learning of a Renaissance prince with the zeal of a Protestant reformer and, more necessary to his immediate task, a mastery of military tactics that surpassed the skill of all his opponents. His ambitions extended beyond the Baltic, and, while he was prepared to accept the subsidies of Richelieu, he knew when to prefer his own interests to those of France. His army carried all before it until Wallenstein, recalled from disgrace, opposed his march near Nürnberg in 1632. Although Gustavus was killed in the ensuing battle of Lützen (*q.v.*), his chancellor, Count Axel Oxenstierna (*q.v.*), continued his policy and linked Sweden with France and Saxony in the league of Heilbronn. The defeat of the Swedes in 1634 obliged Richelieu to commit France to open war against Spain and the empire, and the phase of the war thus came to reflect the underlying contest between France and the Habsburgs.

The cautious policies the rulers of Saxony and Brandenburg had followed in the Thirty Years' War had preserved the independence of the princes, but it had done so at a price which relegated Germany to a subordinate position in Europe for several generations. The ambitions of Maximilian I had finally been frustrated, and the Austrian Habsburgs began to look for compensation to Bohemia and Hungary. Constitutionally and materially the empire was reduced to a shadow. Towns such as Magdeburg, Minden and Göttingen stood in ruins, and the total population had shrunk from about 21,000,000 to about 14,000,000.

By the peace of Westphalia (*q.v.*) in 1648 Calvinism was accepted under the terms by which Lutheranism had obtained recognition in the peace of Augsburg. Germany remained a mosaic of petty states, each with its established religion in conformity with that of its ruler. The independence of the Swiss cantons and the United Provinces was confirmed, and the territory of Brandenburg-Prussia increased in north Germany. France expanded in Alsace, and Sweden was established in Pomerania. The decline of Spain and the emergence of French predominance in Europe were expressed in the peace of the Pyrenees, concluded between the two powers in 1659. Swedish power in the Baltic was again asserted in the treaty of Oliva in 1660. (J. H. McM. S.)

E. The Great Age of Monarchy, 1648–1789

It was symbolic of a new period in European history that no ruler, whether Catholic or Protestant, paid heed to the pope's denunciation of the peace of Westphalia. After centuries of groping, the modern sovereign state had emerged as the continent's most powerful institution. There were still to be political upheavals after 1648, and even the strongest rulers had to tread warily lest they provoke untameable troubles, especially from the nobility. Likewise, although there had been a sweeping invasion of medieval corporate privileges and local immunities, there remained in all countries numerous barriers to a completed authority. None the less, the state had arrived at the point where it allowed little room for rivals and even less for rebels.

1. Control of the State.—The growth of the state had been inextricably bound up with the problem of who controlled it. Deeply rooted in European history was the principle of the contractual nature of the royal power. The need to secure consent to important actions and, concurrently, the need to get help with administrative work resulted in a blanketing of Europe with assemblies at all levels of political life (*see* Estates-General). But the challenge to royal ambitions represented by these assemblies had caused princes to set out to reduce or destroy them and by the mid-17th century only England and the Netherlands showed effective resistance to the symptoms of decay.

Dutch burghers and English gentlemen could not set the pattern for the 17th century. One of the first victims of the Thirty Years' War had been the estates assembly of Bohemia, which had supposed that it had the power to depose and elect kings without consulting the Habsburgs. Soon afterward in Denmark-Norway the king brushed aside his quarreling estates and set up rigorous absolutism. At the same time the princes of Germany took up the offensive, and by the end of the century only five had failed to shake their assemblies. Of all the victorious battles those of Frederick William of Hohenzollern, the Great Elector of Brandenburg-Prussia (1640–88), were the most significant. When he took over the widely scattered family possessions there were assemblies everywhere to block his plans for an army. Against each in turn he used every available device, including soldiers, to win a free hand with taxation.

In the 18th century the estate assemblies continued their downward course. In the three continental lands where they maintained their power they proved to be disastrously incompetent. In the Netherlands, declining economically, the parochial-minded assemblies fought the leadership of the house of Orange, and the country was headed for a revolution of its own when it was overwhelmed by that of France. Sweden, in turn, had been undoubtedly the most advanced state of Europe, but the death of Gustavus Adolphus (1632) opened the door to domestic as well as international troubles. Social quarrels led in 1680 to a virtual absolutism, but the events of Charles XII's reign returned dominant power in 1719 to the *Riksdag*. Continuing social tensions, however, were reflected in quarrels within parliament which opened the door for French and Russian competitive corruption, and the country was saved from disintegration only by a royal *coup d'état* in 1772. The estates of Poland, meanwhile, had brought themselves to the point where a last-minute rescue was impossible. By the 18th century the nobles' zeal for golden liberty had reached the absurdity of the *liberum veto;* one dissident voice could not only block a legislative act but also "explode" the assembly. Surrounded by three hungry and militarily strong neighbours, the Polish nobles by their corruption and their divisions invited the disaster of the partitions from 1772 to 1795.

The decline of assemblies representative of estates was counterbalanced by the rise of absolute monarchy. This shift had not come about simply by mistake and default of the estates; a vital part of the story was a millennium of dynamic energy which Europe's crowned heads had thrown into the struggle for power. Louis XIV was the brilliant embodiment of this royal victory. Not very intelligent, he had, nonetheless, the talent to make his people confuse God's rule and that of the king of France. With insatiable thirst for glory, he made his whole life a ritual of Byzantine majesty. He was a master builder whose works reflected his grandeur; he patronized the arts; he tamed disorders in his own land and he sent his armies abroad. The lesser sovereigns found in him their model. They learned French; they copied, as far as their incomes allowed, the buildings and the manners of Versailles; and, above all, they laboured to imitate Louis' climactic achievement, his despotism.

2. Machinery of the State.—Being an absolute monarch required more than postures and parades. Success largely depended on the nature and the quality of governmental machinery. The tendency of 17th-century Europe was to assign the major tasks of government to councils or commissions. In Sweden a reorganization in 1632 had set up five ministerial colleges that were so effective that they became the model for eastern Europe and notably for Peter the Great in his efforts to reform Russia. In Spain a multiplicity of councils had stemmed from the formative days of the monarchy, and in France Richelieu adopted the Spanish style as a means of allowing his subordinates to restrain each other. Louis XIV, heir to this calculated confusion, did much of his royal business in five separate bodies. Likewise in England, both before and after the Restoration of 1660, boards and commissions were the favoured agencies. Such machinery, evolving in response to the growth of the state's business, was cumbersome, often overlapping and prone to paralysis. As a means for coping with the resultant difficulties, the late 17th century and the 18th turned increasingly to a single minister or secretary of state. Reform not infrequently turned out to be merely an exchange of one confusion for another.

The state's supreme domestic problem was to provide a coherent direction for this inefficient machinery. The English had found an effective answer: policy formulation came from a group, or cabinet, composed of men who individually managed the great departments. The rising importance of political parties tended to bring like-minded men to office, and when George I (1714–27) absented himself from cabinet meetings active leadership fell to a prime minister. (*See* CABINET: *Historical Development.*)

In France Richelieu and Mazarin had served as over-all directing ministers, but Louis XIV found such an office incompatible with his majesty and after the death of Mazarin no one took his place. This jealousy of powerful ministers meant that the king alone not only decided policy but also carried the burdens of a chief co-ordinating administrator. In this respect also, Louis XIV set the precedent for the 18th century. Occasionally there was a chief minister in some capital, but the despotic temper of royalty disliked the expedient. For the same reason royal councils as well as guiding ministers went into decline. Various attempts continued to be made in the 18th century to co-ordinate central functions, but royal impatience blocked the way to success. Accordingly, as continental governments faced ever-mounting obligations the sovereign had no escape from the role of first servant of the state. In England it came to be of relatively little importance who wore the crown, but on the continent the prosperity of a country depended almost entirely on the personal qualities of the monarch.

3. Provincial Government.—Out of special circumstances federalism became established in the two republics that gained their formal independence in 1648, Switzerland and the Netherlands. Elsewhere in Europe the drive had been toward the unitary state, with immediate control from the centre. The difficulties had been enormous, because the great states of Europe were synthetic creations pulled together out of diverse territories whose institutions the king was generally bound to respect.

The English arrived least painfully at a solution. The crucial achievement of medieval English kings had been the supremacy of the law. This enabled the central government to leave local administration to the justices of the peace and the town councils. On the continent local legal variations had been too strongly embedded, and some other means was necessary for achieving unity. That means was the dispatch from the capital of a corps of loyal agents. Sweden could do this as early as the 16th century because of the small influence of feudalism in that country. Ferdinand and Isabella began in Spain by establishing their own *corregidores* in the municipalities, but traditional divisions of the joint monarchy made for slow progress. In France the procedure developed over centuries. The first intendants to go out from Paris were merely reporters; Henry IV and Richelieu used them as temporary agents for dealing with emergencies, and it was not until Louis XIV's reign that they became institutionalized as permanent officials superimposed upon, and thereby handicapped by, the old provincial governments.

While Louis XIV was installing his intendants the Great Elector of Brandenburg-Prussia began to send out war commissioners, at first to receive tax payments but soon to settle down as local officials. A second bureaucratic machine had been set up to administer the elector's patrimonial lands, and in 1723 both were combined under one central *Directorium*. The directory itself was cumbersome and inefficient but its civil service, the best in Europe, took the royal will throughout the kingdom.

In the Habsburg dominions peculiarly formidable barriers stood in the way of a centrally controlled provincial administration because of the unassimilated diversity of titles and crowns. When Maria Theresa (1740–80) undertook the creation of a comprehensive central government she also began to move against the confusion of provincial immunities. Fairly successfully she introduced district captains as immediate agents of Vienna into the western part of her inheritance, but she hesitated to send them into Hungary where high sheriffs drawn from the local nobility administered the counties with more zeal for Magyar liberties than for Viennese autocracy. Her son Joseph II had no patience with such softness and sent in his German-speaking bureaucrats—to the lasting misfortune of central Europe. In these Habsburg lands above all others some accommodation of local differences was essential, but the spirit of 18th-century despotism demanded a rational uniformity.

4. Government Officials.—In the early labours of state-making and administration the medieval kings had had to depend for aid on the nobility and the clergy, neither of which was single-mindedly devoted to the royal interests. With the development of urban life there had appeared as competitors of the higher orders for governmental posts members of the *bourgeoisie* who could offer unqualified loyalty and technical skills in law and finance. Louis XIV, who never forgot the Fronde (*q.v.*), the nobles'

revolt of his childhood, and who was jealous of any light that was not a reflection of his own, went to the extreme of closing the doors of his councils to the French grandees and transacted the state's business with "vile bourgeois." In turn the intendants carried to their offices in the provinces the same nonnoble stamp. But even Louis could not deny the spirit of the times altogether and, following precedent, he sold to his bourgeois officials their positions and titles of nobility. This created a second aristocracy, known as the nobility of the robe because of the gown of office. The old nobility of the sword greeted the robe with a disdain that was fully reciprocated. Before the middle of the 18th century the two nobilities were uniting to oppose the remnants of Louis XIV's bourgeois policy. Increasingly arrogant about the importance of ancestors, they harried the king into giving them a virtual monopoly of the important state posts.

Louis XIV's antinoble prejudice had run aground on the stronger prejudice that nobles were the only fit companions of a king. The simplest means of meeting it was to elevate commoners as they rose in the hierarchy, and this practice became universal. In a number of countries the scale of title-making was such as to constitute, as in France, a new aristocracy. Ivan the Terrible had raised up a second Russian nobility committed to state obligations, but the two groups amalgamated and Peter the Great shortly after 1700 had the task to do over again. He attempted it in substantially the same spirit; the public servant received a status in the hierarchy of nobility in accord with his grade in the state service.

High state office, therefore, was an almost exclusive prerogative of the nobles old and new, except in England. The nobles' attitude toward their right varied from country to country. For the English gentleman public activity was a matter of personal sense of duty. The Prussian nobles, happy in their new alliance, served their king with military devotion. The Russians, resentful of service, wheedled the foolish Peter III into releasing them. The Poles constituted a special anarchistic group, and the Hungarians were potential rebels against Habsburg encroachments. The French nobility was moving into a dissident role before the death of Louis XV (1774). Through the *parlement* of Paris, the chief law court, it was beginning to challenge the king's despotism.

5. Internal Functions of the State.—The growth of governmental machinery reflected the ever-broadening functions of the state. The intensified search for money led to the invasion of provincial and local government and to the endless work of tax-collecting. The maintenance of a properly majestic establishment for royalty was a state duty; though the scale varied from Versailles to a duodecimo copy in a petty German principality, the style of life led to mounting extravagance in almost every land. Yet such expenditures were not to be compared to the outlays for military purposes. One of the most portentous innovations was the nationalization of the European army. By the end of the Thirty Years' War the change from older forms of organization was well advanced and the state was becoming the great consumer of the goods as well as the services and the blood of its subjects.

The oldest state functions had stemmed from the concept of the king as the fountain of justice. Concurrently with the building of a royal court system, the bitter issue of the competence of the king's courts as against traditional noble and church jurisdictions had been fought. By 1648 the major contest had been settled, although important fragments of the old order remained. Over much of Europe the nobles maintained manorial courts in which they administered justice to their peasants and serfs. The church had lost its vast legal competence but continued to bring into its own tribunals religious cases, defined more broadly than a later time would allow.

In the whole contest between church and state the tide continued to favour the latter (*see* CHURCH AND STATE). In Lutheran countries the state churches of the Reformation lived on without political crisis. By an implied contract the princes looked after the administration and the ministers preached submission to God and to his divinely ordained regents on earth. In England the established church was yet more clearly the creature of the state in that even its creed was a product of parliamentary legislation. Temporarily broken during the Commonwealth, restored under Charles II and then seriously menaced by the Roman Catholic James II, Anglicanism was saved by the Glorious Revolution of 1688, and it gratefully put itself at the service of the new order. During their ascendancy in the 18th century the Whigs took over the church as well as the state, and the bishops in the house of lords dutifully supported the government, discouraging any expression of "enthusiasm" that might disturb the religious or political calm.

In Russia the outcome was even more devastating for the authority and the moral position of the church. With a patriarchate of its own after 1589, Russian Christianity had tried, in the face of state encroachments, to argue for dual, independent spiritual and secular realms. An old tension came to the breaking point when the patriarch Nikon, elevated in 1652, introduced modifications in ritual for political ends (*see* NIKON). The government had no quarrel with the reforms but with Nikon's arrogance, and the tsar dismissed him. The church was split, but neither branch was prepared to accept the westernizing infamies of Peter the Great. He responded to opposition by abolishing the patriarchate altogether in 1721 and by putting the church under a Holy synod. Soon a lay procurator became the president of the synod and the controversy was resolved for the life of the tsarist regime: the Russian church was little more than the spiritual arm of the secular order.

In Roman Catholic territories the state, of course, had to reckon with a powerful international organization. The papacy, shaken by the Reformation to its foundations, had been able to restore and even strengthen its position. A proud pope and a Louis XIV believing in divine right were bound to collide. Louis's extension to the whole kingdom of his right to control the affairs of vacant bishoprics precipitated a new chapter of the conflict. In 1682 an assembly of the French clergy, prodded by Louis, stated in four points the so-called ancient liberties of the Gallican church and reasserted the principle of the subordination of the papacy to a general council. The pope again protested vigorously, and Louis threatened to set up an autonomous national church. For ten years relations between Rome and Versailles were suspended. At last, however, Louis's embroilment in war led him to the motions of capitulating to a new pope. The controversy involved no suspicion of heresy; at its height Louis repealed Henry IV's edict of toleration (1685; *see* NANTES, EDICT OF).

Louis's zealous orthodoxy confused and then transformed the whole church-state issue in France. He was so hostile to the doctrines of Jansenism (*q.v.*) that he appealed to the pope for support against a vigorous segment of his own subjects. Papal bulls of condemnation inspired an alliance of Jansenists and nationalist Gallicans, and their resistance to combined royal and papal pressures stirred a discord that lasted until the eve of the Revolution and that did great damage to the spiritual force of official Catholicism. Anticlericalism and deism were the end products.

While France's religious problem was taking this special turn, in the other Catholic countries the debate remained centred on the old issue of the papacy and its powers. A stream of books attacked the allegedly swollen pretensions of Rome. The most noted of them came in 1763 from a Catholic bishop, J. N. von Hontheim, writing under the name of Justinus Febronius. Febronius would allow the pope only the first honour among equals; the church's council stood above him and the time had come to restore the bishops to their usurped position. No less vigorously did he champion the rights of secular princes in the external affairs of the church, and he invited them to suspend obedience to Rome if the pope proved adamant. The clerical opposition within the church and a whole phalanx of Catholic rulers found in Febronius practically every known justification for their ambitions and their encroachments. In 18th-century Naples and Sardinia, in Portugal and Spain the kings extended their powers of church nominations and other functions. During the peace negotiations in 1713 at Utrecht the state delegations excluded the papal nuncio altogether and set a precedent for the rest of the century by ignoring Rome's claims to overlordship over various Italian territories.

Nowhere was the papacy more disappointed than in the Habsburg dominions, where, since the Reformation, the ruling family had vigorously supported Catholicism. Maria Theresa was personally dedicated to the church, but as sovereign she was sensitive to the slightest hint of papal intervention. Joseph II elevated Febronianism into Josephism. He closed monasteries, sent theological candidates to rationalist state seminaries, and even decided matters of church ritual. Joseph, like Louis XIV, lost the support of his people, and his brother Leopold, succeeding him in 1790, had to salvage what he could by abandoning Joseph's excesses.

Meanwhile the Society of Jesus had come under disastrous fire. Some of the Jesuit doctrines and tactics had inspired criticisms within the church and even rebukes from the papacy. Certain Jesuit political teachings and activities had created ill will among secular authorities. Catastrophe began in Portugal, a country the Jesuits had dominated for a century, and at the hands of one of their students, the reforming Marqués de Pombal. Making use of difficulties in Paraguay and alleged complicity in an attack on the king, Pombal expelled the Jesuits from Portugal and its colonies in 1759–61 and for a decade suspended all ties with Rome. In France, Jansenism, Gallicanism and anticlerical rationalism united in a long-sustained attack brought to its climax by the bankruptcy of a Jesuit commercial firm. The Paris law court found the society subversive and ordered its expropriation and expulsion (1764). Three years later the Bourbon king of Spain drove out nearly 6,000 Jesuits and shortly Naples and Parma followed suit. The Bourbon sovereigns then jointly demanded the suppression of the society, and in 1773 Pope Clement XIV, a Franciscan little disposed toward the Jesuits, acquiesced. (*See* also JESUS, SOCIETY OF: *Suppression and Restoration.*)

In addition to these concerns with the administration of the church, the state also had the traditional function of defense of correct Christian belief. The Reformation had given the rulers of Europe an opportunity to decide, each for himself, what was correct, and the starting assumption was that subject would bow to the decision of sovereign. It had been increasingly difficult, however, to impose a single form of faith. In the Low Countries a formal edict of toleration in 1614 arose out of very practical economic and political considerations. In England shifts of official religion from Henry VIII's break with Rome to the Glorious Revolution had strewn the land with a diversity of churches, and the Act of Toleration in 1689 simply regularized an inescapable reality. In Germany the peace of Westphalia had allowed the princes of the Holy Roman empire to impose on their subjects their own choice among three religions. In the regions devastated by the Thirty Years' War there was acute need for population increase, and the Great Elector of Brandenburg-Prussia was more interested in an immigrant's ability to work and pay taxes than in the minutiae of his Christian convictions. In the next century the personal indifference of Frederick II gave new force to old policy; as long as the churches supported the monarchy he was content to allow each individual subject to find his own way to heaven.

The Roman Catholic states lagged behind this Protestant march toward toleration. Louis XIV's economically injurious resolution in 1685 to destroy Protestantism was unrevoked for 102 years, until weariness with finely spun religious quarrels, the undermining thrusts of rationalism, and the commanding power of secularization forced Louis XVI to restore a limited tolerance. In Austria, Maria Theresa saw in tolerance only sinful indifference, but Joseph II, who had no patience with his mother's anxieties, issued a Patent of Tolerance in 1781 that gave to non-Catholics a measure of religious freedom and an improved civil status. This culminating caprice of his religious policy brought the pope himself on a protest mission to Vienna, but to no avail.

6. The State and Economic Enterprise.—In its economic aspect the rise of the modern state had meant an attack on the multiplicity of economic controls that were characteristic of feudalism and a persistent drive to concentrate regulatory powers in agencies of the central government. By 1648 the major outlines of the so-called mercantile system (*q.v.*) had been drawn, and the first task of the new epoch was to develop state intervention still further to enhance the state's power.

Money was fundamental to power, and so interventions for revenue purposes—tolls, tariffs, sale of privileges—were numerous. Even more important and diverse kinds of regulations stemmed from the belief that the protection and development of resources required directing controls. The inherited purpose in agriculture was the prevention of loss of food supplies, but the reforming despots of the 18th century promoted increased production through improved methods and through new crops such as the potato. A flourishing domestic industry, however, was much desired. Industry provided employment and it was the prime counter in international trade; through sufficient export of finished goods could be created a "favourable balance" of trade that would bring in gold and silver, metals so highly esteemed that many enthusiasts considered them the only forms of true wealth. To promote industry, accordingly, the state set up protective tariffs, bounties, monopolies and other benefits. In return the manufacturer was accountable to the state for his methods and the quality of his product. Notably in France under Louis XIV and his minister J. B. Colbert (*q.v.*), but actually wherever there were manufactures, regulating edicts and supervising agencies multiplied.

The principle of state intervention in economic enterprise had governed colonial policies from the outset. A system of exclusive trade connections appeared to have the advantages of avoiding payment of precious metals abroad for necessities, of reserving colonial products and of providing an assured export market for finished goods.

A second consideration leading to numerous kinds of state intervention was a sense of the common good. The secular spirit had deeply eroded the sense of social responsibility that had existed in the middle ages, but the modern state could not completely ignore the claims of the unfortunate for protection. Nor could the state fail to perceive that political strength and the welfare of the people were inevitably linked objectives. As governmental machinery and experience increased and as purposes and means became better understood, intervention directed to public improvement continued to expand. In the time of Frederick II and his contemporaries the European state had become an engine of social betterment.

A third impetus to state intervention came from pressure groups seeking advantages, and in all countries the number of laws embodying their successes was imposing. At times their demands were relatively modest as, for example, when the buttonmaker's guild of Paris, injured by a change of styles, laid its griefs at the feet of Louis XIV. At other times the consequences were of the greatest moment, as illustrated by the English East India company's promotion of the Navigation act of 1651. Of all the interest groups none was more persistent than that of the wool manufacturers in England and France. For 400 years they fought winning battles to control state policy and they were defeated only in the second half of the 18th century by the attacks of the cotton manufacturers.

These manifold economic functions of the state were still expanding when dissatisfactions appeared, foreshadowing a change of opinion. After 1715 the strength and self-confidence of capitalism had so advanced in the west that men of business were becoming more interested in liberty than in advantages that derived from state control. To the same end certain rationalists introduced the doctrine of laissez-faire (*q.v.*)—let nature rather than the state do the regulating—and Adam Smith in his *Wealth of Nations* (1776) sustained the argument.

Discontent and economic theory naturally ran ahead of libertarian achievements. Restraints on internal trade were least defensible. Most of them were relics of the middle ages and impediments to royal government, and consequently the great reformers knocked down these domestic barriers from Russia to France and Spain. Likewise the period after 1750 saw a declining interest in the industrial controls that had seemed so important to Colbert and his contemporaries. Great Britain and France were in the vanguard of abandoning any systematic attempt at enforcement, and both gave up restrictions on the introduction of new machinery, a change that opened the door to the Industrial Revolution (*q.v.*) in England. In the same spirit monopolies came under attack, and the most anachronistic of them, those of the guilds, had almost

disappeared by the end of the century.

This new zeal for freedom had little influence on the old assumptions about closed colonial empires. There were some deviations in fact when the British invaded Portuguese and Spanish America and gave some trading concessions to the North American colonies, but it required a political rebellion in the new world to loosen effectively the imperial bond. Nor was control over international commerce easily shaken. On the continent down to the French Revolution kings and manufacturers were persuaded that industrial expansion required high barriers against foreign competitors. To increase the difficulty there persisted the old concept of trade policy as being fundamentally an extension of war policy. If Colbert for economic purposes had made international exchange difficult, Frederick II and Joseph II and Catherine of Russia for political reasons made it virtually impossible. Likewise Franco-British frictions for decades closed the English channel to all but smugglers. But reversal of this policy began in 1776 when Sweden started a move toward tariff reductions. A decade later a commercial treaty between France and Britain greatly benefited trade.

7. The State in Its External Relations.—The peace of 1648 revealed in unclouded sharpness the sovereign independence of the European state. The old claimants to universal authority, the empire and the papacy, had withered into impotence and by recognized right the state defined its own interests and decided how best to protect them.

For the conduct of peaceful international relations, mechanisms and procedures had been developing since Renaissance days in Italy. The basic institution was the foreign ministry, which emerged out of the administrative uncertainties of the major continental capitals around 1700. England lagged behind; two secretaries of state divided external affairs until 1782. But the evolution of a central agency was no guarantee of system and order. Foreign affairs always had been a jealously guarded prerogative of the crown, and policy decisions normally remained subject to those calculations and accidents that moved a king. The second instrument of diplomacy was the ambassador, by custom a man of high birth. Long experience had defined the ambassador's duties and conferred on him privileges and immunities. The charge to report significant developments gave him the aura—and the functions—of a spy, and the obligation to protect his sovereign's honour made him truculent about the honours due to himself. About 1500 the papacy had devised a strict order of precedence for all emissaries in Rome, but toward 1648 this scheme of relative merit was no longer acceptable to Europe's monarchs, filled as each was with a sense of personal majesty. Fierceness about questions of rank made it necessary to hold two peace conferences in Westphalia, and at each the delegates bickered over places in ceremonial processions and even over how far a host should accompany a retiring guest toward his carriage. Forethought and practical expedients eased some of the tensions in the 18th century but only later times fully solved the problems in terms of the equality which proud kings had so long demanded.

8. Military Establishment.—The organization of a satisfactory military establishment proved a difficult task. The feudal array still existed, although largely on paper, in several parts of Europe until about 1700, but for centuries it had been of little value. The state groped toward a better solution when it began to create an army of its own. There were a few standing companies in France by 1500 and more in 16th-century Spain, but the senior regiments of Prussia and Austria dated from the Thirty Years' War. A fully state-owned and operated army came only slowly in the 18th century.

For the officers' corps tradition designated nobles, but sons of the *bourgeoisie* had moved in when the aristocrats showed incompetence in the skills of artillery and engineering and when, in some armies, commissions were sold as private property. In England this intrusion was merely disliked, but on the continent resentments grew as the lines of social exclusiveness hardened. In France the nobles in 1781 at last wheedled out of Louis XVI an edict limiting commissions to men able to prove 16 quarterings of nobility. Meanwhile they had gained another victory: the multiplication of posts until the corps became a vast program of relief and the higher commands had to be rotated frequently to give a semblance of duty. In Prussia the nobles also enjoyed a strong preference, but the results differed markedly from those in France. Instead of granting demoralizing sinecures the Prussian kings demanded a total dedication to duty and paid for it by recognizing in their brothers-in-arms the foremost order of the realm. In the other European armies the prevailing aristocratic stamp suffered little damage.

The common soldier came from, and lived in, another world. The nationalized army began under the theory of volunteer long-term enlistments but very early free decision did not produce enough soldiers. Public policy disapproved of taking useful workers and other sources had to be tapped. Enlisting foreigners proved an effective resort; their proportion in the continental armies ran at times from one-quarter to two-thirds. Within the country the preferred solution was a round up of the economically nonproductive. Local authorities collected jailbirds and paupers and misfits while recruiting sergeants took in by tricks, and if necessary by violence, the innocent and the unwary. The inadequacy of such devices gradually forced the state into instituting compulsory service. Sweden was an innovator in a systematically operated draft, but the more momentous step occurred in Prussia around 1730: a regularized conscription based on the principle of universal liability. A generation later Joseph II took the same course. The inclusion of peasants and artisans, however, remained a disagreeable necessity; until the Revolution Europe preferred for its armies the local tatterdemalions and the foreign adventurers. No one had any illusions about the quality of this raw material, but ways had been devised to cope with it. The first was discipline so rigid and brutal that the soldier had to fear his officer more than the enemy. The second was drill and yet more drill until the men became cogs in a precision machine. Even so, in tactical operations little was possible beyond marching onto a field and firing volleys; any break in the ranks usually resulted in wholesale desertions.

Strategy was limited. An army was far too expensive to be thrown recklessly into battle or taken far from its base of supplies, since foraging meant decamping. Supplies, in turn, meant transport and hence a campaign had to close at the onset of autumn. This sense of limited objectives both in battles and campaigns meant that in the 18th century no one thought of a total defeat or of exacting severe penalties from a fallen enemy. War, whatever its dimensions, nonetheless continued after 1648 to be a ready resort of state policy. It had its critics as well as its glorifiers. In the 17th century Henry IV's great minister Sully published a project attributed to the king of a "great design" for the reconstruction of Europe and the creation of an international council to restrict warfare to crusades against the Turks. In the generations between Sully and Immanuel Kant men repeatedly issued their own impassioned versions of the great design for an organized peace; the dream would not die, but the reality of those generations paid it no attention.

9. Balance of Power.—The best device for peace and protection seemed to be the balance of power. The principle, as old as western culture, had reappeared as conscious policy in Renaissance Italy and had moved to transalpine Europe about 1500. Its need had become apparent when the emperor Charles V inherited so many dominions on all sides of France. War resulted, and the conflict went on for so long that Franco-Habsburg hostility seemed unending and inevitable. Finally the peace of Westphalia ended any conceivable menace to France from the German Habsburgs and the treaty of the Pyrenees in 1659 won territorial advantages and a royal marriage from the Spanish. These settlements confirmed a shift in the direction of French external ambitions from Italy to the northeastern frontier.

Louis XIV began his personal rule in 1661. The threat to the balance that then appeared was of Louis's own making, and one coalition after another rose up to resist his aggressions. By the time of the peace of Rijswijk in 1697 Louis's drive had been fairly effectively checked, but the death of Charles II in Spain in 1700 reopened the struggle. Obliged to designate an heir to his many crowns, the childless King Charles in his last days chose a grandson of Louis, his brother-in-law. Louis's greedy response indicated

the greatest upset to the balance of power since Charles V, and the widespread War of the Spanish Succession (*q.v.*) ensued.

This multiplication of battlefields was the work of England. In the 16th century imperial friction and religious animus had made Spain the great enemy; in the 17th century commercial competition had made the Dutch the prime enemy of both Cromwell and the Restoration. As Louis XIV began his depredations he bought the support of his cousin, the English king Charles II, but on the question of France, as on so many others, Charles and his subjects drifted apart. Religious feeling, economic and colonial friction, and the French threat to the Low Countries all indicated to English opinion the emergence of a new enemy. In 1688 the Glorious Revolution placed on the throne Louis's most implacable enemy, William of Orange, and before a year was out the new king took England into his old struggle. Thus began that Anglo-French contest that went on until Waterloo in 1815.

When exhaustion and allied cross-purposes brought the War of the Spanish Succession to an end at Utrecht in 1713, the treaties proclaimed a "just equilibrium of power." Historically conceived as a deterrent to war, the principle of the balance as invoked at Utrecht determined the distribution of territorial compensations at the end of a war. Repeatedly thereafter until the final partition of Poland (1795) numerous crises arose from demands to restore the balance in this manner.

The treaty of Utrecht (*q.v.*) introduced a period of co-operation between England and France, but unresolved imperial rivalries allowed it only a passing term. Deteriorating relations led first to the War of Jenkins' Ear between England and Spain in the West Indies in 1739 (*see* JENKINS, ROBERT) and shortly to world-wide conflict between England and France. Meanwhile Frederick II in 1740 had precipitated a second European war, the War of the Austrian Succession (*q.v.*), by overrunning Silesia at the expense of the empress Maria Theresa. For hotheads in Paris here was a golden opportunity for a new attack on the Habsburgs, and soon French military successes in Germany warned of a dangerous upset to the balance of power in the making. British interest required assistance to the young empress, and the two wars intermingled. In 1748 general weariness forced the peace of Aix-la-Chapelle. The British had tired of the Austrian cause and Frederick kept Silesia but otherwise the peace terms required mutual restoration of conquests.

The settlement settled nothing. The only subsequent development not anticipated was that the allies changed partners (the so-called diplomatic revolution). Frederick's assurances beguiled the British into supporting him, and Frederick's perfidies drove the French to Vienna. Turning their backs on two and a half centuries of family quarrels, Louis XV and Maria Theresa joined forces and the empress' daughter Marie Antoinette went to Versailles as a bride to seal the strange compact.

The Seven Years' War (*q.v.*) of 1756–63 erupted in America, India and Europe and produced both epic deeds and great disappointments. When George III abandoned Prussia, Frederick bitterly haggled out of Vienna a return to the *status quo ante*, but the treaty effected no peace of heart between Hohenzollern and Habsburg. Nor was there reconciliation between London and Versailles even though George III had abandoned substantial British gains. In France preparations for a new trial began immediately, and the occasion came in 1778 when Louis allied with the colonists in the American War of Independence. In the peace of 1783 Britain had the humiliation of losing 13 colonies but suffered no organic blow. France won satisfactions, largely intangible, but they were a poor return for a mountain of unmanageable debt.

10. Eastern Europe and the Eastern Question.—A new period in eastern Europe was also opening toward the middle of the 17th century. Poland, once stretching over the vast plains from the Baltic to the Black sea, was beginning to pay the price of its constitutional follies as one neighbour after another carved for itself slices of the country. In contrast Sweden entered the political scene with its drive to dominate the Baltic. High moment and disaster came in rapid succession after Charles XII (*q.v.*) came to the throne in 1697. He struck at foes from Denmark to Russia, but in his second campaign against Russia he turned southward to join the Cossack Mazepa and his army and suffered annihilation at Poltava in 1709. Charles was one of the few to escape, but his glory had departed. The peace of Nystad, made in 1721 after his death, merely confirmed the fact that Russia had taken the eastern shore of the Baltic, where already the tsar was building his new capital of St. Petersburg. Charles by invading the heart of Russia had opened the way of Russia into Europe.

Russia was also in those years expanding southward. In 1677 the Russians stumbled into their first encounter with the Ottoman Turks, and this scuffle launched a new theme of European history. Peter took up the challenge and by 1697, after bitter reverses, had fought his way down to the Sea of Azov. Already, however, a Russo-Turkish war was caught up in a broader context. Habsburg armies were also on the march against the Ottoman empire, and the consequence was a question laden with implications: how were these enemies of Turkey going to deal with each other?

In 1683 the Turks began their second siege of Vienna, and when King John III Sobieski appeared to relieve the city Ottoman power collapsed. In pursuit the Christian forces pushed down the Danube to liberate most of Hungary. The eternal necessity of Austria to face both east and west interrupted the work—Louis XIV was making trouble—and it was not until 1697 that the emperor Leopold was ready to strike a decisive blow. Meanwhile Peter the Great, desirous of promoting a grand crusade, had sent an emissary to Vienna, and there the two enemies of Turkey signed a three-year alliance. The Austrians inflicted a crushing defeat on the sultan at Zenta, but then, to the consternation of Peter, Leopold insisted on making peace; he had to prepare for the death of the king of Spain. In the treaty of Karlowitz (1699) the emperor recovered Hungary and the tsar had to content himself with Azov. This first essay in co-operation, therefore, turned out badly. In 1710, when the sultan declared war on Russia, Peter proposed a new alliance, but the War of the Spanish Succession was absorbing all of Austria's energies and he had to fight alone. Suddenly caught by the Ottomans on the Pruth river, Peter bought his escape by surrendering Azov; when, a few years later, there came an Austro-Turkish explosion the tsar remained at home. There was again, consequently, a separate war, but this time, after Eugene of Savoy had won more brilliant victories, Austria at the peace of Passarowitz (1718) gained possessions in Bosnia and Serbia including the citadel of Belgrade.

After Peter's death (1725) the ruling clique in Russia rekindled the anti-Turkish fires and, in preparation for fresh conflict, negotiated a defensive treaty with Austria. When in 1735 the Russians were the aggressors Austria was not bound to fight but Vienna discovered a powerful argument for war: there would be great danger in allowing Russia to fight its way down to the Danube and there dictate a private peace; co-operation in the war would ensure co-operation in the peacemaking. Calculations and hopes alike went awry. There were bickerings between the allies over their unco-ordinated efforts, and just as the Russians had won a solid victory in Moldavia panic seized the Austrian army; amid great confusions an emissary signed the treaty of Belgrade (1739), which returned to Turkey almost all of the gains of 1718. Unable to carry on alone, the Russians went home and St. Petersburg made peace with the small comfort of recovering Azov.

For a generation the two former allies had no time for Turks, but in 1768 Catherine II's activities against the Poles set off a new Russo-Turkish conflict. Within a year Russian troops stood on the Danube and, in gravest alarm, the Austrians made an alliance with the sultan, pledging themselves to rescue Poland and to defend the territories of Turkey. Instead of fighting, however, Joseph and Maria Theresa turned around and joined Russia and Prussia in the first partition of Poland (1772). Catherine was then free to settle alone with Turkey, and in the peace of Kuchuk-Kainarji (1774) she won spectacular territorial and commercial gains and, most ominous of all, secured a vaguely defined right to make representations on behalf of the Greek Orthodox Christians of Turkey.

Joseph had walked out on his ally but justified himself on the ground that he had moderated Catherine's demands. It seemed to the emperor better to work with the tsarina than against her.

but over several years of negotiations they could not agree on objectives. Nonetheless when Catherine in 1787 provoked a new war, Austrian forces also took to the field. Neither army fought well, however, and suddenly the new king of Prussia struck at them by allying with Turkey. In the midst of these and other misfortunes the dispirited Joseph died and his brother Leopold, abandoning the Russian accomplice, made peace at Sistova in 1791 on the basis of the *status quo ante*. Catherine in the subsequent peace of Jassy (Iasi) settled for a Black sea fortress and its hinterland. Already the French Revolution and a new partition of Poland were capturing Europe's attention, and the Eastern Question (*q.v.*) passed unresolved to a later day.

11. Basic Change and Crisis.—The outward trend of this century and a half seemed to be toward the crystallization of Europe's politico-social structure: an absolute monarch surrounded by proud nobles who, in the chief offices, served him and restrained him in the interest of their own class. Those sovereigns were, in varying degrees, reformers but their purposes were more political than social—to abolish and to introduce in order to make a more effective state. Initiative, even thought, belonged only to one. However, changes in the structure were occurring. For more than a century the population of Europe had remained around 100,000,000; then it began to climb—to about 140,000,000 in 1750 and 165,000,000 in 1789. In England there was an attempt to solve the problem of food supplies. New crops and methods of escape from the ancient poverty of the fallow land were introduced. The material profit of ending the inefficient heritage of small strips and open fields through enclosure was shown (*see* Commons). Textile machines, the steam engine and the attack on the problem of cheap iron heralded a new epoch in industry.

The continental despots were not averse to such changes. They had long sought larger populations, had tried to promote industry, and in the 18th century had tried English methods of agriculture. Recognizing the hampering consequences of serfdom, some had wished to moderate or even destroy that institution. The difficulty for the rulers lay in the unforeseen social consequences of economic change. The rapid growth of urban forms of wealth in the west tipped the scales heavily against Europe's traditional form of wealth—land, the economic base of the aristocracy. Since the continental nobles did not participate in business, they were dependent on royal patronage. On the other hand, it was equally inevitable that men of commercial and industrial fortune should want better recognition. Before 1789 the English middle class had already posed the problem of wider representation in the commons; the French *bourgeoisie*, without a political tradition on which to build, found its first great grievance in inequality rather than in political abuse, but in France, too, political hopes were dawning.

A wide assent to a set of intellectual principles summarized under the term Enlightenment (*q.v.*) made the great age of monarchy also the age of reason. Building on a long tradition which seemed to have a final vindication in Newtonian science, thinkers set out to re-examine all things with the tool of reason. Even though some could accept absolutism provided it was enlightened, the broad effects of rationalism were damaging to the monarchic order. Reason demanded a right of private judgment; its religious speculations injured the Christian basis of society; its legal thought subjected a royal edict to the test of right reason; its economic analysis shut the door on kingly interference; its political philosophy, born of the doctrines of social contract and natural rights, sketched the outlines of the death warrant of the *ancien régime*. The great synthesis was breaking down before 1789 but not before it had helped to prepare the way for a new kind of Europe. (D. Ha.)

F. The French Revolution and Napoleon, 1789–1815

The French Revolution (*q.v.*) was more than a sequence of events in France between 1789 and 1795. It was one of the greatest upheavals of modern history, complex and far-reaching in importance for the whole of European civilization. It was part of a wider trend of democratic movements affecting many other countries, including North America, during the later 18th century. Modern historians see it as a continuing process, beginning in the 1770s, comprising in 1787–88 a reaction of the privileged classes of nobles, clergy and officeholders against Louis XVI and his ministers, and passing through several phases between 1789 and 1799, with later reverberations throughout the 19th century. Events between 1789 and 1795 are the central, especially dramatic, act in a larger drama.

1. The Revolution in France.—The importance of the meeting of the estates-general in May 1789 was that it brought to one focal point all the political and social conflicts of the *ancien régime* in France. From the only nationally representative body known to France, though it had not met since 1614, the king hoped to find an escape from his chronic insolvency. The ecclesiastical, lay and judicial aristocracy hoped, by means of the traditional method of voting by separate estates, to block the reforms they feared. The third estate, especially the middle class of officials, traders and professional people, welcomed this opportunity to ventilate grievances and demand reforms. From this tangle of conflicting expectations derived a series of constitutional and political crises. Nobody wanted a revolution; at first nobody wanted to overthrow the monarchy and set up a republic. The hope was for peaceful changes to satisfy the various sectional interests. Yet within the next four years the surge of events and circumstances swept away first the privileges of the aristocracy and the church, and then the monarchy; produced civil war and a reign of terror; engaged France in war with its neighbours; and spread revolutionary ideals to the rest of Europe. These unforeseen and unintended developments, the classical "French Revolution," came about because of the interplay of conflicting ideological and social forces.

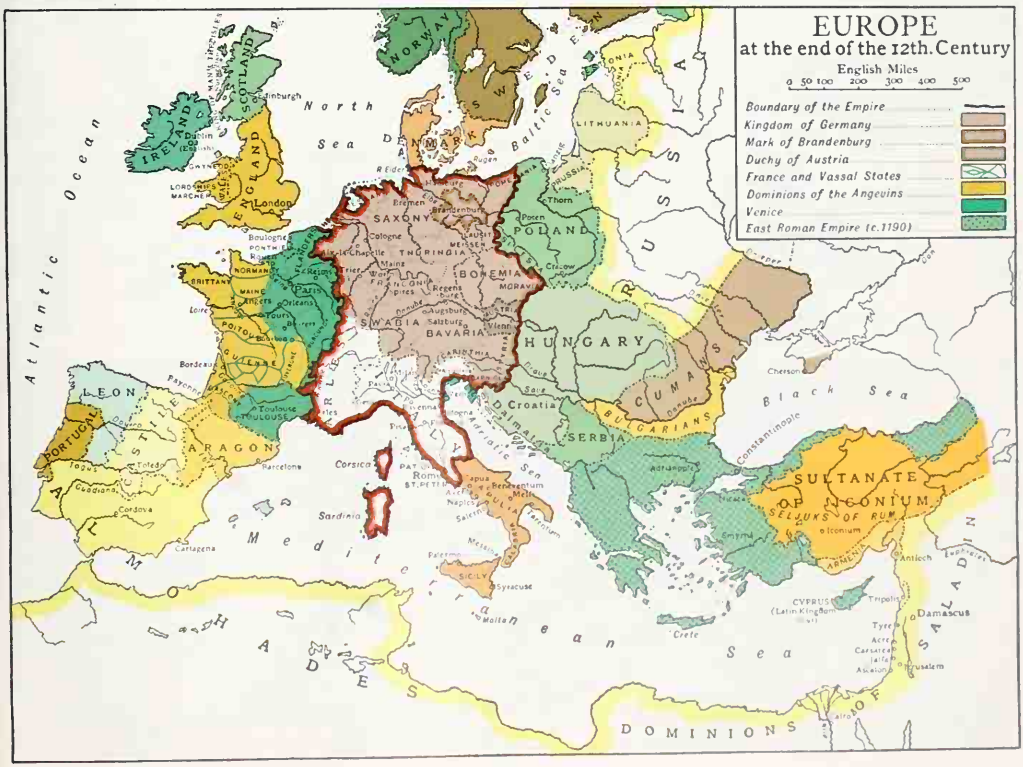
The first group to emerge, the enlightened liberal constitutionalists led by such men as the marquis de Lafayette, Bishop Talleyrand and the abbé Sieyès, triumphed in June 1789 when the third estate constituted itself as the "national assembly," and in July when it undertook to remodel the constitution. The assembly wanted a limited constitutional monarchy on the English model. But it was soon submerged by other more violent elements in French society. The Paris mob scored a symbolic victory with the fall of the Bastille on July 14. The peasants indulged in local frenzies of destruction in the "great fear" of July and August. There were revolts in provincial towns. There began a flight of *émigrés* from France. When representatives of the aristocracy and clergy, on Aug. 4, voluntarily "surrendered" feudal rights and privileges, they were prompted as much by fear as by generous idealism.

The forces of violence drew strength from economic conditions. The years 1787–89 were a time of bad harvests and severe food shortages, of high prices, bad trade and unemployment. There was a slump in the textile industries. The peasants, for two decades and more, had suffered from a long-term inflationary trend that forced down their standard of living. Economic distress, deepened by the upheavals of currency, civil strife and war in the 1790s, was the background to the political turmoil; and it, too, had its roots in the years before 1789.

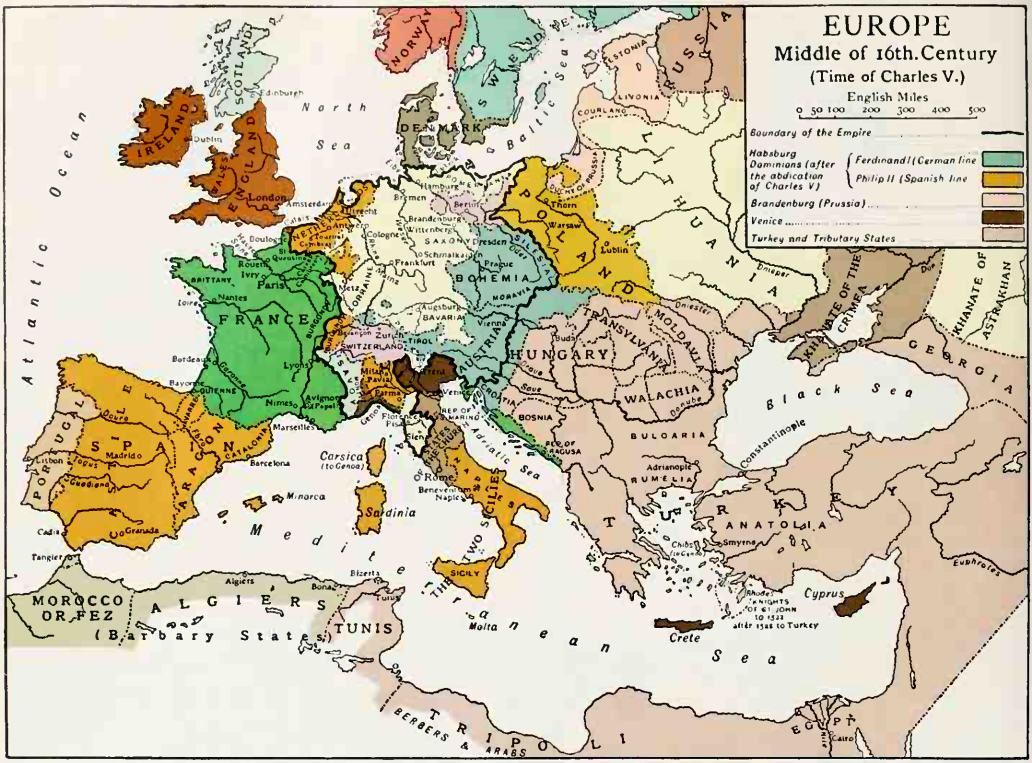
2. The Terror.—There arose within the assembly a more extreme democratic movement, headed by the Jacobins (*q.v.*), demanding the rule of liberty and equality and asserting the principle of "sovereignty of the people." The new constitution proclaimed in Sept. 1791 was, in essence, a parliamentary constitutional monarchy; but by then the royal family, by open support of the *émigrés*, had destroyed the confidence between king and parliament necessary if such a regime was to work. The king was virtually held captive in Paris and in April 1792 war was declared on Austria. The emergencies of war led to the prison massacres of Sept. 1792 and the declaration of a republic. Louis XVI was executed in Jan. 1793. A new assembly, known as the Convention and dominated by the Jacobins, conducted a war of national defense, and in the course of 1793 instituted the Reign of Terror under the dictatorial rule of Maximilien Robespierre (*q.v.*).

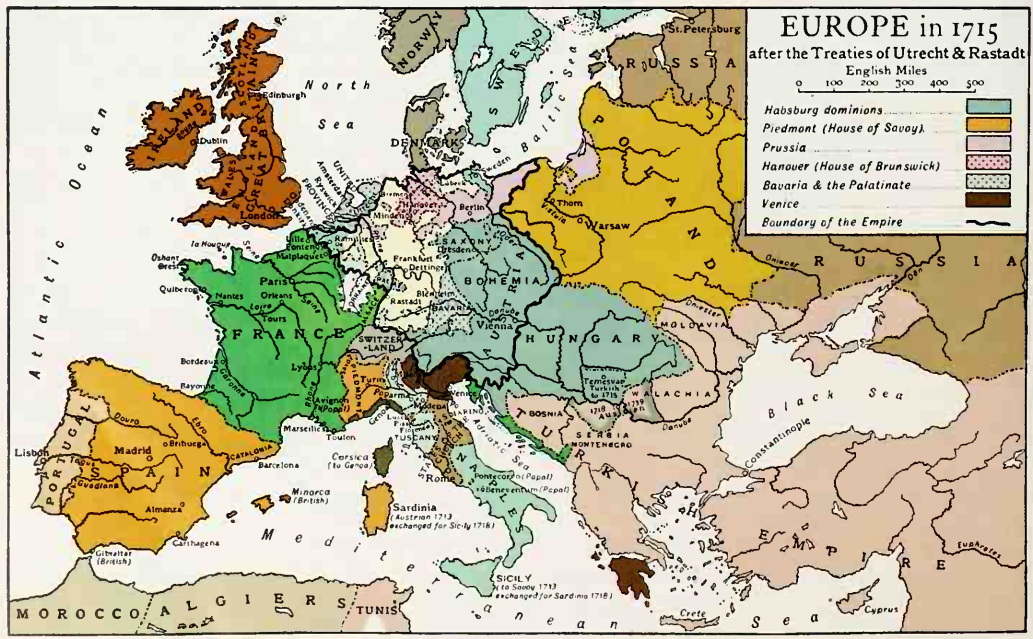
The Terror swelled the tide of *émigrés*, who now included, like the victims of the Terror, people drawn from all classes. Great Britain, the Netherlands, Spain and Sardinia formed a coalition to resist the aggressions of the Revolutionary armies, which had recently occupied Belgium, the left bank of the Rhine and Savoy

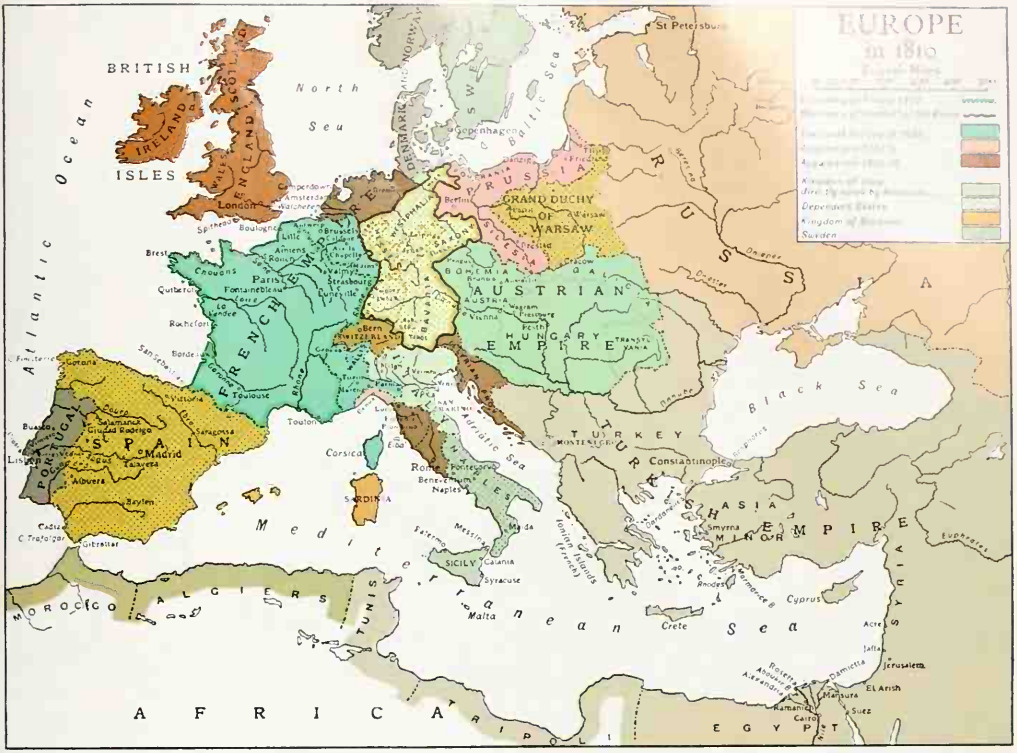
EUROPE
at the end of the 10th. Century
English Miles
0 100 200 300 400 500
Boundary of the Empire
Kingdom of Germany
North Sea
Baltic Sea
Atlantic Ocean
Mediterranean Sea
Black Sea
ENGLAND
IRELAND
SCOTLAND
NORWAY
DENMARK
SWEDEN
LETTS
PRUSSIANS
POMERANIA
RUSSIA
London
Paris
FRANCE
Poitiers
Bordeaux
Toulouse
Marseilles
Vienne
BURGUNDY
Milan
LEON
NAVARRE
CALIPHATE
OF
CORDOVA
Toledo
Cordova
Corsica
Sardinia
Rome
Sicily
BOHEMIA
Moravia
BAVARIA
Salzburg
Vienna
HUNGARY
Dalmatia
Adriatic Sea
Serbia
BULGARIA
Adrianople
Constantinople
Danube
PATZINAKS
EAST ROMAN EMPIRE
DOMINIONS OF THE FATAMITE
Crete
Cyprus

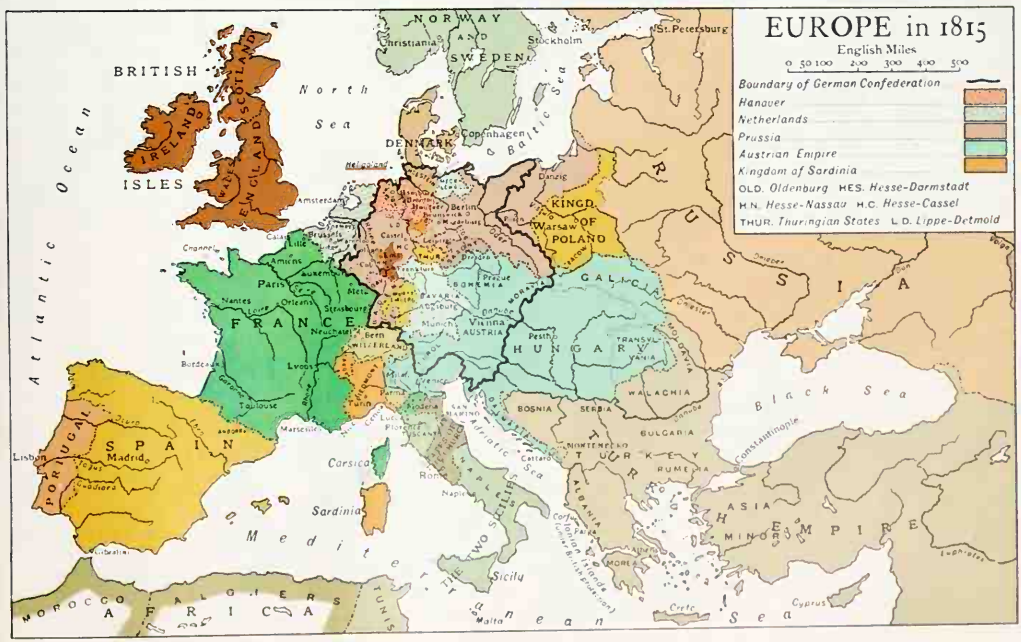

EUROPE
at the end of the 12th. Century
English Miles
0 50 100 200 300 400 500
Boundary of the Empire
Kingdom of Germany
Mark of Brandenburg
Duchy of Austria
France and Vassal States
Dominions of the Angevins
Venice
East Roman Empire (c.1190)
North Sea
Baltic Sea
Atlantic Ocean
Mediterranean Sea
Black Sea
ENGLAND
IRELAND
SCOTLAND
NORWAY
SWEDEN
DENMARK
LITHUANIA
RUSSIA
London
SAXONY
Bremen
Cologne
Poznan
Thorn
POLAND
Prussia
THURINGIA
BOHEMIA
Prague
MORAVIA
Paris
FRANCE
BRITTANY
NORMANDY
Orleans
Tours
Poitiers
Bordeaux
GUIENNE
SWABIA
BAVARIA
Augsburg
Vienna
HUNGARY
Croatia
SERBIA
Adriatic Sea
Rome
Naples
Sicily
Syracuse
Malta
Corsica
Sardinia
LEON
PORTUGAL
ARAGON
Toledo
Cordova
Carthagena
Barcelona
Toulouse
Avignon
Genoa
Milan
Venice
Danube
Constantinople
Adrianople
BULGARIANS
CUMANS
Cherson
SULTANATE
OF
ICONIUM
SELJUKS OF RUM
Iconium
Smyrna
Antioch
Tripoli
Damascus
Tyre
Acre
Jaffa
Ascalon
Jerusalem
CYPRUS
(Latin King)
SALADIN
DOMINIONS OF THE ALMOHADES

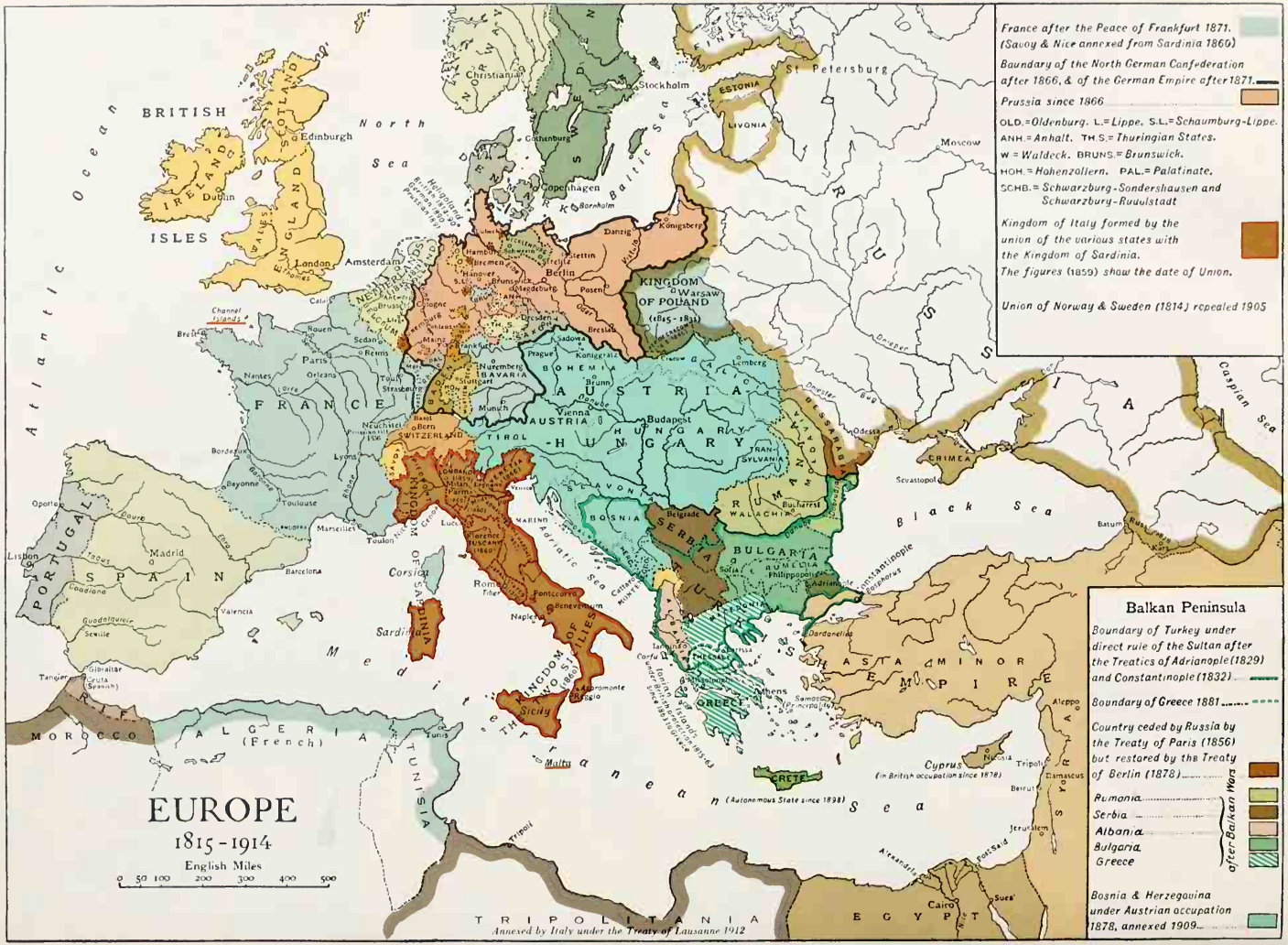
EUROPE


EUROPE
Middle of 16th. Century
(Time of Charles V.)
English Miles
0 50 100 200 300 400 500
Boundary of the Empire
Habsburg Dominions (after the abdication of Charles V)
Ferdinand I (German line)
Philip II (Spanish line)
Brandenburg (Prussia)
Venice
Turkey and Tributary States
North Sea
Baltic Sea
Atlantic Ocean
Black Sea
Mediterranean Sea
Adriatic Sea
IRELAND
ENGLAND
SCOTLAND
DENMARK
NORWAY
SWEDEN
RUSSIA
LIVONIA
ESTONIA
COURLAND
LITHUANIA
POLAND
PRUSSIA
BOHEMIA
SAXONY
BAVARIA
AUSTRIA
SWITZERLAND
FRANCE
BRITTANY
GUIENNE
PROVENCE
CATALONIA
PORTUGAL
SPAIN
HUNGARY
TRANSYLVANIA
MOLDAVIA
WALACHIA
BULGARIA
BOSNIA
RAGUSA
TURKEY
ANATOLIA
RUMELIA
KINGDOM OF THE CRIMEA
KHANATE OF ASTRAKHAN
GEORGIA
MOROCCO OR FEZ
ALGIERS
TUNIS
TRIPOLI
BERBERS & ARABS
EGYPT
Cyprus
Crete
Sardinia
Corsica (to Genoa)
Minorca
Sicily
Rhodes
KNIGHTS OF ST. JOHN TO 1522 after 1566 to Turkey
London
Paris
Madrid
Lisbon
Rome
Naples
Venice
Vienna
Constantinople
Smyrna
Cairo
Algiers
Tunis
Barcelona
Marseilles
Lyons
Bordeaux
Nantes
Amsterdam
Hamburg
Berlin
Dresden
Prague
Frankfurt
Cologne
Augsburg
Zurich
Milan
Parma
Florence
Genoa
Warsaw
Thorn
Lublin
Danzig
Riga
Moscow
Euphrates


EUROPE in 1715
after the Treaties of Utrecht & Rastadt
English Miles
0 100 200 300 400 500
Habsburg dominions
Piedmont (House of Savoy)
Prussia
Hanover (House of Brunswick)
Bavaria & the Palatinate
Venice
Boundary of the Empire
North Sea
Baltic Sea
Atlantic Ocean
Black Sea
Mediterranean Sea
Adriatic Sea
IRELAND
GREAT BRITAIN
NORWAY
SWEDEN
RUSSIA
St. Petersburg
DENMARK
POLAND
PRUSSIA
SAXONY
BOHEMIA
FRANCE
SWITZERLAND
HUNGARY
MOLDAVIA
WALACHIA
BOSNIA
SERVIA
MONTENEGRO
TURKEY
EMPIRE
PORTUGAL
SPAIN
MOROCCO
ALGIERS
TUNIS
Cyprus
Crete
Sardinia (Austrian 1713 exchanged for Sicily 1718)
SICILY (to Savoy 1713 exchanged for Sardinia 1718)
Corsica (to Genoa)
Minorca (British)
Gibraltar (British)
London
Paris
Madrid
Lisbon
Rome
Vienna
Constantinople
Barcelona
Marseilles
Lyons
Bordeaux
Nantes
Amsterdam
Hamburg
Berlin
Dresden
Frankfurt
Cologne
Warsaw
Thorn
Danzig
Riga
Moscow
Euphrates


EUROPE
in 1810
BRITISH ISLES
North Sea
Atlantic Ocean
Baltic Sea
St Petersburg
Copenhagen
Stockholm
Christiania
SWEDEN
NORWAY
DENMARK
PRUSSIA
GRAND DUCHY OF WARSAW
RUSSIA
Dnieper
Dniester
London
Amsterdam
Berlin
Brussels
Boulogne
Brest
Cherbourg
Caen
Paris
Fontainebleau
Orleans
Strasbourg
Rochefort
Bordeaux
Toulouse
Nantes
Lyons
Marseilles
Geneva
Bern
SWITZERLAND
BOHEMIA
AUSTRIA
Vienna
Prague
Munich
Trieste
Venice
Milan
Turin
HUNGARY
Pesth
TRANSYLVANIA
GALICIA
Warsaw
Cracow
Lemberg
Odessa
Black Sea
TURKEY
MONTENEGRO
Constantinople
Adrianople
Smyrna
ASIA MINOR
EMPIRE
Euphrates
Cyprus
Rhodes
Crete
Sea
Corfu
Ionian Islands (French)
Adriatic Sea
Naples
Rome
Benevento
Pontecorvo
Corsica
SARDINIA
Messina
Malta
SICILY
Catania
Syracuse
Mediterranean
Gibraltar
Cadiz
C. Trafalgar
Lisbon
PORTUGAL
SPAIN
Madrid
Talavera
Aldovera
Badajoz
Seville
Granada
Carthagena
Valencia
Saragossa
Barcelona
Tarragona
Victoria
Salamanca
Ciudad Rodrigo
Burgos
Santander
Corunna
MOROCCO
ALGIERS
TUNIS
TRIPOLI
AFRICA
Acre
Jaffa
Jerusalem
Damietta
Alexandria
Rosetta
Aboukir
Ramanieh
Cairo
Mansurah
El Arish
Suez
EGYPT
Kingdom of Italy
Dependent States
Kingdom of Sweden


EUROPE in 1815
English Miles
0 50 100 200 300 400 500
Boundary of German Confederation
Hanover
Netherlands
Prussia
Austrian Empire
Kingdom of Sardinia
OLD. Oldenburg HES. Hesse-Dormstadt
H.N. Hesse-Nassau H.C. Hesse-Cassel
THUR. Thuringian States L.D. Lippe-Detmold
BRITISH ISLES
North Sea
Atlantic Ocean
Baltic Sea
St. Petersburg
Christiania
SWEDEN
Stockholm
NORWAY
DENMARK
Copenhagen
Helgoland
RUSSIA
KINGD. OF Warsaw POLAND
Danzig
Amsterdam
Berlin
Brussels
Cologne
Lille
Amiens
Paris
Nantes
Orleans
Strasbourg
FRANCE
Lyons
Bordeaux
Toulouse
Marseilles
Geneva
Bern
SWITZERLAND
BOHEMIA
Prague
Munich
Vienna
AUSTRIA
Pesth
HUNGARY
TRANSYLVANIA
GALICIA
Lemberg
Cracow
Odessa
Black Sea
WALACHIA
BOSNIA
SERBIA
BULGARIA
TURKEY
RUMELIA
Constantinople
Adrianople
Smyrna
ASIA MINOR
EMPIRE
Euphrates
Cyprus
Crete
Sea
Corfu
Ionian Islands (French)
Adriatic Sea
Naples
Rome
Corsica
Sardinia
Sicily
Malta
Mediterranean
Gibraltar
Cadiz
Lisbon
PORTUGAL
SPAIN
Madrid
Seville
Granada
Carthagena
Valencia
Saragossa
Barcelona
Corunna
MOROCCO
ALGIERS
TUNIS
AFRICA


France after the Peace of Frankfurt 1871. (Savoy & Nice annexed from Sardinia 1860)
Boundary of the North German Confederation after 1866, & of the German Empire after 1871.
Prussia since 1866
OLD. = Oldenburg. L. = Lippe. S.L. = Schaumburg-Lippe.
ANH. = Anhalt. TH.S. = Thuringian States.
W = Waldeck. BRUNS. = Brunswick.
HOH. = Hohenzollern. PAL. = Palatinate.
SCHB. = Schwarzburg-Sondershausen and Schwarzburg-Rudolstadt
Kingdom of Italy formed by the union of the various states with the Kingdom of Sardinia. The figures (1859) show the date of Union.
Union of Norway & Sweden (1814) repealed 1905
Balkan Peninsula
Boundary of Turkey under direct rule of the Sultan after the Treatics of Adrianople (1829) and Constantinople (1832)
Boundary of Greece 1881
Country ceded by Russia by the Treaty of Paris (1856) but restored by the Treaty of Berlin (1878)
Rumania
Serbia
Albania
Bulgaria
Greece
after Balkan Wars
Bosnia & Herzegovina under Austrian occupation 1878, annexed 1909
EUROPE
1815 - 1914
English Miles
0 50 100 200 300 400 500
Annexed by Italy under the Treaty of Lausanne 1912
Atlantic Ocean
North Sea
Baltic Sea
Mediterranean Sea
Black Sea
Caspian Sea
BRITISH ISLES
FRANCE
SPAIN
PORTUGAL
AUSTRIA
HUNGARY
GERMANY
RUSSIA
ITALY
SWITZERLAND
BELGIUM
NETHERLANDS
DENMARK
SWEDEN
NORWAY
FINLAND
GREECE
TURKEY
MOROCCO
ALGERIA (French)
TUNISIA
TRIPOLITANIA
EGYPT
Corsica
Sardinia
Sicily
Malta
Crete
Cyprus (in British occupation since 1878)
Christiania
Stockholm
St Petersburg
Moscow
Copenhagen
Gothenburg
Edinburgh
Dublin
London
Amsterdam
Paris
Berlin
Vienna
Budapest
Prague
Munich
Constantinople
Athens
Belgrade
Bucharest
Sofia
Odessa
Sevastopol
Batum
Cairo
Suez
Jerusalem
Damascus
Beirut
Aleppo
Tripoli
Tunis
Algiers
Madrid
Lisbon
Barcelona
Marseilles
Toulon
Lyons
Strasbourg
Frankfurt
Hamburg
Dresden
Warsaw
KINGDOM OF POLAND (1815-1831)
KINGDOM OF THE TWO SICILIES
KINGDOM OF SARDINIA
BOHEMIA
BAVARIA
TIROL
TRANSYLVANIA
BOSNIA
SERBIA
BULGARIA
EASTERN RUMELIA
WALACHIA
ESTONIA
LIVONIA
CRIMEA
ASIA MINOR
ARMENIA
SYRIA
EGYPT

and had appealed to peoples everywhere to overthrow their rulers. The Jacobins repelled the allies and annexed all territories west of the Rhine and the Alps, claiming these to be France's "natural frontiers." Russia, Prussia and Austria dramatically flouted the idea of nationality by their successive partitions of Poland; by 1797 they obliterated it as a separate state, a process begun in 1775. Now rulers everywhere felt threatened by revolution. The war became a conflict between revolutionary nationalism and dynastic monarchy—though the revolutionary leaders showed little respect for feelings of nationality in the populations they subjected to their own rule. (*See* FRENCH REVOLUTIONARY WARS.)

By the end of 1794 the Terror had spent its force; in July Robespierre himself had been guillotined. Power lapsed more and more to the class of "new rich," who had made fortunes from the upheavals of the time, from war contracts or from speculation in the church lands or in the currency. By Oct. 1795 the Convention was dissolved and the Directory, which now governed, represented the forces of reaction against Jacobinism. The "White Terror" of May–June 1795 crushed the surviving revolutionary elements and inflicted yet more suffering on France. But war went on: having invaded the Netherlands in 1795, the French armies invaded northern Italy and southern Germany in 1796. The hero of the brilliant Italian campaign was a young republican general, Napoleon Bonaparte. The French set up a series of satellite republics, on the French model, in northern Italy, Genoa, Rome, Naples, Holland and Switzerland. Prussia and Spain made peace in 1795, Austria in 1797; Britain alone remained at war with France. Austria re-entered the struggle in March 1799 but was forced to make peace again at Lunéville in 1801. In 1802 Great Britain, too, made peace at Amiens.

3. Napoleon in Power.—Since his military coup of 18 Brumaire (Nov. 1799) Bonaparte had been master of France (*see* NAPOLEON I). The Revolution had run its violent course until only a military dictatorship could ensure order. Because Bonaparte was an administrator of genius as well as a brilliant soldier, the five years of the Consulate (1799–1804) laid the foundations of modern France. By the concordat with the papacy (1801) he ended the long struggle begun by the revolutionaries in 1790 when they made the church in France a department of state. In his new civil codes, his schemes for national education, his overhaul of the administrative and fiscal systems, Napoleon consolidated many of the gains of the Revolution. To the middle classes his dictatorship was a bulwark of order and a guarantee that careers in the public service would remain open to talents. To the peasants it was security against a restoration of feudal privileges and of the land they had now acquired. In 1804 he was crowned emperor of the French.

The peace of Amiens was only a truce. France's neighbours, especially Britain, felt as much menaced by Bonapartist imperialism as by Revolutionary Jacobinism, and could not accept without further challenge the great enlargement of French territory and power in Europe. By July 1805 Britain made alliances with Russia and Austria and, in the battle of Trafalgar (October), established its supremacy at sea. Napoleon, by his victories over Austria (Ulm and Austerlitz) in 1805 and over Prussia at Jena in 1806, asserted French supremacy on land (*see* NAPOLEONIC WARS). The struggle became one between Leviathan and Behemoth. To strike at British trade Napoleon decreed (1806) his continental system (*q.v.*), to which Britain replied with a naval blockade. Checked at sea, Napoleon extended and consolidated his continental empire. By 1810, the zenith of his power, his empire comprised the Low Countries and northern Germany, the Rhineland, northern Italy, and the Illyrian provinces, with the Confederation of the Rhine, Spain, the rest of Italy and the grand duchy of Warsaw as dependent territories. Sweden, Denmark, Norway, Prussia, Russia and Austria were his allies. He controlled the whole of Europe except Britain, Portugal and the Ottoman empire. His invasion of Portugal in 1810 was repelled by the British. His anxiety to keep Russia within the orbit of his power led in 1812 to the fateful Russian campaign. It ended in his disastrous retreat from Moscow with the tattered remains of the *Grande Armée*, recruited from most of Europe.

The Russian campaign shattered Napoleon's prestige as well as his military power. The recovery of Prussia and the resurgence of Austria mobilized against him large military resources. These, backed by British naval strength and subsidies, and coupled with guerrilla resistance in Spain, brought about his eventual defeat and abdication (April 1814). The allies entered Paris, established on the throne of France Louis XVIII, the aging brother of Louis XVI, and then met at Vienna to redraw the map of Europe. In March 1815 Napoleon escaped from Elba and made a dramatic return to France, promising liberal institutions; but he was again defeated at Waterloo (June 1815) and exiled to St. Helena, where he died in 1821.

The peoples and states of Europe that had experienced these events would never be the same again. In 1806 the Holy Roman empire had been destroyed, never to be restored. The multitude of little German states, pounded by Napoleon's conquests, grouped into new formations. Prussia had undergone drastic reorganization and emerged with new prestige and strength. Through the legal codes and administrative systems Napoleon had introduced into conquered territories, new ideas and standards of government had become familiar and popular. Even the revolutionary ideals of popular sovereignty and national rights had been spread by his conquests, and his harsh rule had stirred a new spirit of national solidarity and resentment.

Napoleon's despotism was in many respects paternalist and reminiscent of the prerevolutionary benevolent despots. He patronized the arts and encouraged scientific research. He carried out schemes of public works and rationalized civil administration. In other ways he foreshadowed the tyrants of the 20th century, for he governed through a police state, censorship, manipulated elections and plebiscites. The extent of his conquests was due in part to his genius for war and diplomacy, but in part also to the selfish quarrelsomeness and ambitions of his adversaries. Prussia sought gains at the expense of Austria, Russia at the expense of Turkey and Poland, while Great Britain expanded its commercial empire and its sea power. Only experience of repeated separate defeats forced the powers into coalition against him. But in the end he was defeated by the professional armies of the major powers. It was the old-time dynastic monarchs of Europe who assembled as the victors to reconstruct a shattered Europe. Yet it was a new Europe, seething with new ideas and beset with new needs, that confronted the peacemakers of Vienna in 1815.

G. THE 19TH CENTURY, 1815–1914

1. The Restoration.—A quarter-century of revolutionary ferment and recurrent warfare not only made reconstruction and restoration very difficult: it foredoomed to failure any such labours that did not recognize the extent of the transformation these upheavals had brought about. The generation after Waterloo was one of peace between the powers; to that extent the Vienna settlement succeeded. It was also one of restlessness and revolt within most states and of deep social conflicts; this was the irrepressible aftermath of the events of 1789–1815. These internal disturbances in turn affected relations between the powers and brought about the downfall of the "congress system" by which the peacemakers hoped to preserve the peace. Though there was no war between the major powers until the Crimean War, peace was bought at the price first of general repression and the subordination to Austria of the German confederation, then of periodic crises and realinements of the powers.

The treaty of Paris of Nov. 1815 (*see* PARIS, TREATIES OF) pushed back French power to the frontiers of 1790 and exacted a large indemnity. The treaty of Vienna of June 1815 was agreed to, in effect, by the representatives of the five major powers: the emperor Alexander I of Russia; Metternich, the chancellor of Austria; Karl August von Hardenberg, acting for the king of Prussia; Lord Castlereagh for Great Britain; and Talleyrand, the wily spokesman of France. In the general congress nearly every European state was represented, and its purpose was a general settlement of European affairs (*see* VIENNA, CONGRESS OF).

The guiding principles were to prevent a recurrence of French aggression, to restore territories as far as possible to their legitimist

rulers and to establish a balance of power between the major states. To these ends territories were allocated among the powers regardless of the wishes of their inhabitants. Belgium and Holland were combined into the kingdom of the Netherlands, as a buffer state against France. Norway was transferred from Denmark to Sweden. The Palatinate on the left bank of the Rhine went to Bavaria. The German provinces of Saxony, Westphalia and the Rhineland, as well as the Polish provinces of Poznan (Posen) and Pomorze and the remainder of Swedish Pomerania, went to Prussia. Germany was formed into a confederation of states dominated by Austria, which also took Lombardy and Venetia. Poland again disappeared from the map as an independent state. Russia and Great Britain kept their conquests, Britain thus gaining Malta and Heligoland, the Dutch colonies at the Cape, Ceylon and Mauritius, Russia obtaining Finland, the major part of Poland and Bessarabia.

The spoils having been divided on a nice calculation of mutual gains and a judicious balance of power, the settlement was to be preserved by periodic meetings of the major powers to discuss any unsettling issues that might arise. By the terms of the quadruple alliance, Great Britain, Prussia, Austria and Russia pledged themselves to maintain, by force, the terms of the settlement for a period of 20 years. But the concert of Europe, to be conducted at periodic congresses, soon broke down. Britain differed fundamentally from its allies about their readiness to intervene in the internal affairs of other states (even of France) in order to keep the settlement intact. At the congress of Aix-la-Chapelle in 1818, and again at Troppau in 1820, Britain refused to take part in the policing of Europe. The main task of doing so fell on Austria, which accepted it willingly enough, and Metternich established his famous system of espionage and repression for dealing with any signs of agitation or insurrection.

The greatest weakness of the restoration was that it was wedded to dynastic principles at a time when all the main forces of change were opposed to dynasticism. In France Louis XVIII issued a charter of rights and set up a limited parliamentary regime. Patently owing his restoration to the allies, and arousing more apprehension than loyalty in his subjects, he survived by treading warily and restraining the extremists. Other restored monarchs, less shrewd than he, such as Ferdinand VII of Spain and his uncle Ferdinand I of Naples and Sicily, stored up trouble for themselves or their successors by policies of vindictiveness and blind reaction. Napoleon, by his own upstart dynasticism that led him to put members of his own family on the thrones of conquered countries and to marry, in 1810, a Habsburg archduchess, had done as much as the Revolutionaries to discredit dynastic monarchy and destroy its *mystique*.

The Restoration brought back the church as well as the monarchy, for the alliance between altar and throne was still traditional in Europe. The Roman Catholic Church regained a highly privileged position and where, as in France, it could not recover all its lands and property, it was generously supported by the state and resumed its control over education. It reaped some of the ultimate advantages of having been persecuted, and of the prevalent reaction against the rationalism and free thinking of the 18th century. Faith revived, as did clericalist sympathies; though soon Ultramontane claims met resistance even from the restored monarchs.

2. The Economic Revolution.—The years 1815 and 1816 saw bad harvests and economic distress throughout Great Britain and Europe. The depression spread to industry and business, taxes remained high, thousands of banks and commercial companies failed and there was widespread unemployment. These conditions helped undermine the popularity of the restored order. But they were only symptoms of a much vaster economic change, which, like the political and social changes already described, had been gathering momentum during the generation before 1789. This economic revolution had many aspects—not the least important was a great acceleration in the growth of population that began in the later decades of the 18th century in western Europe and, during the following century, affected the whole of Europe.

In 1750 the population of Europe had been roughly 140,000,000. By 1800 it rose to 180,000,000 and by 1850 to 266,000,000. British and German populations increased by roughly 50% between 1815 and the 1850s. The French increased more slowly, Italian and Spanish more slowly still. Russia's rate exceeded even Britain's; its numbers doubled during the first half of the century, reaching 65,000,000. This rapid shifting of the human basis of society may help explain some of the phenomena of the time: the new romantic spirit in literature and the arts, the increased vigour of economic life, the revival of a restless revolutionary movement of high idealism. For the sustenance of so many millions more Europeans, improved agriculture was needed; and improvements in the techniques of farming, whether tillage or stock raising, occurred in these years. The so-called agricultural revolution—marked in Britain by the rapid enclosure of open fields and improved techniques, in France, Belgium and the Rhineland by the redistribution of the land and the freeing of peasants from the burden of feudal dues, tithes and absentee landlords—made it possible to feed the growing populations.

To these demographic and agrarian developments was added an industrial revolution: the invention and use of power-driven machinery, especially in textile industries and mining, the great expansion of heavy industries (mining and metallurgical) and after 1830 the application of steam power to transport with the coming of the steam locomotive (*see* INDUSTRIAL REVOLUTION). All these changes led by the middle of the century to a growth of larger towns, of factories, mills and ports. Manufacturing involved a great movement of population from countryside to towns, as well as the rise of a large town-bred, urban-dwelling population; and a great movement of goods, whether raw materials, foodstuffs or manufactured products, by roads and canals and eventually by railways and steamships. The concurrent improvements in modes of transport and communication brought far-reaching changes in social life and in mental attitudes, as easier mobility broke down localism.

The great expansion in economic activity of all kinds involved, too, new forms of business and commercial organization, wider use of joint-stock companies for mobilizing large amounts of capital, new methods of accountancy and banking. Such developments affected governments no less than peoples, for new classes of rich and influential bankers, financiers and industrialists demanded co-operation from governments and more voice in the framing and conduct of state policies. Men of business, trade and money replaced nobles, clerics and officials as people whose interests and allegiance governments must safeguard. Old-fashioned dynastic states were usually ill-adapted to the needs of this dynamic new class. At the same time, the problems of administering, keeping public order in and managing a fiscal system for an urban, industrial society taxed the ingenuity of governments slow to adapt their methods to unfamiliar conditions.

Although Great Britain was the first country to experience the full impact of this great economic revolution, followed quickly by Belgium and France, its institutions and traditions were in many respects less fitted to deal with them than those of Belgium, France or Prussia. By mid-century Britain's economic and commercial policy became one of free enterprise and free trade, with the state doing as little as possible. This policy, conceded in response to the demands of the powerful mercantile and manufacturing classes, which had little competition to fear in the shipping and the markets of the world, offered wide scope for the quick and complete exploitation of the great natural advantages the country enjoyed. Its rich resources of coal and iron, of technical skill and business organization, gave Britain a natural lead. So negative a policy was not well suited, however, to protecting the economically weaker elements of society; and the philosophy of laissez-faire was responsible for much of the human suffering and ruthless exploitation of labour that marred the Industrial Revolution in its earlier phases.

In France and Prussia, where traditions and means of firmer governmental control and paternalist legislation were strong, the working classes were spared some of the hardships they endured in Britain. In education, too, Britain lagged behind other European states until late in the century. On the other hand, its representative parliamentary institutions proved flexible and capable of successive reforms; and it was inventive in devising new systems

of local government, police and civil administration. In France, Belgium and Prussia the framework of government remained more centralized and bureaucratic.

The economic revolution helped transform the connections between European peoples and those of other continents. In 1783 Great Britain had lost control of its settlements in America, but it kept Canada and some of the West Indies possessions. In 1807 it had abolished the slave trade in the British empire. By 1815 other countries followed this example and in 1833 Britain abolished slavery as an institution. Spain and Portugal lost the empires they had built in Central and South America. The Latin-American peoples won their independence by 1825, sheltered by the Monroe Doctrine of the United States (1823) and by British naval power. But British rule and trade expanded in India. The Dutch, firmly established in Java, spread their rule over Indonesia. An era of fresh European expansion overseas accompanied the economic revolution of the first half of the century.

For all these reasons Europe by 1815 was entering upon a new phase in its civilization, a phase of rapid economic growth and expansion, in which the achievements of science and technology would be of fundamental importance. Just as advances in medical science and hygiene were primarily responsible for the growth of population, and more scientific farming for the increase in food production, so the applications of science and technology to industry and transport brought about the Industrial Revolution. This progress of science was taking place before the French Revolution and would have continued had that event never happened. The drama of the 19th century was the interplay of forces released not only by the great political upheavals of the revolution, the Napoleonic struggle and the economic revolutions, but also by the steady progress of scientific knowledge and of applied science and technology.

3. Liberalism and Nationalism.—Europe of the Restoration was a collection of old bottles filled with new wine, and trouble was soon fermenting. In Spain, Portugal, Piedmont and Naples in 1820–21 revolts broke out, demanding democratic constitutions, the model for which was the abortive Spanish constitution of 1812. Austrian troops quelled the uprisings in Piedmont and Naples, but those in Spain and Portugal temporarily succeeded, since Great Britain and France refused to join in repressive intervention. In 1821 the Greeks rebelled against their Turkish rulers, raising for the powers the greater peril that Russia would take the chance to attack Turkey in the popular cause of a Christian nationalist movement against infidel misrule. The Spanish and Greek risings were the chief concern of the congress of Verona (*q.v.*) in 1822. In 1823 France invaded Spain and restored King Ferdinand.

Austrian domination was confirmed by the suppression of revolution in Italy and Spain, but Great Britain emerged as a champion of movements for national independence and constitutional government in Greece and Portugal. Both the quadruple alliance and the congress system were disrupted by these events. The Greeks fought on, rousing Philhellenic enthusiasm throughout Europe by their heroism, and when they won independence in 1830 it was because Britain, Russia and France went to war with Turkey and forced it to make concessions in the treaty of Adrianople (Sept. 1829). In 1826–27 Britain had also to intervene in Portugal to preserve the forms of constitutional government.

Disruption of the new system paved the way for wider movements of insurrection in 1830. France raised the flag of revolution in protest against the Ultramontane and aristocratic reaction of Charles X, who had succeeded his more tactful brother in 1824. The liberal opposition could take its stand on the charter of June 1814. In July 1830 republican groups in Paris seized the *hôtel de ville* and raised the tricolour flag of the revolution. By a maneuver of the moderate liberal politicians, journalists and bankers, Charles, who fled to England, was replaced not by a republic but by Louis Philippe, representative of the Orleanist branch of the royal family, a more accommodating, liberal-minded man than his immediate predecessors on the throne.

Parliament gained greater control over government and rested on a wider (though still restricted) franchise. The new regime closely resembled the limited constitutional monarchy of Great Britain after the reform bill of 1832 widened the electorate.

In 1830, too, in addition to the Greeks' attainment of independence, the Belgians won their independence from the Dutch (so destroying one provision of the Vienna settlement). Ripples of revolution, stirred by these events, spread even to lands under the watchful eye and the heavy hand of Metternich. In Brunswick, Hanover, Saxony and other states of the German confederation, local revolts forced concessions from reluctant rulers. In such states as Baden and Bavaria, that already had parliamentary systems, liberal oppositions gained in elections and the press became more openly critical of their governments. In 1832 Prussia and Austria got the diet of the confederation at Frankfurt to pass the "six acts," curbing press and opposition in all German states. By 1835 Metternich's system was again secure in Germany. In Italy, too, it faced revolts in Modena and Parma and in papal territories, but these Austria quickly quelled. A Polish insurrection was similarly crushed by Russia in 1831. Only in free Switzerland did most of the cantons succeed in setting up more liberal constitutions.

By the 1830s domestic conflicts between conservative, dynastic and clericalist forces on one side and liberal, nationalist and republican movements on the other had become a general pattern in European politics. Internationally, this same division was reflected in the realignment of Great Britain, France and Belgium as against Austria, Prussia and Russia. It was a pattern very different from anything envisaged in 1815. Most liberals (seeking parliamentary and more representative constitutions) were also nationalists (seeking the unification and independence of all communities feeling themselves to be distinct nationalities). The ideal of popular sovereignty, born of the French Revolution, implied both internal and external self-government. Government and governed were to be brought into closer reciprocal relations, in the name both of democracy and of national self-determination.

A link between domestic forces of revolution in different lands was formed by the secret societies, which flourished because the police states of the restoration drove all rebels underground. It was a golden age of insurrectionary conspiracies and feverish plots, many of them wild and impracticably idealistic, though the Carbonari in Italy and Spain were at times both active and effective. Metternich's spies were equally active internationally against the secret societies, and if he sometimes exaggerated the danger it was to justify his own penetration of other regimes. In such societies cultured aristocrats, intellectuals and poets mingled strangely with cut-throats and adventurers, many of them spending a large portion of their lives in prison or exile.

This sensational quality of the politics of the time matched the mood of the romantic movement in culture and thought. The origins of romanticism can be traced to Jean Jacques Rousseau or to the novels of Sir Walter Scott, but not until the first half of the new century did it become a recognizable and widespread movement in literature, music, painting and sculpture. It had close affinities with both liberalism and nationalism. The nationalist poets of France, Poland and Greece were romantics, hostile to the restoration as well as to Napoleon because they exulted in individual freedom and personal heroism. They dreamed of a society of free and equal individuals as well as of free and independent peoples. Many, like the French poet Lamartine who became minister of foreign affairs in the second French republic of 1848, assumed leading roles in politics. The prototype of the romantic liberal patriot was Giuseppe Mazzini, a member of the Carbonari and founder of the Young Italy movement for Italian unification under a democratic republic. Such men played a central role in the events of the great year of revolutions—1848.

In 1848, as in 1830, discontent grew from bad economic conditions as well as from liberal and nationalist aspirations. Once again the flames of revolt spread fast from one city to another—for it was mainly a revolt of the towns, where social distress was acute and where organized insurrection could more readily occur. In Jan. 1848 the king of Naples was again faced with open rebellion, and within a month riots took place in nearly all the large Italian cities. In February Paris put up barricades and overthrew Louis Philippe, proclaiming a second republic. Within

the year both types of revolt—national revolt against foreign rule and the dreary repressiveness of Metternich's agents, and democratic revolt against the inadequacies of middle-class government—were repeated elsewhere. Disturbances in Great Britain (where the Chartist movement for radical political reforms had been vainly active for more than a decade) and in Belgium and Switzerland were on the French pattern. The movement failed in Great Britain. It succeeded in France only until 1851, when Louis Napoléon, nephew of the great Napoleon, seized power. Switzerland alone achieved the lasting result of a new federal constitution (1848). The main importance of the revolutions lay in the outburst of national enthusiasm, which spelled the end of Metternich's system in Europe, and in the schisms they caused between liberalism and nationalism.

The Italian risings revealed throughout the whole peninsula enthusiasm for national independence but deep differences as regards means to this end. One party, of conservative liberals, pinned its hopes on Charles Albert, king of Piedmont and Sardinia; another, of Catholic federalists, looked to leadership from the papacy; yet a third, the Mazzinian republicans, fought for a democratic republic. Weakened by these divisions, the revolution was eventually crushed by Austria. Within the nationally diversified empire of the Habsburgs the risings were violent enough in Vienna to force Metternich himself to resign, and to exact the promise of a new constitution; while in Hungary a new national leader appeared in Lajos Kossuth demanding home rule for the Hungarian population at the expense of Slav minorities within the kingdom. Here, too, deep divisions brought disaster to the nationalist cause, and by 1849 the Habsburgs were back in the saddle, ruling through the army and the bureaucracy.

In Germany a body representing the aspirations of all nationally minded Germans met at Frankfurt in May 1848 and sat for a year. But again divisions ruined its cause, for irreconcilable divisions appeared between the "great Germans," who wanted a federal state including all Austrian lands except Hungary, and the "little Germans," who looked to Prussia for leadership and wished to omit the mixed peoples of Austria in order to unify Germany more tightly. Roman Catholics tended to favour Austrian leadership, Protestants to favour Prussian. By 1849 the king of Prussia had withdrawn concessions made within his own kingdom, and refused the crown of a united Germany. Liberal nationalism was discredited throughout Germany.

The paradoxical outcome of the year of revolutions was military rule in Prussia and Austria, a restored confederation in Germany and a revived Bonapartism in France. But though frustrated in 1848, and subjected to even more severe repression by 1850, nationalists everywhere now had a clearer vision of how their aims could be reached. Governments, having tasted the strength of nationalist feeling, were forced to consider how they could harness it instead of merely crushing it. The system of Metternich had gone, and seemed even more old-fashioned than the heroics of romantic poets and secret societies. A mood of grimmer realism pervaded European affairs after 1850.

In 1850, as in 1815, it was the professional armies of monarchical governments that won; the intellectuals and romantic idealists went down in defeat. Again, strong monarchies seemed reaffirmed in their lease of power, and again the power of the Roman Catholic Church revived and its role in education grew. Yet a certain residue of more liberal government remained. Universal suffrage and parliamentary institutions were not abolished in France, though Napoleon III (as Louis Napoléon became) blunted their edge. In Prussia, as in Sardinia, parliamentary institutions remained. In Austria seignioral rights were not restored, and the new rulers of these countries differed in temperament and methods from the traditionalist monarchies of the restoration. Above all, the forces of profound economic change were now at work throughout the whole of Europe, making further changes in politics and in mental attitude inevitable.

4. New National States in the Balkans.—From the end of the 18th century the Turkish power was crumbling in southeastern Europe and the rival ambitions among the powers increased (*see* EASTERN QUESTION). Great Britain and France were interested in Egypt and the Levant, Austria and Russia in the near east and the Balkans. These problems entered a new phase with the stirring of liberal and nationalist movements in eastern Europe, especially in Poland and Greece. Greek independence involved joint action by Britain, France and Russia against Turkey (*see* above). When, in Dec. 1832, an internal crisis in Turkey prompted the sultan to appeal to Russia for help, the tsar Nicholas I sent an army and a fleet to defend Constantinople. In 1833 sultan and tsar signed a treaty closing the straits to other powers and creating a masked Russian protectorate. Seven years later Great Britain joined with Russia, Prussia and Austria to intervene and make a convention (1841) with the sultan whereby he undertook to close the straits to all warships. Each power showed its willingness to interfere in Turkish affairs when it felt that its own interests might be affected, and each feared the machinations of the others in this turbulent area.

It was these entangled interests that led, in 1854, to the first war between the former allies of 1815: the Crimean War (*q.v.*), in which Britain, France and Sardinia, allied to Turkey, made war on Russia. The Russians, defeated by the Turks on the Danube, evacuated the Rumanian principalities they had occupied, but the war continued in the Crimea until the fall of Sebastopol. Austria, as a nonbelligerent, discussed with the allies conditions of peace to be imposed on Russia. In 1855, when the tsar Alexander II succeeded his father, Russia agreed to the terms proposed; and at the congress of Paris (1856) the great powers, with Sardinia and the sultan, made a peace treaty with Russia. The sultan promised reforms in return for a guarantee by the powers of the integrity of his empire. The Black sea was neutralized and closed to ships of war. The Rumanian principalities were declared autonomous, and Russia restored part of Bessarabia to Moldavia. The congress was a revival of the idea of a concert of Europe, and its declaration respecting maritime law revived concern for a public law of Europe. None of the settlement was to prove durable. The chief beneficiaries of the war were Napoleon III and the king of Sardinia, both of whom gained in prestige. The Eastern Question was not settled.

This became apparent in 1876, when the Slav Christian subjects of Turkey in Serbia and Bulgaria again rebelled, and their savage suppression by Turkey attracted the attention of the powers. Russia went to war, and in the treaty of San Stefano (1878) arranged for the formation of an enlarged Bulgaria under Russian protection. Again the concert of Europe revived at the congress of Berlin (*q.v.*), attended by all the great powers, including the new Italy and the new Germany (*see* below), the only two powers that left it without some territorial gains. The settlement, however, left each power dissatisfied and more anxious than before.

5. The Remaking of Central Europe, 1850–71.—The schism between the forces of liberalism and nationalism which appeared by 1850, and the breaking of the spell of peace between the powers by the Crimean War, inaugurated a new era in great-power relationships. To forge nation-states and win their independence became the task of strong leaders of existing states, with armies and economic resources of their own, not the dreams of exiles or conspirators. Diplomacy and war were the new agencies of successful nationalism.

Economic changes reinforced this tendency. Railways made it possible to transport men and goods—including armies and war materials—rapidly from one part of Europe to almost any other. The expansion of industrial productivity opened the way for the rapid growth of the economy of any state ready to take advantage of these new opportunities. France under Napoleon III was well placed to seize these opportunities, and the second empire (1852–70) was an era of growth, prosperity and expansion under a paternalist, if harsh, regime. Exports exceeded imports, so that France, like Great Britain, could invest capital abroad. The Comité des Forges, heading France's great iron industry, was formed in 1864 and extended its interests into Belgium and Germany. A new French colonial empire was built up. Algeria was completely taken over by 1857, Tahiti and the Ivory Coast had been added earlier. Expeditions went to Peking in 1859–60 and to Syria in 1861. Explorers went to west Africa and new settlements

were made in Dahomey and the Guinea coast. New Caledonia was occupied in 1853. From 1859 onward three provinces were annexed in Cochin China and a protectorate was set up over Cambodia (1863). At home the emperor perpetuated Bonapartist traditions by his schemes of public works, notably the replanning of Paris. France in those years became, like Great Britain, a world colonial power and a larger, richer, more powerful state.

In central Europe two states were advantageously placed to seize the new opportunities of growth. One was the constitutional monarchy of Sardinia, Piedmont and Savoy, with its favourable strategic position in northern Italy, its enlightened government anxious to "westernize" the country, and its prestige as a leader of Italian hopes of national unification. The other was the kingdom of Prussia, already advanced in internal administration and in fiscal and military organization. With territories and interests widespread throughout Germany, it was the natural focus for all hopes of national unity excluding Austria. Austria of the Habsburgs was precluded from being a champion of nationalist aspirations anywhere, because its sprawling domains included peoples of many races and national loyalties and its survival as an empire depended on discouraging and suppressing movements of emancipation.

In 1852 Count Camillo di Cavour (*q.v.*) became premier of Piedmont; in 1862 Otto von Bismarck (*q.v.*) became prime minister of Prussia. The reshaping of central Europe, achieved by 1871, was pre-eminently the work of these two states and these two men.

Cavour, as Piedmont's minister of agriculture, commerce and marine in 1850, and soon also as minister of finance and premier, developed intensively his country's whole economy. He built or improved railways, roads, ports, docks; he expanded trade and linked Piedmont to the west by loans and commercial treaties. He spent money generously on the army, especially after his successes in the Crimean War. He agreed with Napoleon III in 1858 to make war jointly on Austria, and then provoked Austria into war (1859). After six weeks of fighting the Austrians were driven from Lombardy, and Napoleon abruptly made terms with Austria, on his own, at Villafranca. But he had precipitated the first triumph of Italian nationalism. Cavour gained Lombardy, but not, as he had hoped, Venetia. In 1860 Parma, Modena, Tuscany and Romagna adhered to his cause, while Garibaldi conquered Sicily and Naples, and Cavour annexed all the central Italian states except Rome itself. When the first Italian parliament met at Turin in Jan. 1861, the new kingdom included all the Italian peninsula except Venetia and the papal city of Rome. Cavour died in June 1861. The business he left unfinished was completed by 1870 under the momentum he had given: Venetia was added in 1866 when Prussia defeated Austria; and in 1870 Germany's defeat of France left open the road to Rome.

Bismarck, as a loyal servant of Prussia, regarded it as his duty to assert Prussian as against Austrian leadership of Germany. To this end he sought alliances with France and Russia. Prussia was growing in population and in power. Since 1815 its population had risen from 11,000,000 to 18,000,000. Its economy had expanded and, like Cavour, Bismarck was convinced of the need to spend lavishly on armaments and the army. He despised parliamentary assemblies and liberal ideas, and believed in the traditional *Junker* values of order, discipline, service and duty. But his task was formidable. Since 1815 Germany had been divided into 38 states, unequal in size and diverse in constitution, bound together loosely within the German confederation (*Bund*) which included Austria as well as Prussia.

Bismarck's first task was to drive Austria from the *Bund*. First Prussia joined Austria in defeating Denmark in 1864. As a result (convention of Gastein, 1865) they took control of the duchies of Schleswig and Holstein, Prussia being responsible for the administration of Schleswig, Austria for that of Holstein. The future of the duchies remained the joint concern of Austria and Prussia —a convenient bone of contention for future use. Then Bismarck isolated Austria diplomatically by promising Italy Venetia in return for Italian help against Austria, and by ensuring French neutrality. In June 1866 he forced the issue by proposing an end of the *Bund* and the election of a German assembly to draft a new constitution excluding Austria. After defeating Austria at Sadowa (Königgratz) on July 2, he made the treaty of Prague in August, annexing both Schleswig and Holstein and setting up a North German confederation under Prussian leadership. The southern German states formed a looser association "with an independent international existence." Austria undertook to claim no further part in German affairs, and surrendered Venetia to Napoleon to hand on to Italy. The efficiency of the Prussian army, the precision of the campaign and the skill of Bismarck's diplomacy achieved no less (and no more) than Bismarck intended. The war was a sharp blow to French prestige. It brought Italian and German unity one step nearer. It helped to bring about reorganization of the Habsburg empire into the dual monarchy of Austria-Hungary in 1867.

The process of German unification was completed by the Franco-German War (*q.v.*) of 1870. Napoleon III, his government at home weakened by growing liberal and republican opposition, his prestige shattered by the failure of his Mexican expedition and by the demonstration of Prussian power at Sadowa, sought feverishly but fumblingly for some form of "compensation." The conventions of diplomacy and the concepts of a balance of power required France to seek some material gains to offset the recent accessions of power to Prussia and Italy. Bismarck used these anxieties to precipitate war with France—a war in which France appeared as the aggressor. The pretext for conflict was the candidature of Leopold of Hohenzollern for the vacant throne of Spain; the possibility of a German prince on the Spanish throne infuriated France. But even when Leopold withdrew, the war parties of Paris and Berlin were so intransigent, the public feelings of the two countries so frantic, that France declared war in July. At the battle of Sedan the main French army, and even Napoleon III himself, were captured; and the war would have ended then had not a new republican government of national defense in France continued hostilities until Feb. 1871.

The war achieved Bismarck's aim. In Jan. 1871 King William of Prussia was proclaimed German emperor at Versailles. Germany was united under Prussian hegemony when the southern states combined with the North German confederation within a new imperial constitution. France, after an interlude of civil war in the Paris Commune (*q.v.*) of 1871, set up the third republic in 1875—a parliamentary regime of a moderate, conservative kind. Russia took the chance to repudiate the Black sea clauses of the treaty of Paris of 1856, and the other powers were too much preoccupied to do more than protest. A completely new balance of power existed in Europe, with Germany overshadowing France and Russia overshadowing Austria.

6. Rise of Socialism.—The history of European Socialism (*q.v.*), like that of liberalism and of nationalism, entered a new phase around 1850. Before mid-century, Socialist ideals derived partly from the democratic ideals of the French Revolution and partly from a wide variety of humanitarian, philanthropic and idealistic trends of thought and action. The leading Socialist thinkers had been men like Robert Owen in England and Charles Fourier in France, stirred by the social injustices and individual hardships of the industrial revolution with its exploitation of unorganized labour. They looked to enlightened employers, public conscience or new modes of economic and political organization to improve the material lot of the working classes. They went beyond this, too, and saw in a transformation of social ideas and habits the only path to a more just way of life for all. They pinned their faith on free association and education to bring progress. They wanted a new kind of society.

By mid-century Socialist thought, like nationalist thought, was moving toward greater reliance on state action and legislation. Louis Blanc in France, and radical reformers in Great Britain, came to urge more positive action by governments to protect workers in factories. They sought to implement the "right to work" by policies of poor relief and provision of employment in "national workshops." These more empirical and realistic proposals were in sharp contrast with two other strands of Socialism: the idealist and visionary Socialists, who encouraged migration to

the American frontier to form simple egalitarian communities practising communism in primitive agrarian conditions; and the Socialists of the secret societies and underground conspiracies, who plotted mass uprisings in the capitals of Europe to overthrow existing regimes and establish extreme democratic republics based on universal suffrage.

All alike suffered severe reverses in the *débâcle* of 1848–49. The insurrectionary followers of Auguste Blanqui in France, the international supporters of Filippo Buonarroti in France, Italy and Belgium, attempted risings that were brutally crushed. Louis Blanc, who served for a time in the provisional government of the second French republic of 1848, found his aims frustrated and discredited. The only successful Socialist movements by mid-century were the moderate trade-union movements and the co-operative movement promoted in Great Britain by Robert Owen. They were not revolutionary and not even political in purpose. The Industrial Revolution had not yet produced the social conditions needed for a broad and effective Socialist movement anywhere. In every European country approximation to universal (male) suffrage gave power to the peasants, who proved, by their support of reaction in France or in Hungary in 1849, to be a predominantly conservative force. But the movement from countryside to town continued everywhere, and the existence of large industrial urban populations would change the whole situation.

In 1848, on the eve of the ill-fated revolutions, Karl Marx and Friedrich Engels issued their Communist Manifesto. It attacked all previous Socialist movements as "utopian" and based its program of revolution on an analysis of the inherent tendencies in capitalist society. Finding in private ownership of the means of production the source of inevitable class war between those who own the means of production (*bourgeoisie*) and those who have only their labour to sell (proletariat), they saw in the growing class of industrial workers and landless peasants the truly revolutionary class of the future. Only when this class seized power, liquidated the *bourgeoisie* and instituted a dictatorship of the proletariat could modern society move forward to a completely classless, Communist society.

Their famous pamphlet had no effect in 1848. But in the aftermath of disillusionment during the 1850s their arguments found readier adherents. An industrial exhibition in London in 1862 brought together British and French working-class leaders, and by 1864 an international working men's association was formed by delegates from Great Britain, France, Germany, Poland, Italy and other countries (*see* INTERNATIONAL, THE). Marx drafted its rules; it organized periodic conferences and became the basis of the First International. Marx deepened and strengthened his analysis in his study of *Das Kapital* (of which the first volume appeared in 1867). He began to attract disciples in various countries, especially in his native Germany.

The events of 1870–71 had both adverse and beneficial effects on the fortunes of Socialism and Communism. The Paris Commune of 1871 was not essentially Marxist in its inspiration, though some Marxists took part in it. Its defeat was a grave reverse for all Socialist and working-class movements in France. Marx depicted it (in *The Civil War in France*) as a prototype of the new proletarian revolutionary technique and hailed it as a sign of hope. In fact, it was the last flicker of the old Jacobin tradition of the barricades, and by leading to the killing or exiling of Socialist leaders and discrediting revolution it secured a highly conservative regime in France for the next 70 years. On the other hand, the creation of unified states in Italy and Germany, each with parliamentary institutions within which Socialist parties could operate, paved the way for the growth of large and powerful Social Democratic parties during the next generation.

In every European country except Russia such parties appeared and thrived on the spread of industrialism and urban development before 1914. In some, such as Great Britain and the Scandinavian countries, they formed close links with trade-union and co-operative movements. In others, such as France and Italy, trade unions grew stronger but remained aloof from political parties. But in most countries Socialist movements were split between those who favoured constitutional parliamentary methods and sought reforms through legislation, and the more orthodox Marxist parties, which shunned the compromises of parliamentary life and aimed at revolution. The First International died in 1876, disrupted by quarrels between Marxists and anarchists. But in 1889, centenary year of the French revolution, the Second International was founded and continued to hold periodic congresses until 1914. The congresses were haunted by dissensions between reformist and revolutionary Socialists and by disputes between the French and German parties on nationalistic lines.

In Russia the severity of the tsarist autocracy made reformist policies useless, and there grew up a variety of revolutionary movements, among them the Russian Social Democratic party, led by the Marxist G. V. Plekhanov. At its meeting in 1903, in London, Lenin (V. I. Ulyanov) succeeded in getting its membership restricted to those who "personally participate in one of the organizations of the party." This decision kept it a tightly disciplined, militant body, unlike the broader Social Democratic parties of the rest of Europe. Lenin's party, the Bolsheviks, was to prove an effective instrument of revolution in 1917 (*see* below).

How far the spread of Socialism, of all kinds, marched in step with the growth of industrialization was shown by events in Russia. The older social revolutionaries concentrated on propaganda among the peasants. What made Marxism more prevalent was a remarkable phase of intensive industrial growth in the late 19th and early 20th centuries. The heroic age of railway building in Russia came after 1890, culminating in completion of the Trans-Siberian railway between 1900 and 1905. Iron, coal and petroleum industries, already started with foreign capital, passed through a boom as a result of railway construction on a vast scale. Depression set in between 1903 and 1908, and in these years too Russia suffered first the Russo-Japanese War (*q.v.*), in which it was defeated, and then the revolution of 1905. From 1908 to 1914 Russia enjoyed a period of rapid industrialization. It was in the western regions, most affected by this process, that Bolshevism and rival Socialist movements made headway. The numbers of workers employed in mines and factories rose from 1,425,000 in 1890 to 2,931,000 in 1912. The prewar years saw the growth of trade unions, though strictly limited by law, and epidemics of great strikes. It was with this new human material that the Bolsheviks worked.

These same years in western and central Europe also brought powerful trade-union movements and recurrent strikes and riots. Syndicalist theories and movements, adapting Marxist theories to the special claims of organized labour movements, found adherents, paradoxically, in the less advanced industrial countries (Russia, Spain and Italy) where social conditions were worst. Already the more advanced industrial societies were reaping benefits in a higher standard of living, and much of the unrest there came either from periodic slumps and dislocations or from a slowing down in the accustomed rate of material improvement. Such conditions favoured reformist rather than revolutionary attitudes.

7. Diplomacy and War, 1878–1914.—For nearly a decade after 1871 France, Italy, Germany and Austria-Hungary were preoccupied in domestic reorganization and consolidation of the new regimes. French resentments over defeat and the loss of Alsace-Lorraine led to sporadic tensions with Germany. Bismarck was anxious to isolate France and keep an unbalance of power in Germany's favour. In 1873 he made the League of the Three Emperors (*Dreikaiserbund*) between Germany, Russia and Austria-Hungary. It was a general agreement to work in concert to preserve the peace. Bismarck, confronted with a series of domestic wrangles with Liberals, the Roman Catholic Church and the growing Social Democratic party, was as anxious now to preserve peace in Europe as he had been ready, in the previous decade, to break it in the interests of Prussia. The same considerations led him to revive the concert of Europe at the congress of Berlin (1878), where he could act as "honest broker" over the Eastern Question.

The resilience and recovery of France, and its refusal to accept as permanent the loss of Alsace-Lorraine, led him to keep France isolated by making positive alliances of his own. In 1879 he made a secret treaty of alliance with Austria-Hungary, for

mutual aid should either be attacked by Russia. This, however, encouraged trends toward an entente between France and Russia, for Russia feared Austrian designs as much as Germany feared French. Seeking a third ally, Bismarck found one in Italy, alienated from France by the establishment of a French protectorate over Tunisia in 1880. In 1882 Italy entered into two secret treaties of defensive alliance with Germany and Austria. If Italy were attacked by France, its two allies would support it, and Italy in return would back its allies against any two powers other than Great Britain. Thus was formed the triple alliance, in basis a reciprocal defensive alliance against the warlike intentions attributed to Russia and France. But it reinforced the hegemony of Germany, the strongest of the three partners, and seemed to give the new German empire a permanent superiority in European affairs.

This predominance of Germany in the heart of Europe was emphasized economically even more than diplomatically; for German industrial, commercial and financial expansion began to outpace that of all its neighbours. Germany had at its disposal the rich mineral resources of the Ruhr, the Saar and Alsace-Lorraine, and the unified, growing labour force of the *Reich*. Within the *Reich* rapid development of all means of transport and communication welded the country into one great economic unit. From 1879 onward its agriculture and industry were protected by tariffs. It developed great new modern chemical and electrical industries, in which its technical education and skills and its rich resources gave it a lead over all other states in Europe. It manufactured for export, so came into open competition with British trade and industry in world markets. As the great central land power in Europe it became the focus of the continental network of communications and transport. Its trade flowed irresistibly into eastern Europe, the Balkans and the near east. It developed a large shipbuilding and shipping industry, based on Bremen and Hamburg, so that by 1914 Germany had the second biggest merchant marine in the world, exceeded only by that of Great Britain.

Against this background of dynamic economic growth, Bismarck's diplomacy of alliance-making seemed especially alarming for the security of Germany's neighbours. The arrangements of the triple alliance were renewed in 1887. There were several obstacles to a *rapprochement* between Russia and France: the contrast between an autocracy and a democratic republic; Bismarck's skilful diplomacy toward Russia, which kept alive some memory of the *Dreikaiserbund* and the fears that had prompted it; Russia's wish to avoid specific commitments. These obstacles diminished with the collapse of Boulangism as a disturbing force in France (*see* BOULANGER, GEORGES ERNEST JEAN MARIE) and with the dismissal of Bismarck in 1890 by the young emperor William II, who took policy more into his own hands. Between 1891 and 1894 France and Russia succeeded in making a mutually defensive agreement comparable with the dual alliance treaty. Military conventions envisaged simultaneous mobilization against Germany in the event of a defensive war. The balance of power was partially redressed.

The impetus of German expansion was much too great to be checked by such diplomatic maneuvers, however, and William II gloried in German growth and power. With a sense of being late in the race and of being restricted in Europe, Germany sought fresh outlets, economic and strategic. Schemes for *Mitteleuropa* (Germanic economic union of central Europe), for a *Drang nach Osten* (expansion into the European and near east), of *Kolonialpolitik* (the acquiring of overseas colonies) were variant expressions of this urge, and all bred anxieties and conflicts with Germany's neighbours. It was the German bid for naval power that did most to bring about British hostility. By the 1890s Germany was indisputably the greatest military power in Europe and the heart of its economy. For the largest continental land power to become also the greatest naval power meant a bid for world supremacy which challenged the security of all other states. Germany's flamboyant and clumsy diplomacy after the dismissal of Bismarck in 1890 did nothing to dispel such fears, and German expansionists made no secret of the aims of their *Weltpolitik*. Anxieties in the colonial and naval fields drove France and Great Britain closer together, while the German drive into the near and the far east alarmed France's ally, Russia. This complex of forces welded, by 1907, the triple entente of France, Russia and Great Britain—a rallying of the only three powers that, together, could act as a counterpoise to the triple alliance. Germany accused it of aiming at its encirclement.

This system of rival grand alliances, completed by 1907, was an important factor in bringing about World War I; therefore the causes and the nature of the system require more precise analysis. The dual alliance, as made by Bismarck and flanked by the *Dreikaiserbund* and his other devices, was intended to keep the peace by isolating France and establishing a deterrent unbalance of power in Germany's favour. Since Austro-Russian relations were uneasy he could not contrive to bring both into a solid triple alliance; but he kept Russia half-affiliated and brought in Italy as a make-weight. His principle was "to try to be a party of three in a world governed by five powers." Once France and Russia reached an entente of a similar kind the main buttress of his system crumbled. Thereafter Germany's only hope of preserving a balance of power in its favour was to keep Great Britain isolated or neutral, if not friendly. This hope persisted until as late as 1904. During the 1880s the scramble for Africa occasioned a series of frictions between France and Great Britain: over Egypt and Morocco, and above all over the Fashoda incident of 1898. William II entered into friendly relations with Great Britain, and in 1890 agreed to give up Heligoland in return for concessions in east Africa. The hope was severely blighted by British experience of its hostile isolation during the South African War (*q.v.*), and by William's anti-British behaviour then. When Great Britain at last broke from its policy of so-called splendid isolation in 1902, it was to make a treaty of alliance not with Germany but with Japan, then emerging as a great expansionist power in the far east. Any surviving hope of reconciliation with Great Britain was ruined by Germany's naval program, regarded as a frontal challenge to British sea power. This made the Franco-British entente inevitable; for, once Great Britain was in the market for European allies, France, in the situation of 1904, was the only possible partner.

Even then, completion of the triple entente was beset with obstacles. In 1904–05 Great Britain's ally, Japan, defeated in war France's ally, Russia. British distrust of Russian aims in both southeastern and eastern Asia was fed by fresh evidence of its interest in Persia, Afghanistan and Tibet. But, when settlement in this area was agreed in 1907, the path was clear for making the triple entente. Once made, German attempts to disrupt it in 1906 at the Algeciras conference about Morocco, in the Balkan crisis (1908) or in the Agadir crisis (1911) served only to strengthen it by showing the entente powers the dangers of their separation.

Although international incidents and tensions caused by the scramble for colonies contributed to the formation of the system of alliances, there is little evidence that colonialism otherwise played a major part in causing the eventual outbreak of war. The gains to be had were so vast that agreements to share were more attractive propositions than armed conflict. In 1875 less than one-tenth of Africa had been turned into European colonies; by 1895 only one-tenth remained unappropriated. At the Berlin conference of 1884–85 the great Congo basin was amicably enough partitioned among the powers concerned, and it was there agreed that any power effectively occupying African territories and duly notifying the other powers could thereby establish possession. Spheres of influence were repeatedly agreed, not only in Africa but also in China, Persia, New Guinea and other parts of the Pacific. It was after 1908, when the active rush for colonies had ceased and there was little left to share, that hostilities between the great powers became more probable.

From this time onward more irreconcilable conflicts dominated international relations: the Anglo-German naval rivalry; the armaments race between the powers, unchecked by proposals for disarmament at the Hague conference of 1899; and, above all, the clash of old-style dynastic imperialisms in eastern Europe, revived by the challenge of insurgent Balkan nationalisms. The tensions

of the Eastern Question, much more than colonial conflicts, were the immediate cause of war in 1914. They were accentuated in 1908, when Austria annexed Bosnia and Hercegovina. When the heir to the Austrian Habsburg throne and his wife were murdered in the Bosnian town of Sarajevo in June 1914, and the government of Vienna set about crushing Serbia as a separate state, it was because it saw in Serbia a new Piedmont or a new Prussia, the nucleus of yet a third major movement for national unification that could triumph only at the expense of Austria-Hungary. Serbia had, indeed, become the core of a movement to unite all southern Slavs (Serbs, Croats and Slovenes) into an independent kingdom, and the movement was backed by Russia. The Austro-Serbian issue raised others of wider interest. With Serbia under Austrian control, German-Austrian influence would prevail from Berlin to Baghdad. An independent Serbia was a wedge in this German-Austrian-Turkish bloc and a channel for the influence of the rival triple entente. The conflict not only revived all the passions of the Eastern Question: it was now a trial of strength between the rival alliances.

The Sarajevo murders were the climax to a sequence of crises. These included Italian occupation of Tripolitania in 1911, which weakened Turkey; the Balkan Wars (*q.v.*) of 1912 and 1913, which further weakened Turkey, strengthened Serbia and were regarded as a reverse for the triple alliance; and increasing Anglo-German naval competition. Accordingly, once Serbia was attacked by Austria-Hungary, Russia went to its help, Germany to that of Austria, France to that of Russia, Great Britain to that of France. In the prevailing mood of intense fear, none dared desert its ally. The alliances, born of a desire for security, brought a decade of insecurity and eventually war. They also ensured that war, when it came, should be a general war of unlimited liabilities.

8. European Civilization in 1914.—The continent that embarked on war in 1914 was entirely different in character from the Europe of 1815. It had a complete and well-founded faith in human progress, and especially in the material progress of European civilization. In little more than a single lifetime it had added immensely to the numbers, longevity, health, wealth, literacy and knowledge of mankind. New continents had been peopled, developed and explored by Europeans. Men who could build Trans-Siberian and Canadian Pacific railways, dig Suez, Kiel and Panama canals, construct skyscrapers and oceanic liners, could surely master the world. The accomplishments of science and technology opened infinite vistas of further progress: man could at last subdue nature to his ends.

New learning was mostly scientific learning. Research into electricity and radioactivity revealed the relationships between energy and matter and led to new theories of atomic structure, of atomic energy and even the possibility of transmutation of elements. Albert Einstein in 1905 first propounded the new mathematics of the theory of relativity, only a year after Ernest Rutherford published his book on *Radioactivity*. Just as Charles Darwin had shattered the old theology as well as the old biology by his theory of evolution through natural selection, so at the turn of the century Sigmund Freud and his Austrian associates gave new understanding of the working of the human mind and of the subconscious impulses in man. At the same time French and German sociologists, pre-eminent among them Émile Durkheim and Max Weber, studied the behaviour of men in society and in social groups. Knowledge of man, as well as of nature, attained new heights when the 20th century dawned. The originality and the far-reaching implications of the new learning produced a crisis of culture and thought.

This crisis was reflected in philosophy and religion as well as in the arts. The philosophers gaining in influence were Friedrich Nietzsche and Henri Bergson, equally at variance with the more traditionalist schools of positivism and idealism. Modernism grew in Catholic, Protestant and Jewish theologies. The novel and drama increasingly concerned themselves with social or psychological issues; painters and poets became obsessively introspective; the arts tended to become esoteric. Conventionalism and established forms were everywhere challenged and discarded. This trend was in conformity with the faith in progress; enterprise, experimentalism, emancipation were the keynote of this amazingly vigorous and versatile civilization. The age had brought emancipation to slaves, to serfs, to nations, to women, to artists. But states and peoples could not emancipate themselves from fear, from the drift to destruction, from war. The civilization that accomplished so much progress, so much emancipation, was also beset with moods of mass hysteria and nihilist violence. The price of its vitality was tension and chronic instability.

H. The 20th Century

1. World War I and the Russian Revolution.—The war that began in 1914 was the first war of the masses, the first between nation-states able to command the energies of all their subjects and the resources of modern industrial technology, the first on a scale large enough to disrupt the world economy which now hinged on Europe. It differed completely from the wars of the second half of the 19th century; these had been engagements limited in purpose and scope, whereas the war of the grand alliances was as complex in purpose as it was unlimited in commitment. It was unprecedented in European history (*see* World War I).

Russia, France and Great Britain being allies, Germany was from the start confronted with a war on two fronts. Its inevitable purpose was to crush France before Russian resources could be fully brought to action in the east. Its initial effort to achieve this by means of the Schlieffen plan failed in 1914. Thereafter, fighting on the western front became a war of attrition, with millions on both sides locked in close and indecisive struggles extravagant in men and materials. Artillery and the machine gun made each attack so costly in lives and munitions that soon each side had to recruit mass armies and expand munitions production on a scale neither had anticipated. On the eastern fronts warfare was more mobile but hardly less expensive in human life.

The strain on national resources, financial and industrial but even more of discipline and morale, was immense. In the end it shattered the cohesion of the weaker dynastic empires of Turkey, Austria-Hungary and Russia, and even the more solid western states experienced mutinies and political crises. Yet the most remarkable feature of the war was the amount of hardship and sacrifice modern society could endure without disintegrating. The German submarine war on British, Allied and eventually neutral shipping, and the Franco-British blockade of the central powers, alike imposed on civilian populations severe shortages of food and supplies. For all these reasons a new concept of total war emerged. Governments controlled or commandeered supplies, trade, credit, transport and manpower more fully than ever before. The war accustomed nations to centralized power and widespread governmental activities. Total war made possible the 20th-century conception of the totalitarian state. Walter Rathenau, the industrialist who organized the German economy for war, was the forerunner of Lenin, who made Russia, after 1917, the first of the single-party totalitarian states.

The war could not be confined to Europe, if only because the British empire drew its strength from all continents and because the other great colonial powers spanned the globe. But Europe remained the great battlefield, and until the United States entered the war against Germany in April 1917, the war at sea was sustained mainly, on the Allied side, by Great Britain and France. Entry of the United States, followed within the year by the withdrawal of Russia, changed the whole character of war aims. The war now became more truly a struggle between the western, democratic, maritime powers and the autocracies of central Europe. Pres. Woodrow Wilson's Fourteen Points (*q.v.*), enunciated in 1918, provided a general ideological purpose that had not been present in 1914. This purpose, rooted in the doctrine of national self-determination and sovereignty, fitted the aims of all the insurgent nationalities released by the collapse of the dynastic empires. It predetermined the general pattern of postwar Europe.

But the postwar situation was shaped, too, by an unforeseen by-product of the struggle: the collapse of the tsarist regime in Russia in March 1917, the seizure of power in November by the Bolshevik party led by Lenin, and the unshakable consolidation of this dic-

tatorial regime during the years of chaos and turmoil after the war. The triumph of the Bolsheviks in Russia and their failure to set up permanent Communist regimes anywhere else introduced an incalculable new factor into the postwar situation. The chain of small nation-states that appeared in eastern Europe, justifiable on the principle of national self-determination and independence, was welcomed by the allies as a *cordon sanitaire*—a quarantine belt against the spread of Communism into Europe. It was never more than a forlorn hope of the western powers that so flimsy a barrier could hold against a Russia reorganized and expanded into a world power by successful revolution. Yet, when efforts to overthrow the regime by a war of intervention had failed by 1921, this alone remained.

Germany's allies were the first to crumble in 1918. To resist the final German spring offensive of 1918 the Allies set up, at last, a unified command under Ferdinand Foch. By now problems of supply, transport and manpower were being solved by the war effort of the United States as well as by the continued efforts of France and the British empire. The tank, first used in 1916, was now used to greater effect. The German offensive failed, the Allied counteroffensive succeeded. On Nov. 9, 1918, William II abdicated and the recently formed liberal government signed an armistice on Nov. 11. It signed on the urgent request of the German high command, though there were no Allied armies on German soil. A peace conference, much more widely representative than that of 1814, assembled in Paris in Jan. 1919. Eventually treaties of peace were signed with Germany (Versailles, June 28), Austria (St. Germain, Sept. 10), Bulgaria (Neuilly, Nov. 27), Hungary (Trianon, June 4, 1920) and Turkey (Lausanne, July 24, 1923).

2. Interwar Years.—The peace settlement was dictated by the Allies, mainly by the leaders of the Big Three powers: Pres. Woodrow Wilson of the United States; Georges Clemenceau, premier of France; and David Lloyd George, prime minister of Great Britain. It was a compromise not only between the aims of these three men, but also between the avowed peace aims of the victors and the brute facts of the situation at the end of the war. Many expectations were inevitably disappointed.

The total number of European states was increased. Czechoslovakia, Poland, Estonia, Latvia, Lithuania and Finland appeared on the map; Austria and Hungary survived as two small, separate, landlocked states, while European Turkey was restricted to a narrow coastal strip along the straits. Serbia, with the addition of the southern Slavs of Croatia and Slovenia, became Yugoslavia (though not adopting that name until 1929). Rumania was increased in size at the expense of Hungary and Russia. France recovered Alsace-Lorraine; Belgium and Poland made gains at the expense of Germany; and the area of the Saar was temporarily taken over by an international commission. Otherwise Germany remained territorially intact and set up the democratic Weimar republic. These arrangements left large national minorities on the wrong sides of frontiers; notably minorities of Germans and Ukrainians in Poland and of Hungarians in Rumania. Although the new succession states made treaties governing their treatment of such minorities, the settlement involved one-sided violations of the principle of national self-determination to the consistent disadvantage of the defeated nations.

The new states, as well as Germany, adopted democratic constitutions. Democracy was in vogue, its reputation enhanced by the victory of the western powers, its ideals trumpeted in Allied proclamations and translated into the covenant of the new League of Nations (*q.v.*) which, on Wilson's insistence, was included in each peace treaty. Events proved that the unfamiliar institutions, procedures and values of democracy were in such countries fragile and ill-fitted to survive the turbulent aftermath of war. By the end of 1926 Hungary, Italy, Portugal, Poland and Lithuania were all under authoritarian or dictatorial regimes, and elsewhere militarist and authoritarian forces challenged the working and the survival of democratic governments. The prevalence of democracy in Europe was short-lived.

The League of Nations, too, quickly disappointed hopes that it would be an effective medium for co-operation among the major powers of the world. The U.S. senate did not ratify the peace treaties by the necessary majority of two-thirds, and so did not endorse the covenant of the League. Not only was the United States never a member of the League, but the U.S.S.R. was not admitted to membership until 1934, and by then Japan and Germany had ceased to be members. The League was much weakened in power and prestige by the absence of never less than two of the world's greatest powers. The task of making it work devolved on the main western Allies, France and Great Britain; and their aims and policies diverged during the testing time of the 1920s.

Because the settlement, in all these ways, was less stable and acceptable than its makers had originally hoped, they set about supplementing and reinforcing it by other means. France made treaties of alliance with Belgium (1920), Poland (1921), Czechoslovakia (1924), Rumania (1926) and Yugoslavia (1927). The last three states also formed a little entente (*q.v.*) among themselves. But these alliances formed no coherent system, since they were not automatic in character and set up no co-ordinated system of mutual defense. Aimed partly against a resurgence of Germany and partly against Hungary, they were effective only so long as Germany was kept disarmed, Russia internally preoccupied and the two states did not join forces. The treaty of Rapallo (*q.v.*), a treaty of friendship signed by Germany and the U.S.S.R. in 1922, was a sign that this last contingency was far from impossible. Neither France nor its eastern allies found firm promise of security in the new system of alliances.

The Geneva protocol for the pacific settlement of international disputes (1924), the Locarno treaties (*see* LOCARNO, PACT OF) of 1925 between France, Germany, Great Britain, Italy and Belgium which paved the way for Germany's entry into membership of the League, the Kellogg-Briand pact of 1928 renouncing aggressive war as an instrument of national policy, were all further expedients to reinforce the already shaky settlement. Such arrangements, if not incompatible in principle with the covenant of the League, carried implications often alien to it. Thus separate national alliances and frontier guarantees had point only if the system of collective security and guarantees embodied in the covenant were regarded as unreliable. The spirit of good will and optimism that prevailed by 1925 had, however, good effects on the pacification of Europe. Supported by an improvement in general economic conditions, the mood of Locarno seemed a happy fulfillment of the high hopes of 1919 and of earlier wartime dreams.

It was economic collapse that most drastically undermined the trend of the 1920s toward stability and peace. The war had wrought havoc in the economic system of Europe, and in Europe's relations with the rest of the world through international trade and investment. Russia was no longer an integral part of Europe's economy; Germany, the economic heart of prewar Europe, was thrown into chaos by the currency collapse of 1923 and by the demands of the Allies for reparations (*q.v.*). France, wonderfully resilient after 1871, was haunted by instability of currency; and most countries suffered from periods of severe unemployment. Economic dislocation proved to be a greater enemy of stability and prosperity than the sheer destruction of wartime. In eastern Europe lack of capital, high birth rates and shrinking death rates, diminished opportunities for emigration, political instability and nationalist unrest, all conspired to perpetuate low standards of living and unemployment. European countries as a whole encountered in world markets increasingly severe competition from the more industrialized nations overseas.

Collapse came with the crash on the New York stock exchange in Oct.–Nov. 1929, though the world economic crisis had causes wider and deeper than the reckless boom that ended so dramatically in Wall street. The chronic dislocations caused by war and its aftermath, which national policies and international co-operation had proved incompetent to overcome, were one major cause. Another was the long-term depression in the prices of primary products, including foodstuffs; another the secular trade cycle itself. The prevalent psychology of panic—born of insecurity and fear—deepened the consequences of the crash. Governments were slow to act, and when they did they often took measures later regarded as wrong.

The social and political consequences of the depression were as

disruptive as the economic consequences of the war. Financial bankruptcies, still further contraction of trade and more mass unemployment brought active social unrest and lent new vigour to extremist political movements. In Germany, especially sensitive and exposed to the slump, both Communists and National Socialists made large gains in the 1930 election to the *Reichstag*. By the beginning of 1932 Germany had more than 6,000,000 unemployed. It stopped paying reparations, which were by agreement liquidated at the conference of Lausanne in June–July 1932. Great Britain, its figures of unemployed exceeding 2,000,000, passed through a political crisis which led to the formation in 1931 of a national government headed by J. Ramsay MacDonald. Britain's abandonment of the gold standard was followed by most other countries. The U.S.S.R., partially sheltered from the economic blizzard by its planned economy (the first five-year plan was launched in 1928), nevertheless passed through a crisis of its own. Collectivization of agriculture led to severe dislocation of production, famine, riots, mass deportations and terror. The dictatorship of Stalin became increasingly savage and tyrannical.

In Europe the aftermath of the economic depression was political revolution, first and most significantly in Germany. On Jan. 30, 1933, Adolf Hitler, leader of the National Socialist party, became chancellor (*see* NATIONAL SOCIALISM). This movement, dating from the years of turmoil immediately after the war, had sporadically gained in strength by exploiting every difficulty of the Weimar republic. Appealing to the aggrieved nationalism of the military and conservative classes, the embittered and ruined middle classes and the mass of unemployed workers, it was avowedly antiparliamentary, antidemocratic, anti-Socialist and anti-Communist. It aimed at total overthrow of the Versailles settlement and the resurgence of Germany as the dominant power in Europe. By its violently anti-Semitic racial theory it exploited, through skilled propaganda, the deepest impulses of mass hysteria and blind hatred. With this dynamic, ruthless, revolutionary movement in control of Germany, the postwar era gave place to a prewar era: a change made doubly certain by the refusal of the western powers to perceive that this was happening and by events in the far east, where in 1931 Japan had attacked Manchuria and in 1933, along with Germany, left the League of Nations.

The recovery of Germany's power in Europe began with one false move, Hitler's attempt to bring about union (*Anschluss*) with Austria in 1934, which was frustrated by Austrian and Italian reactions; and with one legitimate concession—the return to Germany, by plebiscite, of the territory of the Saar (March 1935). On March 16 Hitler announced that Germany would no longer consider itself bound by the military clauses of the Versailles treaty, and embarked openly on rearmament, which had already, for many years, proceeded secretly. In 1936 he similarly repudiated the Locarno agreements and reoccupied the Rhineland with German troops. He began to build the Siegfried line of fortifications, counterpart to the Maginot line constructed by the French along the same frontier.

Germany's program of planned aggressions was facilitated by other developments, both national and international. France was beset by Paris riots in Feb. 1934 which threatened the survival of parliament, and by a series of weak governments and prolonged ministerial crises. When, in 1935, Italy attacked Ethiopia (Abyssinia), a fellow member of the League, the council of the League belatedly and ineffectually imposed economic sanctions on Italy (Nov. 1935). This did not prevent Italy's conquest of Ethiopia, but led in Oct. 1936 to the Rome-Berlin axis, agreements to cooperate made with Hitler by the Fascist government of Benito Mussolini. Meanwhile the reality of this alinement of the dictatorships was proved by their joint support for Gen. Francisco Franco in the Spanish civil war that had begun in July 1936. France, under the popular front government of Léon Blum from May 1936 onward, was preoccupied with overdue domestic reforms. It felt obliged to follow, along with Great Britain, a policy of "nonintervention," though the U.S.S.R. sent assistance to the Spanish republican forces. The U.S.S.R. itself, in these years, was engaged in far-reaching party purges and sensational trials of military leaders. In most European countries native Fascist leagues and paramilitary groups rivaled Communist militants in causing internal disorder. The whole of European society seemed to be lapsing into disorder and disunity.

Taking advantage of these circumstances, Hitler proceeded rapidly with his aggressions. In March 1938 he invaded and annexed Austria. Six months later he launched a campaign against Czechoslovakia, demanding cession of the so-called Sudetenland; and by the Munich agreements of Sept. 1938 he contrived to involve Great Britain, France and Italy in exacting this concession from Czechoslovakia. In March 1939 he invaded the remainder of Czechoslovakia in violation of assurances given; and the following month Mussolini took Albania. In August Hilter stunned the western powers by making a pact with the U.S.S.R., hitherto presented by Nazi propaganda as Germany's deadliest enemy. On Sept. 1, 1939, he attacked Poland. Since Great Britain and France had meanwhile guaranteed Poland against German attack, and had come to see that German aggressions could be checked only by war, they declared war on Germany on Sept. 3. Soviet troops occupied the eastern provinces of Poland, so completing yet a further partition of that unhappy country. Italy declared itself a nonbelligerent. World War II began, therefore, as a conflict between Germany and the western allies of France and Great Britain, backed by the Commonwealth of Nations.

3. World War II.—There were many affinities and parallels between the two great wars that engulfed Europe within a single generation. Both began in eastern Europe; both arose from treaty obligations of great powers to smaller powers; both involved an initial alliance between Great Britain and France against a German-dominated central Europe; both came to implicate Germany in a war on two fronts, and were won by a grand alliance of Great Britain, the U.S.S.R. and the United States; both changed their dimensions and their character as they proceeded, until they engaged most nations of the world; both left behind tangled problems of reconstruction and resettlement and the seeds of future conflict; of each the outcome was largely unforeseen and unintended when it began. There were also important differences. The second, more truly than the first, was a world war, for it involved prolonged fighting in the Pacific as well as the Atlantic, in Asia and Africa as well as Europe, the defeat of Japan as well as Germany (*see* WORLD WAR II).

The war in Poland ended with the surrender of Warsaw (Sept. 27, 1939). A brief war began between the U.S.S.R. and Finland (Nov. 1939–March 1940), for which the U.S.S.R. was expelled from the moribund League of Nations (Dec. 1939). The small Baltic states signed treaties of mutual assistance with the U.S.S.R. Great Britain and France, unable to influence these events, hastily prepared for Hitler's expected onslaught in the west.

Until April 1940 little happened. These early months, known as the "phony war," gave the west more time to prepare but also sapped French morale, for the strain of full mobilization proved considerable. Then with great speed German forces invaded Norway and Denmark in April and in May struck at the Netherlands and France. In Great Britain, Winston Churchill had just succeeded the prewar premier, Neville Chamberlain, unhappily associated with the Munich settlement and the policy of "appeasement." In France, Paul Reynaud tried to rally France for the blow. By June 17 the German army, highly motorized and with good air support, had swept into France and a new government, headed by the aged Marshal Philippe Pétain, sought an armistice. It was signed on June 22. German forces occupied northern France, Paris and a western coastal belt. A French government at Vichy, under Pétain, remained in office to govern France under these severe restrictions.

Since Great Britain, without a foothold on the continent after the retreat from Dunkerque (May 1940), refused to sue for peace, Germany launched a concentrated air bombardment of Britain in preparation for invasion by sea. By September the attempt to gain air supremacy over Britain had failed, and Hitler in October made an abortive offer of peace. Italy, having declared war on defeated France, now attacked Greece (Oct. 28, 1940). But Greek forces inflicted defeat on the Italians, and in April 1941 Germany declared war on both Greece and Yugoslavia. British help proved

of no avail and by May 1941 British troops were expelled from the Balkans. By now, however, movements of internal resistance had formed in all the German-occupied countries. In London Gen. Charles de Gaulle had formed a Free French movement challenging the authority of Vichy, while in Yugoslavia guerrilla war started against the Germans and Italians. (*See* GUERRILLA WARFARE; *World War II.*)

The war entered a new phase during the second half of 1941. On June 22 Germany attacked and invaded the U.S.S.R.—at first winning great victories. On Dec. 7, Japanese aircraft attacked U.S. warships in Pearl Harbor, Hawaii. The United States declared war on Japan (Dec. 8) and three days later Germany and Italy declared war on the United States. With dramatic suddenness the war was transformed from a war in Europe and the Atlantic to a world war involving all the great powers on all the oceans of the world. The United States, already an "arsenal of the democracies" now became an ally.

In Nov. 1942 the Allies, led by U.S. commanders, invaded French north Africa, and British forces won victories in Egypt. The Germans occupied the whole of France, but in Russia they suffered their greatest defeat yet, at Stalingrad. Henceforth the tide of war began to run in favour of the western Allies. Italy, suffering severe defeats at sea and with Sicily invaded (July 1943), underwent an internal revolution that overthrew Mussolini and disbanded the Fascist party. Marshal Pietro Badoglio took charge, on Sept. 8 surrendered and on Oct. 13 declared war on Germany. Germany occupied northern Italy.

The western air forces, with strong bases in north Africa and southern Italy as well as in Britain, launched concentrated bombing attacks on Germany, in preparation for invasion of German-occupied Europe. On June 6, 1944, the Allies landed in Normandy and by Aug. 25 liberated Paris. At the same time Soviet forces, having expelled the Germans from Russian soil, reached the suburbs of Warsaw. The final year of the war brought a mighty convergence of Allied armies upon Germany—U.S.S.R. from the east, Anglo-U.S. from the south and west. The Soviet army occupied Rumania in Aug. 1944 and Bulgaria (although that country had remained neutral) in September. Greece lapsed into civil war between Communists and anti-Communists. As a result of the Yalta conference (Feb. 1945) of Churchill, President Roosevelt and Stalin, Poland fell virtually under Soviet control, yielding eastern territories to the U.S.S.R. but receiving Silesia, Pomerania and the bulk of East Prussia from Germany. In Yugoslavia Tito, the Communist leader of resistance against Germany, was acknowledged ruler of the country. Stalin installed a Communist government in Hungary soon after the Yalta meeting. Soviet power had replaced German power in eastern Europe even before the war ended.

It was May 2, 1945, before fighting stopped in Italy, May 7 before the German high command surrendered unconditionally. Hitler had committed suicide when the Russians reached the heart of Berlin. The extensive campaigns against Japan in the far east and the Pacific, which had raged throughout 1944 and 1945, continued. Two days after the first U.S. atomic bomb had been dropped on Hiroshima (Aug. 6) the U.S.S.R. declared war on Japan and invaded Manchuria. Japan surrendered on Aug. 14. The settlement had to be as global in character as war itself had been. To achieve such an aim a new international organization, the United Nations, was prepared at the conference of San Francisco (June 1945). Its charter was signed by representatives of 50 nation-states. Special agencies, such as the UN Relief and Rehabilitation administration (UNRRA) of 1943, already existed to deal with immediate relief work in a war-torn world. (*See* also UNITED NATIONS.)

4. Communism and Europe.—The wartime agreements of the Allies envisaged a lengthy postwar period of concerted action for a wide variety of tasks: during the "temporary period of instability in liberated Europe" (Yalta, Feb. 1945); for the joint administration of occupied Germany (Potsdam, Aug. 1945); for conclusion of treaties of peace with other former enemy states; and, within the United Nations and its several agencies, for continued co-operation of many kinds. This policy failed after two years, but during those years it set the political pattern of the postwar world.

It settled, first, the territorial pattern of state frontiers. During 1946 the major powers prepared peace treaties with Finland, Hungary, Rumania, Bulgaria and Italy; these were signed in Feb. 1947 in Paris. The provisional settlement with Germany was dictated at Potsdam. The general outcome was that the U.S.S.R. acquired direct control over the whole belt of the eastern marshlands comprising Petsamo and eastern Karelia; the three Baltic states of Estonia, Latvia and Lithuania; the northern part of East Prussia; the eastern part of Poland; the Carpatho-Ukrainian area of Czechoslovakia, Bessarabia and northern Bukovina. Apart from adjustments to the Polish and Czechoslovak and the Italian and Yugoslav frontiers, prewar frontiers were broadly restored. In western Europe the Saar was at first (1947) established as an autonomous territory in economic union with France, but in 1957 it was reunited with western Germany. Germany, at first divided into four separate zones of occupation by Soviet, U.S., British and French forces, was in 1949 constituted as two distinct states, the German Federal Republic comprising the three western zones, and the German Democratic Republic comprising the eastern zone. Berlin, itself divided at first into four zones and jointly administered by the four powers, was similarly split into two zones—west and east—by 1949.

The policy determined secondly, the broad pattern of immediate postwar regimes in Europe. In most liberated or conquered countries governments were coalitions of Communists, Socialists and Christian Democrats or Peasant parties. In most western countries (including France) and some eastern countries (Hungary, Czechoslovakia, Poland) such governments were set up by more or less free elections. They usually carried through important reforms, redistributing the land or nationalizing industries and instituting provisions for social welfare. In these years the welfare state became the prevalent pattern in Europe. In Great Britain active Labour governments (1945–51) achieved comparable reorganization.

But after 1947 the contrasts between east and west became sharper in domestic as well as in international affairs. In eastern lands, under Soviet control or influence, non-Communists were removed from power and the whole area east of the Elbe and the Adriatic, with the exception of Greece and Turkey, fell under the control of Communist parties. In Oct. 1947 the Third International, dissolved in 1943 as a wartime gesture, was revived as the Communist Information bureau (Cominform), to co-ordinate the activities of its member parties. In the west the contrary process occurred. Communist parties were excluded from a share in power in France, Belgium and Italy in the spring of 1947, and in the Federal German Republic in 1949. They had never been strong in the Scandinavian countries and were suppressed in Portugal and Spain. Internal economic policies changed correspondingly in the two halves of the continent: increasingly restorative and moderate in the west, increasingly revolutionary in the east. In Germany economic divergence culminated in the introduction of separate currencies in June 1948. By 1949 the era of "cold war" between east and west had set in, and the "iron curtain" from Stettin to Trieste divided Europe into two worlds.

The schism was symbolized by the creation, in April 1949, of the North Atlantic Treaty organization (NATO), a defensive alliance based on the western powers and providing for common strategy and joint military forces under a unified command. From the first it included the United States, Canada, the United Kingdom, France, Belgium, Luxembourg, Denmark, Norway, Iceland, Italy, the Netherlands and Portugal; later it was joined by Greece and Turkey (1952) and the German Federal Republic (1955). The Warsaw Treaty organization (May 1955) created its formal counterpart for eastern Europe, comprising the Soviet Union, Poland, Czechoslovakia, Hungary, Rumania, Bulgaria, Albania and the German Democratic Republic. The reality of concerted defense under Soviet army control existed long before. The settlement was one of partition and armed deadlock.

The economy of Europe, so gravely weakened by the war, suffered a process of partition parallel to political partition. The

POLITICAL BOUNDARIES, CAPITALS, MAJOR

MOUNTAINS AND RIVERS OF EUROPE

U.S. offer of economic aid to Europe (Marshall plan; *q.v.*) of 1947 led to the European Recovery program (E.R.P.) and to the Organization for European Economic Co-operation (O.E.E.C.), later supplemented by the European Payments union (E.P.U.). O.E.E.C. comprised not only the western European states of NATO but also Austria (with which a peace treaty, restoring its independence, was made only in 1955), the Republic of Ireland and neutral Sweden and Switzerland. The states of the Soviet orbit formed a closed economic bloc in 1949, its economic planning being directed by the Council for Mutual Economic Aid. After the Communist revolution in China at the end of 1949 the U.S.S.R. also made a treaty of friendship, alliance and mutual assistance (Feb. 1950) with the new regime there.

The partition of Europe, thus consolidated by mid-century, brought about new relationships between Europe as a whole and the continents of Asia and Africa. The colonial revolution, dramatically advanced by the war, had already gone far. Former colonial territories asserted their independence and joined the United Nations as separate states until, by 1962, members reached 104 in number.

From 1950 onward the colonial revolution for independence was overshadowed by (and at times overlapped) the cold war between east and west. There was open warfare in Korea (1950), in which UN forces (mainly U.S.) restored the 38th parallel as the frontier between North and South Korea (*see* KOREAN WAR). A similar partition was imposed in Vietnam. When all French African colonies (except Algeria) were granted or promised independence within the new French Community in 1958, and the Belgian administration withdrew abruptly from the Congo in 1960, the Afro-Asian members of the United Nations constituted a bloc large enough for both sides in the European cold war to seek their support ardently. The diminished status of Europe in the world was, indeed, emphasized by the growing importance and solidarity, in world affairs, of the new Afro-Asian states. At the conference of Afro-Asian countries at Bandung (April 1955) it was Communist China, not the U.S.S.R., that took the leading role; the conference included no European representatives at all (*see* BANDUNG CONFERENCE). The entry of the ever-growing number of Afro-Asian states into the United Nations expressed even more forcibly the changing balance of world power.

Awareness both of the threat of Communism and of this changing balance of world forces fostered, during the 1950s, movements for closer European union. The Council of Europe dated from 1949, the European Coal and Steel Community (*q.v.*) from 1951. Western European union (W.E.U.) was set up in 1955, organizations for a common market (*see* ECONOMIC UNION; *The European Economic Community*) and for joint development of atomic energy (EURATOM) followed in 1958. In economic, military and social co-operation the peoples of western Europe equipped themselves with abundant machinery and facilities. By the early 1960s their economic recovery seemed complete, and European productivity surpassed all former levels. The postwar years brought a rise in birth rate and a fall in death rate in most European countries, so that population increased even in countries, such as France, which had previously faced a prospect of decline in population. Social services and welfare provisions made for a higher standard of living and of social security. Despite lavish expenditure on armaments and national defense, the peoples of western Europe experienced unprecedented material prosperity. But over them loomed the possibility of nuclear war and equally unprecedented destruction, and the knowledge that, while scientists conquered outer space, peace on earth hung by a thread. (D. TN.)

VI. POPULATION

1. Component States.—Europe consists of 27 states (28 if Germany is regarded as two states), together with a number of small though interesting units with special status—Andorra, Gibraltar, Liechtenstein, Monaco, San Marino and Vatican City. The pattern of the frontiers has changed constantly through the centuries, usually after each period of warfare. The broad division of the world between Communist and non-Communist countries is reflected in the division between eastern and western Europe, known as the "iron curtain." The Communist countries form a belt across Europe from the Baltic to the Black sea. They form together with the U.S.S.R., a military alliance—the Warsaw Treaty organization (*q.v.*). From 1949 their economic life was directed by collective planning supervised by the Council for Mutual Economic Aid. After 1945 the countries of western Europe became grouped in various ways. The military grouping was culminated in the North Atlantic Treaty organization (*q.v.*). The economic groupings include the Organization for Economic Co-operation and Development; the Benelux (*q.v.*) union; the European Coal and Steel Community (*q.v.*); and the European Economic Community or "common market." Some progress has been made toward a "European parliament" by the establishment of the Council of Europe. Several countries are neutral in their political and military relations, notably Switzerland, Sweden, the Irish Republic, Austria and Finland. Spain and Yugoslavia lie outside the two main power blocs, but Spain has a military treaty with the United States. (*See* also PAN-EUROPEAN MOVEMENT.)

Population of European Countries

Country	Area (in sq.mi.)	Population*	Density per sq.mi.	Capital	Population*
Albania	11,100	1,625,378	146.4	Tirane	136,400
Andorra	175	8,392	48.0	Andorra la Vella	2,012†
Austria	32,374	7,020,626	216.9	Vienna	1,619,668
Belgium	11,779	9,178,154	779.2	Brussels	170,054
Bulgaria	42,831	7,905,500	184.6	Sofia	766,400
Czechoslovakia	49,366	13,741,529	278.4	Prague	998,193
Denmark	16,619	4,585,256	275.9	Copenhagen	721,381
Faeroe Islands	540	34,000	63.0	Thorshavn	6,067‡
Finland	130,119	4,448,575	34.2	Helsinki	452,800
France	212,821	45,500,000	213.8	Paris	2,850,189§
Germany, Democratic Rep. (East)	41,635	17,188,488	412.8	Berlin	1,071,775
Germany, Federal Rep. (West)	95,920	56,163,400	585.5	Bonn	146,889
Gibraltar	2.25	24,502	10,889.8	—	—
Greece	50,547	8,357,526	165.3	Athens	621,741
Hungary	35,919	9,976,530	277.8	Budapest	1,807,299
Iceland	39,809	177,292	4.5	Reykjavik	72,407
Ireland	27,136	2,814,703	103.7	Dublin	535,488
Italy	116,300	50,463,762	433.9	Rome	2,160,773
Liechtenstein	61	16,628	272.6	Vaduz	3,398
Luxembourg	999	314,889	315.2	Luxembourg	71,653
Malta	122	328,785	2,695.0	Valetta	18,433
Monaco	0.6	22,297	37,161.7	Monaco-Ville	1,774
Netherlands	12,616	11,556,008	916.0	Amsterdam‖	866,342
Norway¶	149,166	3,574,210	24.0	Oslo	471,511
Poland	120,359	29,731,000	247.0	Warsaw	1,136,000
Portugal	35,340	9,130,410	258.4	Lisbon	818,382
Rumania	91,699	18,403,414	200.7	Bucharest	1,225,507
San Marino	24	16,700	695.8	San Marino	2,908?
Spain	194,880	30,128,056	155.2	Madrid	1,966,070
Sweden	173,665	7,495,129	43.2	Stockholm	808,484
Switzerland	15,941	5,429,061	340.6	Bern	163,172
Turkey in Europe	9,121	2,293,582	251.5	Ankara§	646,151
United Kingdom[2]	94,211	52,834,229	560.8	London	3,195,114
U.S.S.R. in Europe	2,079,563	158,721,418†	76.3	Moscow	5,045,905†
Vatican City	0.17	1,000	5,882.4	—	—
Yugoslavia	98,766	18,512,805	187.4	Belgrade	587,899
Total	3,991,529	587,723,234	147.2	—	—

*Census or estimate, 1960 or 1961. †1959 est. ‡1955 census. §1954 census. ‖Official capital. De facto capital is The Hague. ¶Includes Svalbard and Jan Mayen. ?1957 est. §Not a part of Europe. [2]Includes Channel Islands and Isle of Man.

2. Distribution of Population.—Europe (excluding the European portion of the U.S.S.R.) has an area of about 1,911,000 sq.mi., and an estimated population of slightly more than 429,000,000 (1961) or about 15% of the world total. It is the most densely populated continent, with an average density of 225 per square mile as compared with 56 for the world as a whole. The rate of annual increase in the 1950s was, however, less than that of any other continent, being only 0.7% compared with 1.6% for the world as a whole. The European portion of the U.S.S.R., with an area of more than 2,000,000 sq.mi.—slightly exceeding that of the rest of the continent—contains about 159,000,000 persons, with an average density of 76 per square mile. These average figures conceal wide variations ranging from the virtually uninhabited areas of the higher parts of the uplands and mountain ranges to the large urban and industrial agglomerations of western Europe. (*See* accompanying Table.)

Norway, consisting of a high *fjeld* plateau falling steeply to a deeply indented and island-fringed coast, with a moist cloudy climate and only 4% of its area under farmland, has the lowest average density of any European country. More than half of the people live in the southeastern lowlands around Oslo fiord and in the Glomma valley. Finland and Sweden also have low densities,

the result of the high proportions of tundra and waste, lakes and forests. In each case the bulk of the people live in the lowlands in the south, where the climate is more amenable and the soils better. More than half the Finns, for example, live in a narrow strip along the south coast, where the capital, Helsinki, is located. Almost three-quarters of Iceland is uninhabited and population density is as low as 4.5 per square mile, mainly concentrated on the western and southern coasts.

While parts of western and northern Denmark consist of thinly populated sandy heathland, the fertile soils of eastern Jutland and the islands lead to an intensive agriculture based on dairying. A quarter of the people work on the land and another quarter obtain livelihood through the commerce and industry associated with agriculture. More than a fifth of the people live in the capital city of Copenhagen.

The three Benelux countries—Belgium, Netherlands, Luxembourg—situated in a vital part of western Europe, with a highly developed agricultural, industrial and commercial life, have an importance out of proportion to their sizes; the Netherlands and Belgium rank high in order of average density. More than 70% of the Netherlands is under cultivation, a proportion achieved only by centuries of effort, yet urban life is highly developed: 14 towns have more than 100,000 persons and another 20 have more than 50,000. The province of South Holland has a density exceeding 2,500 per square mile, one of the most densely populated areas in the world. In Belgium, though the over-all average is high, contrasts range from the Plain of Flanders, the low plateau of Brabant around Brussels, and the industrial region of the Sambre-Meuse valley containing numerous large towns with a prosperous industrial and commercial life going back to the middle ages, to the forested uplands of the Ardennes and the extensive Campine heathlands. In Luxembourg only 15% of the people live in the northern third (the Ardennes), the rest in the Bon Pays, with its more fertile soils and flourishing steel industry.

The density of population in France is below the European average (excluding European U.S.S.R.), although there are several major industrial regions, and 24 towns have populations of more than 100,000, the result of many centuries of growth; the population of metropolitan Paris is about 7,600,000. This low density is partly because several scantily populated regions occupy a large area (much of the Massif Central, central Brittany, the Alps, the Pyrenees and the Landes). In addition, over 40% of the inhabitants are classified as rural dwellers; more than one-third of the surface is cultivated. The heavy war losses and the low birth rate for many years have in fact made France short of manpower, and there are about 1,500,000 foreigners. Spain, too, has an uneven distribution of population, the result of the large extent of plateau and high sierra, much of a semiarid character, with a small area of well-watered but limited basins, valleys and coastal lowlands; 30% of the country is cultivated, though only a quarter of this can be said to be fertile. Yet 25 cities contain more than 100,000 persons. Parts of Portugal are well populated, as around Lisbon (the capital) and in the north where the rural population lives on intensively cultivated small holdings, but in the eastern uplands and the southern plains (where the land is still held in large estates), the density is low. In many respects the country is not well developed, especially industrially.

Italy has a population of more than 50,500,000—about a third lives in the northern plain. This is partly accounted for by a prosperous agriculture, partly by the presence of numerous industrial towns, notably Milan (1,400,000 [commune]), Turin (890,000) and Venice (337,000). The mountain backbone of the peninsula, the Apennines, is thinly populated, the result of bare limestone slopes, cold bleak winters and parched and arid summers. Population is concentrated on the small areas of coastal plain, especially around Rome (2,150,000), Naples (1,120,000), Florence (412,000), Leghorn (155,000) and others. In Sicily a dense population lives along the north and east coasts, with the large towns of Palermo (570,000), Catania (345,000) and Messina (242,000); in spite of the extensive plateau occupying the centre of the island, the density per square mile is over 480. By contrast, that of Sardinia, an island of rugged granite plateaus with some marshy coastal plains is only about 150 per square mile.

Much of Yugoslavia consists of limestone plateaus and ranges, though the capital city, Belgrade (588,000), lies in the most densely populated area in the Danube plains to the northeast. Though a large part of the country is upland with poor soils, nearly three-quarters of the working population is engaged in agriculture. Albania, much of which consists of high ranges and remote basins and valleys, is still one of the most undeveloped countries in Europe, despite the post-World War II influence of the Soviet Union. In Greece at least half the population consists of peasant farmers; less than 10% is engaged in industry; and only three towns, Athens, Piraeus and Salonika, have a population of more than 100,000. Yet the country is mountainous and not very fertile, standards of living are low and poverty is rife. Similarly, Bulgaria is still a poor land, dependent on agriculture (with 75% of the people so engaged), though the extensive Balkan and Rhodope mountains limit the farmland to one-third of the total area, mainly in the valleys and basins. Of the total population of 2,293,000 of Turkey-in-Europe, 1,460,000 live in Istanbul. The Balkans as a whole stand out as a thinly populated region, with few large towns other than the capital cities.

Austria and Switzerland may be described as the Alpine lands, but are very different in many ways. Austria is more than twice as large, but its population is only 1,600,000 greater, and it includes 1,620,000 in its capital city Vienna, once the centre of a great empire. About one-third of the population lives in five large towns, the remainder in small towns and villages in the valleys and basins between mountain ranges. Two-thirds of Switzerland lies within the Alpine system and one-tenth in the Jura, the remaining Swiss plateau with both intensive agriculture and large industrial towns contains two-thirds of the population. Zürich, the largest city, has 440,000 inhabitants; Basel 207,000; Geneva 176,000; and Bern (the federal capital) 163,000.

Hungary, one of the two most densely populated eastern European countries, has an average density of over 275 per sq.mi., the result of close agricultural settlement, with many large villages and small towns distributed evenly about the lowlands. Budapest contains nearly one-fifth of the population (1,810,000), and other towns, such as Miskolc (143,000), Pécs (115,000), Debrecen (130,000) and Szeged (99,000) grew steadily after 1945 with industrialization. Rumania is about the size of the United Kingdom, though its population is only about a third. This is unevenly spread, since the Carpathians and the higher parts of the Transylvanian basin are thinly populated, while the Walachian lowlands and the southwestern plains have a fairly dense rural population. About three-quarters of the workers are engaged in agriculture, but some degree of industrialization has helped the towns—12 have populations exceeding 100,000 compared with 4 before 1939.

The flat agricultural plain of Poland, with an average density of about 260 per square mile was the scene of the greatest population movements during and after World War II, chiefly the result of boundary changes and persecution. The population is increasingly urban and by the early 1960s there were more than 20 towns with populations exceeding 100,000. Czechoslovakia, similarly, with a large urban and industrial population, has a density of over 275 per square mile. The chief towns are Prague (1,000,000), Brno (323,000), Bratislava (258,000), Ostrava (264,000) and Plzen (142,000).

Germany in 1939 had a population of about 80,000,000 and was one of the world's four great industrial countries. After World War II it was divided into two parts. Germany as a whole has an estimated population of 73,500,000 persons. The distribution of population shows a remarkable contrast between the thinly populated northern plains, central uplands and Alpine fringe on the one hand, and the densely populated industrial districts (notably the Ruhr and Saxony) of the Rhineland, the more fertile valleys and basins and the rich Börde. *See* also *Population* sections of articles on the individual countries.

See also references under "Europe" in the Index volume.

(F. J. M.)

BIBLIOGRAPHY.—*Physical Geography*: V. de la Blache and L. Gallois, *Géographie Universelle* (1926–48); S. van Valkenburg and E. Hunting-

ton, *Europe* (1936); M. R. Shackleton, *Europe*, 6th ed. (1958); F. J. Monkhouse, *A Regional Geography of Western Europe* (1959); L. D. Stamp, *A Regional Geography for Advanced Scholarship Courses*, pt. V, *Europe and the Mediterranean*, 9th ed. (1960); A. Sömme, *The Geography of Norden* (1960); A. C. O'Dell, *The Scandinavian World* (1957); A. F. A. Mutton, *Central Europe* (1961); D. S. Walker, *Mediterranean Lands* (1960); T. Shabad, *Geography of the U.S.S.R.* (1951); W. G. Kendrew, *Climates of the Continents*, 4th ed. (1953); C. F. Nyman, *Conspectus Florae Europaeae* (1878–82); *Flora Europaea* (in English) in several volumes (1962–); S. Taylor (ed.), *Handbook of Central Europe* (1937); W. Bieger and A. Wahlström, *Die wildebenden Säugetiere Mitteleuropas* (1938); R. Hainard, *Les Mammifères sauvages d'Europe*, 2 vol. (1949); J. Aubert, *Papillons d'Europe* (1949); R. Peterson, G. Mountfort and P. A. D. Hollom, *Field Guide to the Birds of Britain and Europe*, 3rd rev. ed. (1958); K. H. Voous, *Atlas of European Birds* (1961); A. C. Hardy, *The Open Sea*, vol. ii, "Fish and Fisheries" (1959).

Natural Resources: G. W. Robinson, *Soils, Their Origin, Constitution and Classification*, 3rd ed. (1951); W. C. Kubiëna, *Soils of Europe* (1953); N. Pounds and W. N. Parker, *Coal and Steel in Western Europe* (1957); N. Pounds, *The Ruhr* (1953); W. R. Mead, *A Regional Geography of the Scandinavian States and Finland* (1959); G. East, *An Historical Geography of Europe*, 4th ed. (1955); I. Frank, *European Common Market* (1961); L. Lister, *Europe's Coal and Steel Community* (1961); Political and Economic Planning, *Tariffs and Trade in Western Europe* (1959), *European Organizations* (1959).

Anthropology: W. Z. Ripley, *Races of Europe* (1899); H. J. Fleure *et al.*, articles in *J. R. Anthrop. Inst.* (1916, 1920, 1936, 1958); V. G. Childe, *Dawn of European Civilization*, 6th ed. (1957); C. S. Coon, *Races of Europe* (1939); L. Dominian, *Frontiers of Language and Nationality in Europe* (1917); W. Koppers (ed.), *Die Indogermanen-Und Germanenfrage* (1936); A. E. Mourant, A. C. Kopec and K. D. Sobczak, *The A.B.O. Blood Groups* (1958); C. M. Arensberg and S. T. Kimball, *Family and Community in Ireland* (1940); E. E. Evans, *Irish Folkways* (1957); I. F. Grant, *Highland Folkways* (1961); J. A. Pitt-Rivers, *The People of the Sierra* (1954); A. D. Rees, *Life in a Welsh Countryside* (1950). The following periodicals may be consulted for useful articles: *Folk-Liv* (Stockholm); *Zeitschrift für Ethnologie* (Berlin); *Volkskunde* (Holland); *Scottish Studies*; *Gwerin*.

Archaeology: V. G. Childe, *The Danube in Prehistory* (1929), *The Dawn of European Civilization*, 6th ed. (1957), *Prehistoric Migrations in Europe* (1950); C. F. C. Hawkes, *Prehistoric Foundations of Europe* (1940); R. Pittioni, *Die Urgeschichtliche Grundlagen der Europäischen Kultur* (1949); *Vsemirnaya istoriya*, vol. i (1955); A. L. Mongait, *Arkheologiya v S.S.S.R.* (1955; Eng. trans. by M. W. Thompson, *Archaeology in the U.S.S.R.*, 1961); *Naryasy starodavniovi istorii Ukrainskoi R.S.R.* (1957); P. I. Boriskovski, Paleolit Ukrainy, *"Materialy i issledovaniya po arkheologi," MIA*, no. 40 (1953); A. Ya Bryusov, *Ocherki po istorii plemen Evropeiskoi chasti S.S.S.R. v Neolitcheskuyu epokhu* (1952); T. S. Passek, "Periodizatsiya Tripolskikh poseleny," *MIA*, no. 10 (1949); M. I. Artamonov and O. A. Grakova, "Bronzovy Vek," *Bolshaya Sovietskaya Entsiklopediya*, vol. 6 (1951); B. I. Grakov, "Kamenskoe gorodishche na Dniepre," *MIA*, no. 36 (1954); A. V. Zbrueva, "Istoriya naseleniya Prikamya v Ananinskuyu epokhu," *MIA*, no. 30 (1950); A. Iessen, *Grecheskaya kolonizatsiya severnogo Prichernomorya* (1947); P. N. Tretiakov, *Vostochnoslavyanskie plemena* (1953); M. Gimbutas, *Prehistory of Eastern Europe*, pt. I (1956); T. Talbot Rice, *The Scythians* (1957); M. I. Rostovtsev, *Iranians and Greeks in S. Russia* (1922).

History: Max Beloff, *The Foreign Policy of Soviet Russia, 1929–1941*, 2 vol. (1947–49); E. H. Carr, *Britain: A Study in Foreign Affairs From the Treaty of Versailles to the Outbreak of War* (1940); S. B. Fay, *The Origins of the World War*, 2 vol., 2nd ed. (1930); Edmund Glaise-Horstenau, *The Collapse of the Austro-Hungarian Empire* (1930); H. N. Howard, *The Partition of Turkey* (1931); C. L. R. James, *World Revolution 1917–1936: The Rise and Fall of the Communist International* (1937); H. Johnson, *Vatican Diplomacy in the World War* (1933); J. I. Knudson, *A History of the League of Nations* (1938); R. H. Lutz (ed.), *Fall of the German Empire 1914–1918* (1932); M. H. H. Macartney and P. Cremona, *Italy's Foreign and Colonial Policy, 1914–1937* (1938); W. N. Medlicott, *British Foreign Policy Since Versailles* (1940); L. Robbins, *The Great Depression* (1934); B. E. Schmitt, *The Coming of the War, 1914*, 2 vol. (1930), *From Versailles to Munich 1918–1938* (1939); K. Smogorzewski, *Poland's Access to the Sea* (1934); R. Strausz-Hupé and S. T. Possony, *International Relations in the Age of the Conflict Between Democracy and Dictatorship* (1950); T. A. Taracouzio, *War and Peace in Soviet Diplomacy* (1940); Hans Kohn, *The Twentieth Century* (1949); J. W. Wheeler-Bennett, *The Forgotten Peace: Brest-Litovsk* (1938), *Munich: Prologue to Tragedy* (1948); G. East, *Historical Geography of Europe*, 3rd ed. (1949); J. Bowle, *Unity of European History* (1950); D. G. Ayerst, *Europe in the Nineteenth Century*, 2nd ed. (1952); A. L. C. Bullock and A. J. P. Taylor (eds.), *Select List of Books on European History 1815–1914*, 2nd ed. (1957); R. H. C. Davis, *History of Medieval Europe* (1957); C. H. J. Hayes, *Contemporary Europe Since 1870* (1958); R. M. Rayner, *Concise History of Modern Europe 1789–1914* (1958); H. L. Peacock, *History of Modern Europe 1789–1939* (1958); H. A. L. Fisher, *History of Europe* (1960); H. W. C. Davis, *Medieval Europe*, 2nd ed. (1960); P. J. Helm, *History of Europe 1450–1660* (1961); *The Europa Year Book*, vol. i (annually); K. Linday, *Towards a European Parliament* (1958).

Areas and populations of European countries are summarized annually in *Britannica Book of the Year*.

EUROPEAN COAL AND STEEL COMMUNITY. The E.C.S.C. was established in 1952 by the governments of France, the German Federal Republic, Italy, Belgium, the Netherlands and Luxembourg as an administrative agency designed to pool the coal and steel resources of those nations. It subsequently became the nucleus of the European Economic Community (E.E.C.), established by the same six powers in 1957 to work for a common European market, and the European Atomic Energy Community (EURATOM), established at the same time to pool the resources of the six nations for the development of peaceful uses of atomic energy. *See* PAN-EUROPEAN MOVEMENT; FEDERAL GOVERNMENT.

EUROPEAN COMMON MARKET: *see* ECONOMIC UNION.

EUROPEAN ECONOMIC COMMUNITY (EUROPEAN COMMON MARKET): *see* ECONOMIC UNION; PAN-EUROPEAN MOVEMENT.

EUROPEAN MOVEMENT: *see* PAN-EUROPEAN MOVEMENT.

EUROPEAN PAYMENTS UNION. The E.P.U. was set up in 1950 by the Organization for European Economic Cooperation (O.E.E.C.), the co-ordinating agency for the European Recovery Program, for the multilateral clearing of payments among its members, the United States contributing a working capital fund of $350,000,000. Each member country was allotted a quota based on its transactions with the other members. Each opened with, and received from, the E.P.U. lines of credit equal to 60% of its quota. Members included Austria, Belgium-Luxembourg, Denmark, France, the German Federal Republic, Greece, Iceland, Italy, the Netherlands, Norway, Portugal, Sweden, Turkey and the sterling area countries.

Each central bank reported monthly to the Bank for International Settlements (*q.v.*) at Basel, Switz., its claims on, and liabilities to, the other members. Settlement was made in units equivalent to U.S. dollars through the E.P.U. on an agreed formula which from July 1955 was 75% gold (or currency acceptable to the E.P.U.) and 25% credit with the E.P.U. Clearing was multilateral, eliminating bilateral settlements between countries. Inclusion of the sterling area widened its geographical scope.

In 1950 members of the European Payments union agreed to eliminate quotas on private trade equal to 50% of their 1948 transactions with other members. The percentage of elimination later considerably exceeded the amount, but this did not apply to official transactions, and tariffs sometimes replaced quotas. The E.P.U. ended with the inauguration of the European Common Market in Jan. 1959. *See* also INTERNATIONAL PAYMENTS.

See Stephen Enke and Virgil Salera, *International Economics*, 3rd ed., pp. 195–219, with bibliography (1957); Managing Board, European Payments Union, *Annual Reports*. (J. B. CE.)

EUROPEAN RECOVERY PROGRAM (E.R.P.), a program for European economic recovery worked out by 17 countries of western Europe, in response to a suggestion made by U.S. Secretary of State George C. Marshall, at Harvard university on June 5, 1947. Also known as the Marshall plan, E.R.P. was an attempt to replace by co-operative efforts, and with large-scale coordinated assistance from the United States, the earlier individual attempts to promote economic recovery from the consequences of World War II. The program was run in Paris by the Organization for European Economic Cooperation (O.E.E.C.). Economic assistance from the United States was administered by the Economic Cooperation administration (E.C.A.). *See* FOREIGN AID PROGRAMS. (C. P. K.)

EUROPIUM is a metallic element of the rare-earth group. It is the least dense, the softest and most volatile member of the group and is also one of the least abundant members. The primary use of europium has been for research purposes. Because of its high cross section for thermal neutrons, it may prove to be of use in the construction of such materials as nuclear reactor control rods. It has been used as an agent in the manufacture of

fluorescent glass, as a phosphor activator and as a component of certain electronic materials. The symbol of europium is Eu.; its atomic number is 63; and its atomic weight is 151.96. The naturally occurring stable isotopes are: Eu^{151} (47.77%) and Eu^{153} (52.23%).

The element was discovered by E. Demarçay in 1896 and named for Europe; he isolated a pure fraction in 1901. It occurs in minute amounts in many rare-earth minerals such as monazite, samarskite and also in the products of atomic fission. The element is usually purified by means of its easily attainable oxidation state of 2^+; europium possesses the most easily produced 2^+ oxidation state of the rare earths. Trivalent europium solutions can be reduced by zinc metal and hydrochloric acid to give Eu^{2+} in solution; the ion is stable in dilute hydrochloric acid if oxygen from the air is excluded. A series of white to pale yellow or green divalent europous salts are known, such as $EuSO_4$, $EuCl_2$, $EuBr_2$, $Eu(OH)_2$ and $EuCO_3$. The divalent halides may also be prepared by hydrogen reduction of the anhydrous trivalent halides at about 500° C., by alkali metal amalgams or by electrolysis with a mercury pool cathode.

In its trivalent state, europium behaves as a typical rare earth, forming a series of white salts and colourless solutions. Its ionic radius is 0.950. Solutions show discrete absorption peaks in the ultraviolet and visible regions of the spectrum, which may be used for quantitative analysis. The Eu^{3+} ion is paramagnetic because of the presence of unpaired electrons.

The metal has been prepared by electrolysis of the fused halides or by thermoreduction of the halides with alkali or alkaline earth metals. It has been prepared by reduction of its oxide, Eu_2O_3, by lanthanum metal followed by distillation of the europium metal. It is chemically reactive and is attacked by oxygen or water. Its melting point is 826° C.; its density for the body-centred cubic structure is 5.245 g. per cubic centimetre. *See* RARE EARTHS.

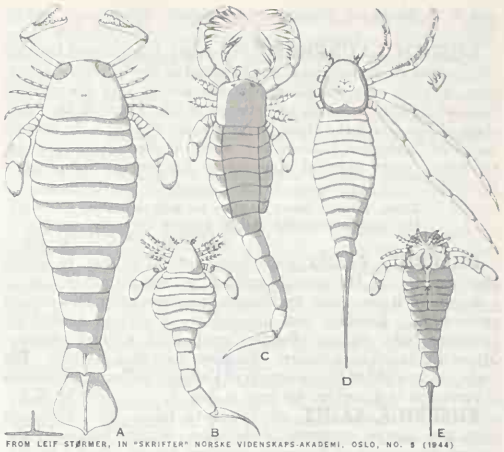
(LD. B. A.)

EURYDICE, in Greek mythology, the wife of Orpheus (*q.v.*).

EURYPTERIDA, an evolutionally important subclass of extinct aquatic arachnoids belonging to the Paleozoic Arthropoda (*q.v.*). Their oldest representatives are found in Lower Ordovician rocks, some 400,000,000 years old. There has been much speculation concerning their relationships, evolution and mode of life, but it is generally accepted that the Eurypterida, together with the living Xiphosura (*q.v.;* horseshoe crabs), constitute the class Merostomata, only distantly related to the trilobites and arachnids (spiders, mites, etc.; *see* TRILOBITA; ARACHNIDA). Probable ancestors of the eurypterids are found among Cambrian aglaspias, some of which are structurally intermediate between xiphosurans and eurypterids. Although eurypterids were not inhabitants of the open sea, the enclosing rocks and associated fossil remains suggest a habitat of brackish and fresh waters. They reached the height of their diversity and abundance in the Upper Silurian, showed an abrupt decline in the Devonian and finally died out in the Permian, about 230,000,000 years ago.

The Eurypterida form a small subclass, of which only about 10 families and 200 species are known. Giants among arthropods, the eurypterids ranged from about one-half foot to more than seven feet in length. Their flattened body with variously modified appendages suggests a bottom-dwelling existence. They probably crawled upon and burrowed into the surface and occasionally swam. They were probably carnivores, the largest of them being perhaps predators.

The body of eurypterids, as reconstructed from fossil remains, is scorpionlike and clearly divided into two regions, the unsegmented anterior cephalothorax, or prosoma, and a many-jointed abdomen, or opisthosoma. The prosoma is covered with a carapace (a dorsal horny shield) bearing various structures useful in classification, such as sutures, tubercles, etc.; a pair of median ocelli; and a pair of larger, lateral, compound eyes. The mouth is situated on the ventral surface behind the anterior edge, and is bordered posteriorly by a small plate, which is almost completely covered by a much larger plate situated between the last pair of coxae (joints articulating with the body) and variously interpreted either as the first abdominal sternite or as the fused remnant of a first pair of abdominal appendages. The abdomen in the adult is composed of 12 segments and a postanal spine, or telson. The shape of the opisthosoma varies in different groups. The telson may be dorsoventrally flattened and provided with a median keel, as in *Pterygotus;* or assume a spikelike shape, as in *Eurypterus* and *Stylonurus,* similar to that in horseshoe crabs; or a curved spine, as in *Carcinosoma,* adding to the outward resemblance of such eurypterids to scorpions. Although 12 dorsal plates (tergites) are present on the surface of the opisthosoma, only 11 ventral plates are present because the first one became incorporated in the prosoma and appears as a separate plate, the metastoma, enclosed between the hind coxae. The five sternites following upon the metastoma bear a pair of book gills each. The first sternite, corresponding to the second tergite and representing the second abdominal segment, in both sexes bears also the genital operculum with a median genital appendage and the opening of the genital ducts.



FROM LEIF STØRMER, IN "SKRIFTER" NORSKE VIDENSKAPS-AKADEMI, OSLO, NO. 5 (1944)

REPRESENTATIVE SPECIES OF EURYPTERIDA

(A) *Pterygotus rhenaniae,* Lower Devonian, length 1–1.8 m.; (B) *Carcinosoma scorpionis,* Upper Silurian, length 0.4–1.0 m.; (C) *Mixopterus kiaeri,* Uppermost Silurian, length 0.67–1.0 m.; (D) *Stylonurus* (*Ctenopterus*) *excelsior,* Middle Devonian, length 1–1.5 m.; (E) *Eurypterus fischeri,* Upper Silurian, 0.24–1.0 m.

All Eurypterida possess six pairs of prosomatic appendages homologous with similar appendages in Xiphosura and Arachnida. The first pair, the chelicerae, present considerable variations in structure, which have been correlated with probable function. In some they are small and inconspicuous, whereas in others they may project beyond the carapace and act as a grasping organ or "hand," formed by the second joint with its immovable finger and the third joint as the movable finger and provided with teeth. The following four pairs of appendages, often similar in appearance, function as walking legs. But the first is sometimes modified as a tactile organ, and the second may be greatly enlarged as a grasping and clasping organ. Not counting the terminal claw, the first leg is six-jointed, the second and third are seven-jointed, the fourth eight-jointed. The coxae of these four pairs of appendages have a mouth appendage (gnathite) used in mastication and sometimes termed epicoxite. The sixth and last pair of prosomatic appendages present the greatest modifications correlated with the eurypterid mode of life. They are variously developed as walking legs. Not counting the terminal claw, they are composed of eight joints. In certain genera the legs are modified for swimming, several of the joints being flattened to form the "oar-blade" or palette of the leg. The opisthosomatic appendages are represented by the five pairs of book gills arising from the posterior surface of sternal plates.

BIBLIOGRAPHY.—J. M. Clarke and R. Ruedemann, *The Eurypterida of New York* (1912); L. Størmer, "Merostomata," in Arthropoda 2, *Treatise on Invertebrate Paleontology,* pp. 4–41 (1955); R. R. Shrock

and W. H. Twenhofel, *Principles of Invertebrate Paleontology*, 2nd ed. (1953). (A. PH.; N. D. N.)

EUSDEN, LAURENCE (1688–1730), English poet laureate satirized by Alexander Pope, was baptized at Spofforth, Yorkshire, Sept. 6, 1688. He was made poet laureate in 1718 through the duke of Newcastle, whom he flattered in his "Poem on the Marriage of His Grace the Duke of Newcastle" (1717). He became rector of Coningsby, Lincolnshire, where he died, Sept. 27, 1730. Of Pope's many satirical allusions to Eusden the following is typical:

> Know, Eusden thirsts no more for sack or praise;
> He sleeps among the dull of ancient days.
> *Dunciad*, bk. i, II, 293–294.

EUSEBIUS, SAINT, pope from April 18 to Aug. 17 of the year 309 or 310. His epitaph, written by Pope Damasus, tells of a dispute in Rome about readmitting apostates after Diocletian's persecution. Eusebius was opposed by a faction headed by Heraclius, who wanted offenders readmitted without penance. Maxentius exiled both leaders. Eusebius soon died in Sicily. His body was brought to the catacomb of Callixtus in Rome. Eusebius is venerated as a martyr; his feast is Aug. 17. (W. M. K.)

EUSEBIUS, SAINT, OF SAMOSATA (d. *c.* 379), Christian saint and martyr, opponent of Arianism, became bishop of the ancient Syrian city of Samosata (modern Samsat, Turk.) in 361, and his subsequent career accurately reflects the major crises threatening the Christian church of the 4th century.

In 361 the Arian emperor Constantius threatened Eusebius with the loss of his right hand if he persisted in refusing to surrender the official record of St. Melitius' election to the see of Antioch. The latter had been supported by the Arian bishops under the mistaken notion that he would prove sympathetic to their cause; when he immediately made plain his orthodoxy, they prevailed upon the emperor to extort the evidence of Melitius' election from Eusebius in order to destroy it. The emperor withdrew his threat when Eusebius calmly offered both his hands for amputation. During the persecution of Valens, Eusebius traveled in disguise through Syria and Palestine, establishing orthodox bishops and priests in the dioceses which had lost them. In 374 he himself suffered banishment to Thrace, where he remained until the death of Valens in 378, when he was welcomed back by his faithful flock. He died *c.* 379 from a head wound inflicted by an Arian woman, but only after extracting an oath from his friends that they would not prosecute her. His feast day is June 21 in the west, June 22 in the east.

BIBLIOGRAPHY.—The principal source for the life of Eusebius is Theodoret, *Ecclesiastical History*. *See* also O. Bardenhewer, *Geschichte der altkirchlichen Literatur*, vol. iv, p. 388 (1924); H. Thurston and D. Attwater (eds.), *Butler's Lives of the Saints*, vol. ii, pp. 607–608 (1956). (T. V. BM.)

EUSEBIUS, SAINT, OF VERCELLI (d. 370 or 371), a supporter of Athanasius (*q.v.*) and a restorer of the Nicene faith, was born in Sardinia. At first a lector in the church of Rome, in 345 he became the first bishop of Vercelli. Living in community with his priests, he was the first western bishop to unite monastic life with the priestly ministry. At the council of Milan (355) which he had helped to summon he refused to sign the condemnation of Athanasius, and was exiled to the east. Allowed by the edict of Julian to return home, he attended the synod of Alexandria (362), whose decrees he thereafter promulgated, thus doing much to restore the orthodox faith and unity throughout the empire. His feast day is Dec. 16.

Eusebius wrote three letters during his exile. The first seven books of the *De Trinitate* long attributed to either Athanasius or Vigilius of Thapsus are now generally accepted as Eusebius' work, though some scholars still contest his authorship. The 4th-century *Codex Vercellensis*, which presents a notable Old Latin text of the Gospels, is probably in Eusebius' handwriting. The Athanasian creed has been ascribed to Eusebius but without solid grounds.

BIBLIOGRAPHY.—Letters and *De trinitate*, ed. by V. Bulhart, *Corpus Christianorum, Series Latina*, vol. ix (1957); also in J. P. Migne, *Patrologia Latina*, vol. x, col. 713–714 (1845), vol. xii, col. 947–954 (1845), and vol. lxii, col. 237–286 (1848). *See* also C. H. Turner and A. E. Burn, "Notes on Eusebius of Vercelli," *Journal of Theological Studies*, vol. i, pp. 126–128, 592–599 (1900); O. Bardenhewer, *Geschichte der altkirchlichen Literatur*, vol. iii, pp. 486–490 (1923); A. Hamman, *Migne Patrologia Latina, Supplementum*, vol. i, col. 305–307 (1958). (V. C. DE C.)

EUSEBIUS OF CAESAREA (*c.* 260–*c.* 340), bishop and scholar whose ecclesiastical history is the chief primary source for the history of the church up to 324. Caesarea in Palestine is the only city with which he is known to have been associated, and there he was trained in the Alexandrian tradition of Christian scholarship as mediated by the learned Pamphilus, to whom he was bound by ties of deep respect and affection, and from whom was derived the title Eusebius Pamphili, by which he was commonly known. In Diocletian's great persecution of the Christians which began in 303, Pamphilus spent a long time in prison and died as a martyr in 310. Eusebius also was imprisoned and was taunted many years later with having escaped by performing some act of submission.

He became bishop of Caesarea about 314. Around 316 he preached at the dedication of a new church at Tyre. By the time the Arian controversy began *c.* 318 (*see* ARIANISM) he was a leading Christian scholar and writer. He did not go so far as Arius in subordinating the Son to the Father, but Arius could regard him, with some justice, as one of his supporters. At the Council of Antioch (late 324 or early 325) Eusebius with two others was provisionally excommunicated for Arian views. His case was held over to a great synod already called to meet at Ancyra, but Constantine I removed this council to Nicaea (*see* COUNCIL). There Eusebius produced his Caesarean creed as evidence for his own correct belief, and he was exonerated from the charge of heresy. He was hesitant about accepting the creed sanctioned by the council, in which the epithet *homoousios* ("of one substance" with the Father) applied to the Son appeared to be Sabellian (*see* CREED). Eventually he signed the creed, though the letter he sent to his church to explain the situation shows that he had difficulty in understanding his own position or the points at issue.

In the years following the council Eusebius took part in the expulsion of Eustathius of Antioch, though he refused translation to the see of that city, of Athanasius of Alexandria and of Marcellus of Ancyra, all opponents of Arianism. In 335 Eusebius delivered an oration at the *tricennalia* (30th anniversary of accession) of the emperor. It was from Eusebius that Constantine commissioned copies of Scripture for the churches of Constantinople. He was well-known as a corrector of biblical texts and surviving manuscripts bear witness to his activity. Nothing is known of his life between 335 and his death.

Eusebius wrote voluminously as apologist, chronographer, historian, exegete and controversialist, but his vast erudition is not matched by clarity of thought or attractiveness of presentation. He is primarily an apologist: even his ecclesiastical history is designed to show, by historical facts, that the church had vindicated her cause against heathen and heretics. It was probably originally completed in 303 (in seven books), and then enlarged (to ten books) in successive editions to relate, together with a separate work on the martyrs of Palestine, the persecutions of his own time and the triumph of Constantine. Eusebius constantly quotes or paraphrases his sources, and thus has preserved portions of earlier works no longer extant. Before writing his history he had compiled his chronicle, an epitome of universal history now extant in an Armenian version. The first part of this chronicle is a survey of the history of various nations; the second consists of the events of history arranged in a chronological table, which Jerome worked into his own Latin chronicle. Eusebius treats the narrative in his ecclesiastical history annalistically, and interrupts it to insert the accessions of the emperors and of the bishops of the four great sees. The result is that his treatment of various subjects tends to be fragmentary; even when he reaches contemporary history (from book viii onward) his narrative is not without confusion and shows that he was not a great historical writer.

After Constantine's death (337) Eusebius wrote a life of the emperor, which, on his own admission, is panegyric, but it is not without historical value and contains original documents. In this and in the anniversary oration, Eusebius developed a theory of

the Christian empire to suit the change of circumstance inherent in Constantine's policy. The theory is largely based on the views of political writers of the Hellenistic age.

Eusebius wrote several specifically apologetic works, of which that against the philosopher Porphyry has not survived. His treatise against Hierocles refutes Hierocles' comparison of Apollonius of Tyana with Christ. The *Praeparatio evangelica* refutes heathen mythology, oracles and astrology, expounds the Jewish Scriptures and argues that whatever was best in Greek thought coincided with or was derived from the Hebrews. The general theme of the *Demonstratio evangelica* (10 books out of 20 are extant) is the fulfillment of Hebrew prophecy in Christ. The *Theophania* (extant in a Syriac version), which deals with the incarnation of the divine word, is largely a repetition of material found in other works.

As a biblical scholar Eusebius used both literal and allegorical exegesis. He had intense admiration for Origen and co-operated with Pamphilus in writing a defense of Origen, now mostly lost. But he did not regard Origen's text of the Greek Old Testament as authoritative and frequently preferred other versions, particularly that of Aquila. In doing so he sought for the text most capable of giving a clear messianic interpretation. He wrote commentaries on various books of the Bible; considerable portions survive of those on the Psalms, Isaiah and Luke. He also produced two books of Gospel questions and solutions, and a series of tables (known as Eusebian canons) listing the parallel passages in the four Gospels. Finally, he compiled a book on biblical topography called the *Onomasticon*.

In his two theological works, *Contra Marcellum* and *De ecclesiastica theologia*, Eusebius exposed and refuted the supposed Sabellianism of Marcellus. A few of Eusebius' epistles also are extant.

BIBLIOGRAPHY.—All extant works in J. P. Migne, *Patrologia Graeca*, vol. xix–xxiv (1857). Critical edition of most works in 8 vol. (1902–56) in *Griechische Christliche Schriftsteller*. Text of the church history reprinted from this edition with Eng. trans. by K. Lake and J. E. L. Oulton in the "Loeb Series," 2 vol. (1926–32); Eng. trans. with introduction and notes by H. J. Lawlor and J. E. L. Oulton, 2 vol. (1927–28); older translation and commentary by A. C. McGiffert (1890) in *Nicene and Post-Nicene Fathers* are valuable; text with French trans. by G. Bardy, 4 vol., in *Sources chrétiennes* (1952–60). The *Praeparatio evangelica* is also ed. by E. H. Gifford with Eng. trans. and notes, 4 vol. (1903). Eng. trans. of the *Demonstratio evangelica* by W. J. Ferrar, 2 vol. (1920). Syriac version of the *Theophania* ed. by S. Lee (1842); Eng. trans. by the same (1843). For editions and studies to date *see* B. Altaner, *Patrologie* (1958); and the annual bibliographies in J. Marouzeau, *L'Année philologique* (since 1920). *See* also J. B. Lightfoot in W. Smith and H. Wace (eds.), *Dictionary of Christian Biography*, vol. i, pp. 308–348 (1880); H. J. Lawlor, *Eusebiana* (1912); D. S. Wallace-Hadrill, *Eusebius of Caesarea* (1960); J. N. D. Kelly, *Early Christian Doctrines* (1958); N. H. Baynes, "Eusebius and the Christian Empire," first published 1933, reprinted in his *Byzantine Studies* (1955); K. M. Setton, *The Christian Attitude Towards the Emperor in the Fourth Century* (1941).

(Js. S.)

EUSEBIUS OF DORYLAEUM, 5th-century bishop of Dorylaeum in Asia Minor (the modern Turkish town of Eskisehir), enters history in the role he was to fill throughout his life, that of an impulsive but courageous champion of the great Christological doctrines which were to receive definitive formulation at the ecumenical Council of Chalcedon. In 429, though still a layman, he was the first to challenge publicly the teaching of his bishop, Nestorius, patriarch of Constantinople, that Mary was legitimately called Mother of Christ but not Mother of God. To Eusebius this denial was implicitly a revival of the heresy of Paul of Samosata that Christ was merely a human person upon whom the Divine Person of the Word rested. Not content with a ringing protest from the congregation while Nestorius was preaching, Eusebius posted throughout Constantinople his famous *Contestatio*, summoning the reader to arouse all the faithful against the patriarch. His action led to the condemnation of Nestorius by the Council of Ephesus in 431.

In 448 Eusebius, now bishop of Dorylaeum, was the first to oppose the heresy which originated in an excessive fear of Nestorianism. Eutyches, archimandrite of a monastery of Constantinople and a friend of Eusebius, was teaching a doctrine later known as the heresy of Monophysitism, maintaining that the nature of Christ was single and not a human nature identical with that of other men. For his efforts in securing the excommunication of Eutyches, Eusebius was deposed by the Latrocinium or Robber Synod of Ephesus in 449. Immediately appealing to Pope Leo I, he was given refuge in Rome. In 451 he was rehabilitated by the Council of Chalcedon, whose classic definitions of the person and natures of Christ he assisted in drafting. *See* also COUNCIL; EUTYCHES; MONOPHYSITES.

For a full treatment of Eusebius and complete bibliography, *see* G. Bareille in the *Dictionnaire de Théologie Catholique*, vol. v, col. 1532–37 (1913).

(T. V. BM.)

EUSEBIUS OF EMESA (d. *c.* 359), 4th-century semi-Arian writer on doctrinal subjects, was born of a noble family of Edessa and endowed with a brilliant mind. He was widely educated, first under the great Eusebius of Caesarea (*q.v.*), from whom he may have imbibed his Arian leanings, and then in the ecclesiastical centres of Antioch and Alexandria. He won the respect of the emperor Constantius, who often had Eusebius accompany him on his expeditions against the Persians. Eusebius was appointed about 339 to the see of Emesa (the modern Syrian city of Homs) after prudently declining the bishopric of Alexandria, from which Athanasius, champion of Nicene orthodoxy, had been deposed by the semi-Arian synod of Antioch with the assistance of imperial troops. Realizing the Alexandrians' affection for Athanasius Eusebius refused the synod's request and accepted Emesa, only to be expelled by the inhabitants there. He was finally reinstated after taking refuge with his close friend, George, bishop of Laodicea. The latter, a partisan of Arius, wrote a biography of Eusebius, fragments of which are preserved in the writings of the 5th-century church historians Socrates and Sozomen. St. Jerome, to whom Eusebius was "a standard-bearer of the Arian faction," disparagingly describes his prolific literary output as noteworthy chiefly for rhetorical exhibitionism.

For the few surviving works of Eusebius, and especially for the 17 homilies ascribed to him, *see* the full bibliography under "Eusebius" in the *Oxford Dictionary of the Christian Church*, p. 474 (1957).

(T. V. BM.)

EUSEBIUS OF LAODICEA was bishop of the Mediterranean seaport of Laodicea (modern Latakia, Syria) *c.* 264–*c.* 269. A former pupil of the illustrious Origen and a deacon of the church in his native Alexandria, he had risked his life in the service of the Christian martyrs there during the Decian persecution of 250 and the Valerian persecution of 257. No less noteworthy was his deliverance of many fellow citizens from starvation during the civil strife which followed the death of Valerian in 261, when Aemilianus, prefect of Alexandria, was contesting the imperial succession with Gallienus. Many Alexandrians had been trapped in the Pyrucheum, the strong quarter of the city, where Aemilianus was besieged by Gallienus' general. Eusebius negotiated the release of noncombatants and met them with food and medicines.

In 264 Eusebius represented his ailing bishop, St. Dionysius of Alexandria, at the synod of Antioch. On his return journey Eusebius was detained by the Laodiceans and persuaded to become their bishop at a time when Christians were eager for saintly leaders whose learning would equal that of their pagan rivals.

Modern knowledge of Eusebius is derived principally from the *Ecclesiastical History* of Eusebius of Caesarea, which contains testimonies in his praise by Dionysius of Alexandria. For the dates of Eusebius *see* the commentary in the translation of the *Ecclesiastical History* by H. J. Lawlor and J. E. L. Oulton (1928; reprinted 1954).

(T. V. BM.)

EUSEBIUS OF MYNDUS (4th century A.D.), Neoplatonist philosopher, was a pupil of Aedesius of Pergamum. He was distinguished from the other members of the Pergamene school by his comparative sobriety and rationality and by his contempt for the religious magic or theurgy to which other members of the school were addicted. He was too sober for the future emperor Julian ("the Apostate"), who turned from his philosophical teachings to the sensations provided by the wonder-worker Maximus of Smyrna.

(A. H. AG.)

EUSEBIUS OF NICOMEDIA (d. *c.* 342), Greek bishop and theologian, was a supporter of Arius and leader of an Arian group called Eusebians after him. At first bishop of Beirut, about 318 he be-

came bishop of Nicomedia (modern Izmit, Turk.). He wrote a letter in defense of Arius to Paulinus, bishop of Tyre, which is preserved in the *Church History* of Theodoret. Eusebius of Nicomedia appears to have been agreed with Eusebius of Caesarea in placing Christ above all created beings, the only begotten of the Father, but in refusing to recognize him to be "of the same substance" with the Father, who is alone in essence and absolute being. Thus at the Council of Nicaea (325) he earnestly opposed the insertion of the Homoousian clause, which states that Father and Son are "of one substance" and was designed to exclude the Arians. Nevertheless, after being defeated in his object, he signed the creed. He refused, however, to sign the anathema directed against the Arians. After the council his zeal against the Athanasians led to a brief banishment from his see. In 339 he was promoted to the see of Constantinople and became the leader of the anti-Nicene party, which position he maintained till his death.

EUSTACE, the name of four counts of Boulogne.

EUSTACE I (d. 1046–1047), who became count in 1023 or 1027, was a son of Arnulf I (d. *c.* 993) and a great-grandson of Baldwin II, count of Flanders (d. 918).

His son EUSTACE II (d. *c.* 1087) was the husband of Goda (d. before 1056), daughter of the English king Aethelred II the Unready and sister of Edward the Confessor. When Eustace visited England (1051) a brawl between his retinue and the men of Dover provoked a serious quarrel between King Edward and Godwin, whose refusal, as earl of Wessex, to punish the town by a harrying expedition, led to his exile (1051–52). Eustace accompanied William, duke of Normandy (afterward William I), to England (1066) and fought at the battle of Hastings. Possibly dissatisfied with his share of English territory, he crossed to England again (1067) and led a rebellion in Kent; but his attack on Dover castle was repelled by the garrison. He was later reconciled with William. By his second wife, Ida of Lower Lorraine, Eustace became the father of Eustace III and of the crusaders Godfrey of Bouillon and Baldwin I, count of Edessa and first king of Jerusalem.

EUSTACE III (d. after 1125) supported Robert II (Curthose), duke of Normandy, in his attempt (1088) to wrest the English crown from his brother William II. He accompanied Robert on the first crusade (1096), returning home late in 1100. On the death of his brother Baldwin (1118), he was elected king of Jerusalem by the crusaders there, but he had only reached Italy when he heard that his nephew Baldwin II had been put forward in his place. He retired to the Cluniac priory of St. Peter at Rumilly in the Boulonnais in 1125, leaving his county of Boulogne to his daughter Matilda and her husband Stephen of Blois, afterward king of England.

EUSTACE IV (d. 1153), their son, was given the county of Boulogne in 1150. He did homage for Normandy (1137) to Louis VII, king of France, whose sister Constance he later married (1140), and he was several times used by the king as an opponent to the claims on the duchy made by the counts of Anjou. At a council held in London on April 6, 1152, King Stephen induced a small number of the English barons to do homage to Eustace as their future king; but Theobald, archbishop of Canterbury, refused to crown him. Eustace died on about Aug. 17, 1153, while engaged in plundering the lands of the abbey of Bury St. Edmunds. His death made possible a peaceful settlement between Henry of Anjou (afterward Henry II) and Stephen, who made no attempt to designate the throne to his younger son, William, earl of Surrey.

See E. Rigaux, "Recherches sur les premiers comtes de Boulogne," *Bulletin de la société académique de Boulogne-sur-Mer,* vol. v (1891–99); P. Héliot, *Histoire de Boulogne et du Boulonnais* (1937).
(G. W. S. B.)

EUSTATHIUS, SAINT, OF ANTIOCH (d. *c.* 337), sometimes styled THE GREAT, was a native of Side in Pamphylia, Asia Minor. He was bishop of Beroea around 320 and became bishop of Antioch shortly before the Council of Nicaea (325). In that assembly he ably opposed the Arians. The subsequent intrigues of the pro-Arian Eusebius of Caesarea (*q.v.*) led to Eustathius' deposition by an Antiochene synod (probably 326) and banishment to Thrace by the emperor Constantine. The resistance of his followers in Antioch developed ultimately into the Melitian schism. Eustathius' later fortunes are unknown. He died probably around 337, and is commemorated on July 16 (Latin Church) and Feb. 21 (Greek Church).

Of Eustathius' many works, written in a much-admired classical style, little is preserved, while several pieces attributed to him are most probably spurious. They may be found in J. P. Migne (ed.), *Patrologia Graeca,* vol. xviii, 609–1066.

See R. V. Sellers, *Eustathius of Antioch and His Place in the Early History of Christian Doctrine* (1928); M. Spanneut, *Recherches sur les écrits d'Eustathe d'Antioche* (1948). (JN. SR.)

EUSTATHIUS OF THESSALONICA (died *c.* 1194), Byzantine scholar and author, is well known for his commentaries on various classical texts, those on the *Iliad* and the *Odyssey* containing valuable extracts from the scholia of other critics whose works are lost. Brought up in Constantinople, he was designated archbishop of Myra in Lycia but was transferred to Thessalonica in 1175. His commentary on the geographical epic of Dionysius Periegetes preserves much of Stephanus of Byzantium and the lost writings of Arrian. Of his commentary on Pindar only the preface, which contains an essay on lyric poetry, a life of Pindar and an account of the Olympic games, remains.

Eustathius was a notable figure in the Byzantine world of his day. He was one of the ecclesiastics who opposed Manuel I's attempt to make it easier for Muslims to become Christian. He was outspoken in his criticism of contemporary monastic laxity. He has left a valuable eyewitness account of the sack of Thessalonica by the Normans of Sicily in 1185, as well as several orations on events during the reigns of Manuel I and Isaac II Angelus.

BIBLIOGRAPHY.—For the commentary on Homer, the preface to the commentary on Pindar and the commentary on Dionysius Periegetes, *see* respectively G. Stalbaum (ed.), 7 vol. (1825–30), F. W. Schneidewin (ed.) (1837), C. W. Müller (ed.), in *Geographici Graeci Minores,* ii (1882). For the Pentecostal hymn *see* A. Mai (ed.), in *Spicilegium Romanum,* v, 2 (1841). For the orations *De capta Thessalonica,* etc. *see* T. L. Tafel (ed.) (1832 and 1839); for the *De Capta Thessalonica* see I. Bekker (ed.) in the Leo Grammaticus volume of the Bonn *Corpus scriptorum Byz. hist.* (1842) and German trans. by H. Hunger, *Die Normannen in Thessalonika* (1955); J. P. Migne, *Patrologia Graeca,* cxxxv, cxxxvi (1864–65). *See* also K. Krumbacher, *Gesch. der byz. Litt.,* pp. 536–541 (1897); G. Moravcsik, *Byzantinoturcica,* i, pp. 135–137, 2nd ed. (1958); K. Bonis, *Eustathios Arkhiepiskopos Thessalonikes* (1950). (J. M. HY.)

EUTECTIC MIXTURE. The eutectic mixture of two substances is that mixture which has the lowest freezing point and is further distinguished by the fact that its two components crystallize out simultaneously in the proportions in which they were present in the liquid. *See* METALLOGRAPHY: *Metals and Alloy Structure.*

EUTHANASIA (MERCY KILLING) usually refers to voluntary death in cases of painful and incurable disease. Since there is no specific provision for it in Anglo-U.S. law, it is accounted either suicide (if performed by the patient himself) or murder (if performed by another). However, a physician may lawfully decide not to prolong life where there is extreme suffering; and he may administer drugs to relieve pain, in the increasing doses necessary to overcome habituation, even though he knows that this may shorten the patient's life. The only country at the beginning of the 1960s to give direct sanction to a form of euthanasia was Switzerland; under the federal code of 1937, lending assistance with a view to suicide was punishable only if the assister was impelled by selfish motives, so that a physician might provide, but not administer, poison at the request of his suffering patient.

The opinion that euthanasia is morally permissible goes back to Socrates, Plato and the Stoics. It is rejected in traditional Christian belief, chiefly because it is thought to come within the prohibition of murder in the sixth commandment. Some Christian writers allow it, however, the earliest specific treatment being in Sir Thomas More's *Utopia* (1516). The organized movement for legalization commenced in England in 1932, when C. Killick Millard founded the Voluntary Euthanasia Legalisation society (later called the Euthanasia society). The society's bill was defeated in the house of lords in 1936, and so also was a motion on the same subject in the house of lords in 1950. The Euthanasia Society of America, founded in 1938, sponsored a similar measure

which was defeated in the Nebraska assembly in 1938; there is also a Voluntary Euthanasia society of Connecticut.

See G. L. Williams, *The Sanctity of Life and the Criminal Law*, ch. 8 (1957 and 1958). (G. L. Ws.)

EUTHYDEMUS, king of Bactria in the late 3rd century B.C. At an uncertain date he overthrew Diodotus II (one of whose satraps he may have been) and took his place as king. In 208 he was attacked by the Seleucid king Antiochus III (*q.v.*) and a long war was fought between them which has been described by the Greek historian Polybius. Euthydemus, having failed in his attempt to defend the line of the Arius (Hari Rud) river, fell back upon his capital, Bactra (Balkh in northern Afghanistan), where he stood a siege lasting two years. In the end, after negotiations had been opened and Euthydemus had stressed the danger which overhung the country from the nomad peoples to the north of Bactria, peace was concluded upon terms by which Euthydemus kept his kingdom while acknowledging Seleucid overlordship. Later, Euthydemus took some territory from Parthia. According to F. Altheim he occupied Sogdiana, Arachosia, Drangiana and Aria at the beginning of his reign. There are clear indications that he won the active support of the Iranian element in his kingdom. The date of his death is unknown. *See* BACTRIA, and for bibliography works cited there. (R. H. SI.)

EUTROPIUS (4th century A.D.), minister of the eastern Roman emperor Arcadius (*q.v.*) and a eunuch, arranged Arcadius' marriage with Eudoxia (*q.v.*) so as to thwart his rival Rufinus, who had wished to marry his own daughter to the emperor. After Rufinus' murder in 395 he was the most influential man in the eastern empire and brought much discredit on himself by his corrupt methods of government. He repelled an invasion of Asia Minor by the Huns in 398, and was nominated consul for 399, the first eunuch to hold the office. In 399 he was overthrown by his political rivals and banished to Cyprus: shortly afterward he was beheaded. He is attacked in two extant poems by the poet Claudianus (*q.v.*). (E. A. T.)

EUTROPIUS (4th century A.D.), Roman historian, was the author of *Breviarium ab urbe condita*, a compendium, in ten books, of Roman history from the foundation of Rome to the accession in 364 of the emperor Valens, to whom he dedicated it. He accompanied Julian on his expedition against the Persians (363) and held the office of secretary (*magister memoriae*) to Valens at Constantinople. The *Breviarium* was compiled from the epitome of Livy, Suetonius and a (lost) series of biographies of emperors, and shows sound judgment and impartiality. The work became a favourite schoolbook. Its independent value is small, but it sometimes fills a gap left by the more authoritative records. The *Breviarium* was enlarged and continued by Paulus Diaconus (*q.v.*) and by Landolfus Sagax (*c.* 1000).

BIBLIOGRAPHY.—There are Greek translations by Capito Lycius and Paeanius, and an almost complete version of the latter is extant. The best edition of Eutropius is by H. Droysen (1878), containing the Greek version and the enlarged editions of Paulus Diaconus and Landolfus. There are smaller critical editions by C. Wagener (1884) and F. Rühl (1887). J. Sorn's *Der Sprachgebrauch des Historikers Eutropius* (1892) treats the grammar and style of Eutropius. (A. H. McD.)

EUTYCHES (*c.* 378–after 451), a revered archimandrite at Constantinople who gave his name to the Eutychian heresy. Brought up in the Alexandrian doctrine of Christ's Person, he confessed that the Word incarnate is "one nature" (*i.e.*, one concrete existence). As a staunch anti-Nestorian, he probably considered that the acceptance by Cyril of Alexandria in 433 of the Antiochenes' "formulary of reunion," with its mention of "two natures" (*i.e.*, the two "substances" of manhood and Godhead) in Christ, favoured the "Nestorianizers" (*see* NESTORIUS). To answer the charge brought against him by the heresy hunter Eusebius of Dorylaeum that he would not recognize the formulary, Eutyches was summoned by Flavian, bishop of Constantinople, himself a supporter of "two natures," to a meeting of the standing synod of Constantinople (Nov. 448). There the aged monk refused to discuss Christ's "natures." His was the faith of Nicaea (325), he declared, and he was following the Fathers in confessing one nature. But he repeatedly affirmed "two natures before, one after the incarnation." He was deposed. Historically, he is regarded as the founder of Eutychianism, an extreme form of monophysitism according to which Christ's human nature was absorbed by his Godhead (*see* MONOPHYSITES). But this verdict seems one-sided; when pressed, Eutyches confessed that Christ's manhood was "consubstantial" with ours and could speak of a "complete" incarnation. Clearly, he was an inexperienced thinker. His appeal to Pope Leo of Rome resulted in the forthright condemnation of Eutychianism in Leo's celebrated *Tome*. Reinstated by the "robber synod" of Ephesus (449), Eutyches was banished as a result of the Council of Chalcedon. *See* also COUNCIL.

BIBLIOGRAPHY.—For text of Eutyches' trial and related correspondence *see* E. Schwartz, *Der Prozess des Eutyches*, in *Sitzungsberichte der bayerischen Akademie der Wissenschaften zu München*, Philos.-hist. Abteilung, part 5 for 1929. *See* also R. Draguet, "La Christologie d'Eutychès," *Byzantion*, vol. vi, pp. 441–457 (1931); T. Camelot, "De Nestorius à Eutychès," in A. Grillmeier and H. Bacht (eds.), *Das Konzil von Chalkedon*, vol. i, pp. 213–242 (1951); R. V. Sellers, *The Council of Chalcedon* (1953). (R. V. S.)

EUTYCHIANUS (EUTYCHIAN), **SAINT,** pope from about 275 to about 283. Fragments of his original epitaph were discovered in the catacombs, but nothing more is known of him. His feast day is Dec. 7.

EUTYCHIDES OF SICYON (fl. early 3rd century B.C.), Greek sculptor, was a pupil of Lysippus. His most noted work was a statue of "Fortune," which he made for the city of Antioch (founded 300 B.C.). The goddess, who embodies the idea of the city, was represented seated on a rock, with the Orontes river at her feet. Copies of this group have been identified in a statue in the Vatican and in several bronze, silver and marble statuettes in New York, Paris and Budapest. The composition also occurs on Syrian coins of Tigranes (83 B.C.).

See Margarete Bieber, *Sculpture of the Hellenistic Age* (1955). (D. M. R.)

EVAGORAS, son of Nicocles, king of Salamis in Cyprus about 410–374 B.C. whose policy was one of friendship with Athens and the promotion of Hellenism in Cyprus. He claimed descent from the Greek hero Teucer and his family had long ruled Salamis until supplanted by a Phoenician exile. He regained the throne about 410. According to Isocrates, Evagoras was a model ruler, whose aim was to promote the welfare of his state and of his subjects by the cultivation of Greek refinement and civilization, which had been almost obliterated in Salamis by a long period of barbarian rule. He cultivated the friendship of the Athenians, and his services were recognized by the gift of Athenian citizenship. After the defeat of Conon (*q.v.*) at Aegospotami (405 B.C.) he afforded him refuge and hospitality. For a time he also maintained friendly relations with Persia, and secured the aid of the king Artaxerxes II for Athens against Sparta. He took part in the battle of Cnidus (394), in which the Spartan fleet was defeated, and for this service his statue was placed by the Athenians side by side with that of Conon in the Ceramicus. But the energy and enterprise of Evagoras soon aroused the jealousy of the Great King, and relations between them became strained. From 391 they were virtually at war. Aided by the Athenians and the Egyptian king Achoris, Evagoras extended his rule over the greater part of Cyprus, crossed over to Asia Minor, took several cities in Phoenicia, and persuaded the Cilicians to revolt. After the peace of Antalcidas (386), to which he refused to agree, the Athenians withdrew their support, since by its terms they recognized the lordship of Persia over Cyprus. For ten years Evagoras carried on hostilities single-handed, except for occasional aid from Egypt. At last he was totally defeated at Citium (Larnaca) in 381 and compelled to flee to Salamis. Here, although closely blockaded, he managed to hold his ground and took advantage of a quarrel between the Persian generals to conclude peace. Evagoras was allowed to remain nominally king of Salamis, but in reality a vassal of the king of Persia, to whom he was to pay a yearly tribute. In 374 he was assassinated by a eunuch.

The chief authority for the life of Evagoras is the panegyric of Isocrates addressed to his son Nicocles; *see* also K. Spyridakis, *Euagoras I von Salamis* (1935).

EVAGRIUS PONTICUS (346–399), a monk whose ascetical writings strongly influenced the spiritual teaching of the church,

was born in Pontus. Made a lector by Basil of Caesarea and a deacon in Constantinople by Gregory of Nazianzus, he remained in Constantinople under Nestorius until, after a spiritual crisis, he left for Jerusalem where he became a monk. Later he went to Egypt, living first in the desert and then in a colony of hermits. He died in Scete (Wadi en Natrun).

His writings, all monastic, survive almost complete in Syriac translation, but only in excerpts in the original Greek. They include a work against the eight principal sins, 100 maxims for uneducated monks (*Capita practica*), 50 maxims for educated monks (extant in a Syriac version), advice to monks and advice to nuns. Some of his writings survive under the names of other Fathers, such as Basil and Nilus; a fragmentary commentary on the Psalms attributed to Origen may be his work. In his own day Evagrius was highly regarded as a spiritual guide to monks and hermits. He exerted a great influence on later spiritual writers, notably Dionysius the Areopagite and Maximus the Confessor. His authority, attested by Jerome and Palladius, was unquestioned until in the 6th century he shared in the anathemas against Origen, whose teaching he followed.

Bibliography.—Greek works (incomplete) in J. P. Migne, *Patrologia Graeca*, vol. xl (1858); advice to monks and to nuns ed. by H. Gressmann, *Texte und Untersuchungen*, vol. xxxix, part 4b (1913). Syriac version partially ed. by W. Frankenburg (1912) and J. Muyldermans (1952). *See* also W. Bousset, *Apophthegmata*, pp. 281–336 (1923); S. I. Hausherr, "Les Versions syriaque et armenienne d'Evagre le Pontique," *Orientalia Christiana*, 22:69–118 (1931); M. J. Rondeau, "Le Commentaire sur les psaumes d'Evagre le Pontique," *Orientalia Christiana periodica*, 26:307–348 (1960). (A. O.)

EVAGRIUS SCHOLASTICUS (*c.* 536–early 7th century), Byzantine historian, was the author of an *Ecclesiastical History* in six books extending from the third general council at Ephesus in 431 to the year 594. Born at Epiphania in Coele-Syria, he became a lawyer and probably practised at Antioch. He was the legal adviser of Gregory, patriarch of that city, whom he successfully defended at Constantinople against certain serious charges. This brought him to the notice of the emperor Tiberius Constantinus, who honoured him with the rank of quaestorian; Mauricius Tiberius made him master of the rolls. His history continues the work of Eusebius, Socrates, Sozomen and Theodoret. Though somewhat credulous it is on the whole impartial. Its main value lies not in its independent literary merit but in its frequent excerpts from excellent sources which have thus been preserved. It is particularly useful for the history of dogma during the 5th and 6th centuries. Evagrius also drew on the writings of Eustathius Epiphaneus, Joannes Epiphaneus, Joannes Malalas, Procopius and probably Menander Protector.

Bibliography.—A critical edition of *The Ecclesiastical History of Evagrius, with the Scholia*, by J. Bidez and L. Parmentier (1898); J. P. Migne, *Patrologia Graeca*, lxxxvi (1860). *See* also K. Krumbacher, *Geschichte der byzantinischen Litteratur*, pp. 245–247 (1897); O. Bardenhewer, *Geschichte der altkirchlichen Litteratur*, vol. v, p. 119 ff. (1932); G. Moravcsik, *Byzantinoturcica*, i, pp. 257–59, 2nd ed. (1958). (J. M. Hy.)

EVANDER, in classical mythology, is represented in a late and artificial legend as an Arcadian who settled very early on the Palatine, at Rome, where he founded a town named Pallantion, after his native place. He also had a son, Pallas. He instituted the Lupercalia (*q.v.*), in imitation of the Arcadian Lycaea, and introduced some of the blessings of civilization, including writing. He hospitably received Hercules and Aeneas. His mother was the goddess Carmentis or Carmenta. His father was Hermes.

The one bit of reality in all this is the name of Carmentis (Carmenta); the rest is bad etymologies of *Palatinus* and *Lupercalia;* perhaps *Euandros* is intended to mean "strong man," in allusion to the name Roma, identified with Greek *rhome*, "strength."

EVANGELICAL ALLIANCE, a British association of churches, Christian societies and individual Christians of different denominations who "hold and maintain what are usually understood to be evangelical views in regard to matters of doctrine." The alliance came into being as the outcome of a conference of Christian leaders held in London in 1846. It was stated at the outset that its purpose was "not to create Christian union, but to confess the unity which the Church of Christ possessed as His Body." The following nine tenets of evangelical faith formed its doctrinal basis: the divine inspiration, authority and sufficiency of the Scriptures; the right and duty of private judgment in the interpretation of them; the unity of the Godhead and the Trinity of Persons; the utter depravity of human nature; the incarnation of the Son of God and his atonement for the sins of all men; the justification of sinners by faith alone; the work of the Holy Spirit as sanctifier; the immortality of the soul, the resurrection of the body and the final judgment by Jesus Christ; and the divine institution of the Christian ministry.

The alliance is primarily concerned with spiritual unity, to which end it instituted (1846) the universal week of prayer observed in the first full week of January, and with the welfare of persecuted Christians and the furtherance of Christian liberty and religious toleration. Its efforts in the latter direction were particularly successful in the 19th century, notably in Sweden, Russia and Turkey.

In England, where it has its headquarters, the alliance has played an active part in various nationwide evangelistic campaigns and was one of the pioneers of film evangelism in prisons. It has also set up hostels in central London for foreign students. In association with kindred societies, the alliance seeks to emphasize the reformed character of the Church of England.

In 1958 the alliance was instrumental in bringing into being the Evangelical Missionary alliance, the object of which is "to provide a medium of fellowship and effective co-operation in the interest of evangelical missionary work and service overseas." Each member mission—denominational, interdenominational and undenominational societies may join—enjoys complete autonomy but subscribes to a common doctrinal basis.

The Evangelical alliance is closely associated with similar alliances in many different countries of the world. In the United States its counterpart is the National Association of Evangelicals, while in most European countries there are evangelical alliances. There is close liaison with the evangelical fellowships of India and Ceylon, as well as those in other Asian countries.

See J. W. Ewing, *Goodly Fellowship* (1946); J. E. Orr, *The Second Evangelical Awakening in Britain* (1949). (Gi. W. K.)

EVANGELICAL AND REFORMED CHURCH, a Protestant church established June 26, 1934, by union of the Evangelical Synod of North America and the Reformed Church in the United States. In July 1961 it was united with the Congregational Christian Churches to form the United Church of Christ. For the earlier history of the Reformed body *see* Reformed Churches: *United States: The Evangelical and Reformed Church; see* also Congregationalism: *United States: United Church of Christ.*

EVANGELICAL UNITED BRETHREN CHURCH was formed in 1946 by the merger of the Evangelical Church and the Church of the United Brethren in Christ. Its members, numbering around 750,000 in North America, are found chiefly in the northern states and in the Canadian provinces; overseas operations are maintained in the orient, South America, Africa and Europe.

The church originated as a German-language movement in Maryland and southeastern Pennsylvania in the period following the American Revolution. The three initial leaders came from Reformed, Mennonite and Lutheran backgrounds, respectively; but Methodism exerted the most formative influence upon its life.

The polity of this church centres around "conferences" at local (congregational), annual (regional) and general (international) levels. Local conferences elect lay delegates who, with the ordained ministers, constitute the annual conferences. Quadrennially, the annual conferences elect an equal number of clergymen and laymen who constitute the general conference, the ultimate authority on all matters in the church. Ministers (elders) are ordained but once. General conference elects bishops from among the elders to superintend the life and programs of the denomination.

This church has never been doctrinaire. Its theological foundation is affirmed in "The Confession of Faith" in the *Book of Discipline.* The major emphasis is on personal salvation. It has proclaimed conversion, personal and experiential, as the occasion when by God's initiating grace man responds to the Saviour's

invitation and enters the Kingdom and the church. In the mid-20th century there was a growing appreciation of Christian worship, liturgy and architecture.

The denomination's evangelical rather than doctrinal or ecclesiastical emphases go far to explain its traditionally ecumenical outlook. It is a member of all the major interchurch programs and movements, evangelistic, missionary and ecclesiastical. In 1963 a plan for its union with The Methodist Church was submitted for study. If a constitution was agreed upon by the two bodies, merger would take place in 1968. (P. H. EL.)

EVANGELISM: *see* REVIVALISM.

EVANS, SIR ARTHUR JOHN (1851–1941), British archaeologist whose most celebrated discovery was the Minoan civilization of Crete, was born at Nash Mills, Hertfordshire, on July 8, 1851, eldest son of Sir John Evans (*q.v.*). He was educated at Harrow, at Brasenose college, Oxford, to which he was elected a fellow, and at Göttingen university. Evans' lifelong interest in the archaeology and anthropology of the Balkans was inspired by a visit in 1875, and he went to live at Ragusa (modern Dubrovnik) in 1877. When Serbia and Montenegro declared war on Turkey, Evans became *Manchester Guardian* correspondent in the area, leaving Ragusa in 1882 after seven weeks' imprisonment by the Austrians on a charge of complicity in a rebellion in Hercegovina. Evans was appointed curator of the Ashmolean museum, Oxford, in 1884, superintending the removal to new buildings and the reintegration of the scattered collections. On retirement from this position, 1908, he became extraordinary professor of prehistoric archaeology.

Evans' singular ability to grasp the significance of details, and his great independence and determination in the execution of his ideas, were illustrated well in his Cretan work. His attention was directed toward the island through his interest in numismatics; ascertaining the Cretan origin of certain engraved gems seen in Athens, he visited Crete in 1894. In 1899, after much negotiation, Evans bought the Kephala site at Knossos. Excavations resulted in the discovery of a hitherto unknown Bronze Age civilization which Evans named Minoan after a legendary Cretan king. (For an account of the excavations of the palace *see* CRETE: *Archaeology.*) Distinguished by the originality and high development of its art and architecture, the Minoan culture is considered to be to some degree parental to the Mycenaean of Greece and there are indications of its contact with Europe. Evans' discoveries proved invaluable to the study of European, and especially of eastern Mediterranean, prehistory. Many examples of a pictographic script were found and two forms, distinguished as Linear A and B, were identified (*see* MINOAN LINEAR SCRIPTS). In 1953 M. Ventris and J. Chadwick announced their decipherment of Linear B, suggesting that it is an archaic form of Greek; the nature of Linear A is controversial. The validity of Evans' dating and of his conclusions was questioned in 1960, but the great importance of the Cretan excavations, upon which he spent a large amount of his personal income, was not diminished.

Evans' work was recognized by university honours. Elected a fellow of the Royal society (1901), Evans was knighted in 1911 and was president of the Society of Antiquaries, 1914–19, and of the British association, 1916–19. He died near Oxford on July 11, 1941. His most important publications include *The "Horsemen" of Tarentum* (1889); *Cretan Pictographs and Prae-Phoenician Script* (1895); *Scripta Minoa*, vol. 1 (1909), vol. 2, ed. by J. L. Myres (1952); and *The Palace of Minos* (1921–36).

(J. M. WI.)

EVANS, SIR JOHN (1823–1908), English archaeologist, geologist and numismatist, specialized in studies of early Britain. The son of the Rev. A. B. Evans, headmaster of Market Bosworth grammar school, he was born at Britwell Court, Buckinghamshire, on Nov. 17, 1823, and educated at his father's school. He was the author of three books, standard in their respective departments: *The Coins of the Ancient Britons* (1864, repr. 1890); *The Ancient Stone Implements, Weapons and Ornaments of Great Britain* (1872, 2nd ed. 1897); and *The Ancient Bronze Implements, Weapons and Ornaments of Great Britain and Ireland* (1881). He was president of the Numismatic society (1872–1908) and treasurer of the Royal society (1878–98). As president of the Society of Antiquaries he was ex officio a trustee of the British museum, and subsequently became a permanent trustee. He was created knight commander (of the order) of the Bath in 1892. He died at Berkhamsted on May 31, 1908. One of his sons, Sir Arthur John Evans (*q.v.*), became the famous excavator of the palace of Knossos, Crete.

EVANS, OLIVER (1755–1819), 18th-century American inventor who anticipated 20th-century methods of automation, was born on a farm near Newport, Del., in 1755.

Apprenticed to a wagonmaker at 15, Evans studied mathematics and mechanics in his spare time. In 1777, while employed in the manufacture of card teeth for carding wool, he invented a machine that turned out 1,500 cards a minute. In 1785 he built and put into successful operation a water-powered flour mill, run by one man, that performed every operation in the conversion of wheat into flour. A system of conveyors and elevators moved raw material, meal and finished product on a continuous production line.

Evans' most important contribution was in devising improvements of the steam engine that enabled operation at high pressure. In 1804 he built a steam-powered dredging scow that anticipated both the automobile and the steamboat. He steered the machinery-laden scow through the streets of Philadelphia from his shop to the Schuylkill river.

He died at New York city on April 21, 1819.

EVANSTON, the second largest city of Cook county, in northeastern Illinois, U.S., is on Lake Michigan immediately north of the city of Chicago, with which it is connected by expressways, rapid transit bus and elevated-subway lines and railway commuter service.

Evanston grew up around Northwestern university, which became the city's largest employer. The university campuses in Evanston are on the shore of Lake Michigan, where a group of Chicago businessmen first purchased property for the university in 1853 under a special state charter of 1851. The city was named for John Evans, one of the founders of the university. Incorporated as a town in 1863, and as a city in 1892, Evanston increased in population until it reached 19,259 in 1900 and 79,283 in 1960. It is noted as an educational and religious centre. The second assembly of the World Council of Churches was held in Evanston in 1954, and the Ecumenical institute was established there in 1958. The elementary and secondary schools are nationally recognized for excellence. Northwestern university, which is privately controlled, had about 7,000 full-time students on its Evanston campus by the second half of the 20th century. Its Chicago campus for professional schools and evening classes was dedicated in 1926, and the Technological institute building in Evanston was completed in 1942. Other institutions of higher education in Evanston include Garrett Biblical institute, a Methodist graduate theological school; Kendall (junior) college; the National College of Education for nursery and elementary school teachers, founded in Chicago in 1886 and moved to Evanston in 1926; and Seabury-Western Theological seminary (Episcopal). A number of national organizations established headquarters there, including the American Academy of Pediatrics, the National School Boards association, the National Merit Scholarship corporation, Rotary International and the Woman's Christian Temperance union. Evanston was the home of the temperance advocate Frances E. Willard (*q.v.*).

Though the city is primarily a residential suburb, about 30% of its employed residents earn their living by working within the city. Manufactures include steel products, hospital supplies, paint, chemicals, food products and books. According to the U.S. census at mid-century, Evanston had the highest percentage of college graduates of any city of over 50,000 population.

The council-manager form of government was adopted in 1952 and retained by a referendum in 1959. The city council comprised a mayor (part-time official) elected at large and 18 aldermen elected from nine wards; the council appointed the city manager. Evanston won a number of awards for traffic safety, including national recognition for consecutive days without a traffic fatality.

(B. W. J.)

EVANSVILLE, seat of Vanderburgh county, Ind., U.S., is located on a sweeping bend of the Ohio river in the southwestern part of the state. Coal deposits and oil fields in an area of fertile farm lands of the Wabash-Ohio valley surround Evansville and, together with the city's excellent location for transportation purposes, contribute substantially to its economic development. The important industries in the city and the immediate area produce aluminum, air-conditioning equipment, nutritional (especially foods for infants) and pharmaceutical products, excavators and cranes, agricultural equipment, furniture, casters, cigars, cake and pastry mixes, office furniture, structural steel, plastic items, uniforms, processed meats, refrigerators and freezers. Evansville has a modern river terminal which provides facilities for interchange of barge, rail and truck traffic.

In 1812 Hugh McGary, the founder of Evansville, purchased 200 ac. of land from the federal government and operated a ferry across the Ohio river from a location which now is downtown Evansville. McGary, hoping that his site would become the county seat, sought the advice and assistance of Robert Morgan Evans, a colonel of the militia and member of the territorial legislature, for whom the city was eventually named. In 1818 Evansville became the seat of Vanderburgh county and in 1847 it was chartered as a city. The Wabash and Erie canal was completed to Evansville, its southern terminus, in 1853 and connected Lake Erie with the Ohio river by way of Toledo, Fort Wayne and Terre Haute.

Evansville college, moved in 1919 from Moores Hill, Ind., where it had been founded in 1854, is a Methodist-related liberal arts institution. The city has a number of parks, including an excellent zoo. Pop. (1960), city 141,543; standard metropolitan statistical area (which includes all of Vanderburgh county and Henderson county, Ky.) 199,313. For comparative population figures *see* table in INDIANA: *Population.*

See James Morlock, *The Evansville Story* (1956). (O. J. J.)

EVARISTUS, SAINT, pope from about A.D. 97 to about 105. He is usually called a martyr, but his martyrdom is not proved. His feast day is Oct. 26.

EVARTS, WILLIAM MAXWELL (1818–1901), U.S. lawyer and statesman who took part in the three greatest public cases of his generation, serving as counsel for Pres. Andrew Johnson in the impeachment trial before the United States senate (1868), counsel for the United States in the "Alabama" arbitration at Geneva, Switz. (1872), and chief counsel for the Republican party in the disputed Hayes-Tilden presidential election of 1876. He was born in Boston, Mass., on Feb. 6, 1818, into a prominent New England family. Educated at Yale college and Harvard law school, Evarts was admitted to the bar in New York in 1841. He quickly took high rank in his profession and during the next 40 years appeared in many important cases including *People* v. *Draper* (1857), the Legal Tender cases in the United States supreme court (1868) and *Theodore Tilton* v. *Henry Ward Beecher* (1875).

At the same time that Evarts was rising to the top of the legal profession he was increasingly active in politics, first as a Whig and then as a Republican. In 1860 he was chairman of the New York delegation to the Republican national convention. On two occasions during the Civil War the state department sent him on special missions to England. From July 15, 1868, to March 1869 he served as attorney general under President Johnson and during Pres. Rutherford B. Hayes's administration (1877–81) he was secretary of state. In 1885 he was elected United States senator from New York and served one term.

Evarts was a witty after-dinner speaker and noted orator. On July 4, 1876, he delivered the chief oration at the national centennial celebration of the signing of the Declaration of Independence, and seven years later at the final centennial celebration of the events of the American Revolution he was again the orator of the occasion.

With his vision seriously impaired, Evarts retired from the senate in 1891 and withdrew from both political and professional life. After ten years in retirement he died in New York city on Feb. 28, 1901.

See C. L. Barrows, *William M. Evarts* (1941); Brainerd Dyer, *The Public Career of William M. Evarts* (1933). (B. Dy.)

EVATT, HERBERT VERE (1894–), Australian statesman and judge, famous especially for his work during World War II and his successful opposition to the banning of the Australian Communist party, was born at East Maitland, New South Wales, on April 30, 1894. He secured a brilliant academic record at the University of Sydney. He entered the New South Wales legislature as a Labor member in 1925, and five years later was appointed justice of the high court of Australia. In 1940 he took the unprecedented and dramatic step of re-entering politics, and won election to the federal parliament. Evatt was a dominant member of the Labor governments of John Curtin and J. B. Chifley from 1941 to 1949. As attorney general he sought, not altogether successfully, to strengthen the power of the federal government vis-à-vis the states. As minister for external affairs and a member of the advisory war council, he played a leading part in the direction of Australia's war policy and in the foundation of the United Nations organization, of which he was president in 1948. Few Australians have made a comparable mark in international politics.

In 1951 Evatt became leader of the opposition, and the first "intellectual" to head the federal Labor party. The only appropriate climax to his career was the prime ministership, but he never achieved it. A major reason was a split in his party between the left and the right (largely Roman Catholic) wings. Having attempted conciliation, Evatt finally drove out the extreme rightists. Some such ending was inevitable, yet Evatt was criticized for waiting overlong to act, and then acting overfiercely. He finally withdrew from politics in 1960 to become chief justice of New South Wales.

Evatt's enigmatic quality derived in part from the nature of his political belief. He was a radical liberal, never a common creed in Australia, and increasingly out of fashion. Nevertheless, it inspired one of Evatt's major triumphs: his leadership of the successful opposition to R. G. Menzies' plan to outlaw the Communist party of Australia. This was a highlight of many years' consistent advocacy of complete civil liberty.

Evatt maintained many sporting and cultural interests. He wrote on Australian history, constitutional law and international affairs. In all fields his contribution was substantial, although—as in his politics—an occasional false note obtruded. Historians may long debate whether, all in all, Evatt was faintly ridiculous or nearly sublime. (O. M. R.)

EVE: *see* ADAM AND EVE.

EVELYN, JOHN (1620–1706), English man of letters and author of a famous *Diary,* was born on Oct. 31, 1620, at Wotton, near Dorking, Surrey. His grandfather had introduced into England the manufacture of gunpowder; his father was a wealthy and cultivated landowner. The young Evelyn, the second son, was brought up by his mother's family at Lewes. At 16 he entered the Middle Temple, principally for its social training, and went to Balliol college, Oxford. He also attracted the attention of Thomas Howard, earl of Arundel, the art collector, who interested him in the visual arts; Evelyn's appreciation of landscape, whether or not spontaneous, was probably encouraged in Arundel's household. Although a devout Anglican he did not join the king in the Civil War, mainly for fear of endangering his brother's estate at Wotton, which was in parliamentary territory. In 1643 he went abroad, first to France, thence by Marseilles and Genoa to Rome, and thence to Venice and Padua, returning by the Simplon pass to Paris in 1646. In 1647 he married Mary, daughter of Sir Richard Browne, the king's diplomatic representative there. Evelyn had studied anatomy and physiology in Padua, and in Paris he attended courses in chemistry and in the cultivation and use of plants. In 1652, when Charles II's fortunes were at their lowest, he returned to England to await better times. He acquired his father-in-law's estate, Sayes Court, in Deptford, at that time a Thames-side village some miles from London, with a naval dockyard. His wife bore him five sons, of whom only one reached maturity, and, later, three daughters. For some time he attended divine service in Deptford parish church, with special services conducted in his house by orthodox Anglican divines, but he preferred the orthodox services in London, and in 1657 became a member of the most important congregation, that using Exeter House chapel in the Strand.

Evelyn had started as author in 1649 with *Of Liberty and Servitude,* a short moral treatise translated from the French of de la Mothe le Vayer. His most important venture in these years was a translation of the first book of Lucretius' *De Rerum Natura,* with a commentary based on the writings of recent philosophical atomists (1656). He took no part in royalist conspiracies, but in 1659 tried to induce an old schoolfellow, governor of the Tower of London, to declare for the king, and he published two pamphlets on the king's behalf.

At the Restoration, because of his father-in-law's work and his own merits, he was well received by Charles II and, while only occasionally holding office at court, spent much time in Whitehall palace. He served on various commissions, including those concerned with London street improvements (1662), the Royal mint (1663) and the repair of Old St. Paul's (1666). Far more important was the commission for Sick and Wounded Mariners and for Prisoners of War in Charles II's Dutch Wars (1664–67, 1672–74); Evelyn was individually responsible for its local activities in Kent and Sussex. Work was especially heavy in 1665, when there were several engagements off the Kentish coast and the plague had driven the court from London. Evelyn received much help from one of the navy officials, the diarist Samuel Pepys (*q.v.*), with whom he formed a lifelong friendship. He served on a council for colonial affairs from 1671 to 1674.

Evelyn had by now made his mark as an author. In 1661 he became a member of the society of scientists which became in 1662 the Royal society (*q.v.*). He was appointed to its council by its first and second charters, and was a keen and active member of the society for the rest of his life. In 1662 the commissioners of the navy, alarmed by the depletion of standing timber in England in the last 20 years, addressed some queries on the subject to the new society. They were referred to Evelyn and three other fellows, and Evelyn elaborated the replies, publishing the result in 1664 as *Sylva, or a Discourse of Forest-Trees, and the Propagation of Timber in His Majesties Dominions.* The volume also contains *Pomona,* which deals with fruit for cider, and *Kalendarium Hortense,* a gardening almanac. *Sylva* was intended for estate owners; it describes the cultivation of the various kinds of trees, and the uses of their timber, but in language unsuitable for practical woodmen and with many anecdotes and classical quotations. A second edition of the three pieces, *Sylva* enlarged, appeared in 1670; a third in 1679, *Sylva* further enlarged, with a new piece added, *Terra,* about various kinds of soil; the fourth edition, now called *Silva,* still further enlarged, with yet another piece added, *Acetaria,* about all plants suitable for salads, was published in 1706, shortly after Evelyn's death. A fifth edition followed in 1729, and in 1776 *Silva* reappeared by itself, much altered by Alexander Hunter and more scientific in character. This, with *Terra* added, went through four more editions, the latest 1825. The record of editions proves the book's usefulness.

In 1662 Evelyn published a small book on engraving and etching, *Sculptura,* in which he announced a new process, mezzotint. Prince Rupert, its first great practitioner, supplied a specimen plate for the book. In 1664 there appeared *A Parallel of the Antient Architecture with the Modern,* translated by Evelyn from the French of Roland Fréart de Chambray, with further matter, in part original; the most important part of the book is the illustration of columns of the five classical orders, showing the proportions of the parts. The book appeared before its time and sold badly, but Evelyn prepared a new edition, which appeared in 1707, and further editions followed in 1723 and 1733. Besides several smaller pieces, original or translated, Evelyn also undertook, by desire of Charles II, a history of the Second Dutch War, but he was too dilatory or too thorough for Charles, who wanted to use the book as propaganda, and its introduction alone was published, as *Navigation and Commerce* (1674).

About 1670, when Evelyn was about 50 and his eldest daughter five years old, he met a maid of honour at court, Margaret Blagge. The orphan daughter of an impoverished royalist, devout and virtuous, she was in love with but unwilling to marry Sidney Godolphin, the future lord high treasurer and first earl of Godolphin. She wanted the practical advice and the moral support that a father could give; Evelyn was enraptured by the attention that she paid him; in 1672 they formed a compact of friendship and met regularly for prayer and conversation. Eventually Godolphin prevailed on her to marry him secretly, and she did not announce the marriage to Evelyn for some months. Her relationship with him, at first very close and fervent, had already broadened into a family friendship; little affected by her concealment of the marriage, this friendship was now extended to Godolphin and his kin. She died after giving birth to a child in 1678. A few years later Evelyn wrote an account of her life for one of her friends, which remained unpublished until 1847 and is now regarded as one of the most moving of 17th-century biographies.

In 1685, some months after James II's accession, Evelyn was appointed one of three commissioners of the Privy Seal, an office without political influence which he held for 15 months. It was lucrative, but proved irksome when he was called on to seal documents of questionable legality. Subservient as he was by nature, he had begun to lose his loyalty to the Stuarts as a result of Charles II's conduct in his last years, his attack on civil liberties and his acquiescence in Louis XIV's persecution of his Protestant subjects and lawless extension of French territory toward the Rhine. James II's attack on the Church of England, with the Revocation of the Edict of Nantes to suggest its tendency, completed Evelyn's alienation. He accepted fully the new settlement achieved by the Glorious Revolution, with only occasional misgivings when the war news was bad, and was pleased by the promotion of the great Latitudinarian divines. In 1694 he left Sayes Court. The garden there, formal in plan, was now distinguished by its great variety of fine trees; a special feature was a holly hedge 400 ft. long and 9 ft. high. Peter the Great occupied the house for about two months in 1698. Evelyn joined his elder brother at Wotton. In 1695 he was appointed treasurer of the foundation fund of Greenwich hospital. His last important book, *Numismata. A Discourse of Medals, Antient and Modern,* was published in 1697; while too diffuse, it is interesting for historical portraiture. Evelyn died on Feb. 21, 1706, and was buried at Wotton, which he had inherited seven years earlier.

Assessment of Evelyn's Works.—Evelyn published in all about 30 books. One group of writings relates to the fine arts and general culture, a second to silviculture and gardening, the remainder are religious and occasional pieces (many of the smaller pieces are reprinted in his *Miscellaneous Writings,* 1825). He also left much in manuscript. Although he wrote so much, his writing is generally undistinguished. He was observant and inquisitive, but lacked powers of synthesis and a sense of relevance. His defects are most noticeable in *Sylva* and *Numismata;* he is at his best where his material requires little manipulation, as in *The Life of Mrs. Godolphin* and in his diary.

The diary began when Evelyn was 11 years old with notes in almanacs. He made longer notes during his travels, and about 1650 formed the habit of making notes every few days, which he continued until within a month of his death. About 1660 he began to copy his old notes, amplifying them from memory and from books; the volume in which he made this copy he called "Kalendarium." He broke off when he had reached the year 1645, resumed again about 1682, and probably brought his copying up to date in 1684, after which he made his original notes in the Kalendarium. About 1700 he began a fresh recension, which only covered the years up to 1644. There survive a few of the original notes, the Kalendarium (less a few leaves), the recension of 1700 and old transcripts of some sections. Evelyn made his notes for himself alone. He perhaps began the Kalendarium for his son, but forgot about him as he advanced; the late recension was probably written for Evelyn's grandson and his successors; but no part of the diary was written for publication.

The *Diary* was first published in 1818 and an improved edition appeared in the following year. About half the Kalendarium was printed, but much rehandled and rewritten. This, apart from a few additions, was the diary as known to 19th-century readers. It was much loved as an exemplary biography and as a middle term between history and historical fiction. The first complete and faithful edition appeared in 1955. Evelyn's notes range from bald

memoranda to elaborate set pieces: descriptions of places and events, in which, both in his original notes and when copying them, he used guidebooks and newspapers freely; characters of persons; and reports of sermons, which occupy a large part of the diary. While there is much heavy-going, many passages—such as the wonderful account of Charles II's last fortnight—have a freshness to atone for the absence of finer qualities. The diary bears witness to the history of more than half-a-century, its social life and culture, its religious feeling and its political transformation; the life of the court; the early days of the Royal society; Christian and charitable endeavour; and especially the life of the Anglicans during the Commonwealth and Protectorate, and the change in public opinion preceding and underlying the Glorious Revolution of 1688. While the diary contains relatively little about Evelyn himself, it was obviously created by a good and fair-minded man, which is what gives it its value as historical evidence.

BIBLIOGRAPHY.—*The Diary* was ed. by E. S. de Beer, 6 vol. (1955); *The Life of Mrs. Godolphin* was ed. by H. Sampson (1939); there is a bibliography of Evelyn's writings by Sir Geoffrey Keynes (with full histories of the various items), *John Evelyn* (1937); his correspondence is published in *Diary and Correspondence of John Evelyn*, 4 vol., Bohn's Historical library (1859, etc.). For an adverse view of Evelyn's character *see* W. G. Hiscock, *John Evelyn and Mrs. Godolphin* (1951), and *John Evelyn and His Family Circle* (1955); the case against Hiscock is to be found in E. S. de Beer's articles in *Notes and Queries*, "New Series," vol. 7, no. 6, 7, 8 (June–Aug. 1960). (E. S. DE B.)

EVENING SCHOOL, a school open during the evening hours to provide educational facilities for those who work during the day. *See* ADULT EDUCATION.

EVENKI NATIONAL OKRUG, part of Krasnoyarsk *krai* (*q.v.*) of the Russian Soviet Federated Socialist Republic, U.S.S.R., covers much of the central Siberian plateau. Area 287,645 sq.mi. It is drained northward by the headwaters of the Khatanga and westward by tributaries of the Yenisei, notably the Nizhnyaya and Podkamennaya Tunguska. In the northwest the Putorana mountains rise to 6,683 ft. in Mt. Kamen. Apart from higher areas with tundra vegetation the *okrug* is covered by coniferous forest, dominated by Dahurian larch, but thinning to the north, with stunted trees. Permafrost underlies the poor podzol soils. The climate is severely continental, with long, extremely cold winters (January average about −17° F.). Rainfall is 12–15 in. a year, although higher in the Putorana.

Throughout the *okrug*, which was formed in 1930, there lived 10,320 persons in 1959 (0.04 per square mile). Almost all are Evenki, a Tungus-Manchurian tribe, believed to have some connection with the American Indian. They live by reindeer herding, hunting and fishing. Only herders now move with the reindeer, the bulk of the population having been settled in the 16 villages of the *okrug*. There is one urban district, Tura (pop. 2,000), the administrative centre on the Nizhnyaya Tunguska. The timber resources and deposits of coal and graphite are scarcely exploited. Communications are by river in summer and by winter trails and the Tura airstrip. (R. A. F.)

EVERDINGEN, ALLART VAN (1621–1675), Dutch painter and engraver, was born in Alkmaar. According to A. Houbraken, he studied under Roelant Savery at Utrecht and under Pieter de Molyn at Haarlem. He married in 1645, and later settled in Amsterdam, where he died in 1675. His earlier works include some sea pieces, but his characteristic pictures are mountain landscapes, with forests, log huts and waterfalls. These are northern scenes, the result of a visit to Sweden soon after 1640, and they inspired Jacob van Ruisdael's work in the same vein. Though artificial by comparison with Ruisdael, betraying influence of Savery's more primitive style, Everdingen's work has considerable quality and vivacity of handling. (R. E. W. J.)

EVEREST, SIR GEORGE (1790–1866), English surveyor, mathematician and geodesist, famous for his measurement of the great meridional arc in India, by which he calculated the figure of the earth, and for the gridiron system of geodetic triangulation on which the accurate survey of the subcontinent of India depends. He was born at Greenwich, Eng., on July 4, 1790, son of W. T. Everest. He was educated at Marlow and the Royal Military academy, Woolwich. He reached India in 1806, and served as a lieutenant in the Bengal artillery from 1813 to 1816 in Java, where his surveys led to his selection as assistant to Col. W. Lambton, founder of the great trigonometrical survey of India. Entrusted with the principal triangulation in central India he worked untiringly until forced by ill-health to spend a year in South Africa. There he studied the short arc measured by N. de Lacaille in 1750–52. Everest succeeded Lambton in 1823, but two years later returned to England after 19 years in the east. While at home he discussed with leading mathematicians his proposals for India, being elected a fellow of the Royal society in 1827. He returned to India in 1830 to become surveyor general. He completely reorganized the survey, introduced the most accurate instruments of the day and carried his measurements and calculations to a successful end. He retired to England in 1843. Everest was knighted in 1861 and died in Paddington, London, on Dec. 1, 1866. The highest mountain in the world, Mt. Everest, was named after him in 1863; but his great arc, from Cape Comorin to the Himalayas, is his true memorial. (K. M.)

EVEREST, MOUNT, the highest mountain in the world, a rock and snow peak of the Himalayas, stands on the border of Nepal and Tibet in about latitude 28° N. and longitude 87° E. Its height is 29,028 ft. The peak was named after Sir George Everest (*q.v.*). Its Tibetan name is Chomolungma.

The expedition which first reached the summit, in 1953, was the eighth to make the attempt; there had also been three reconnaissance expeditions.

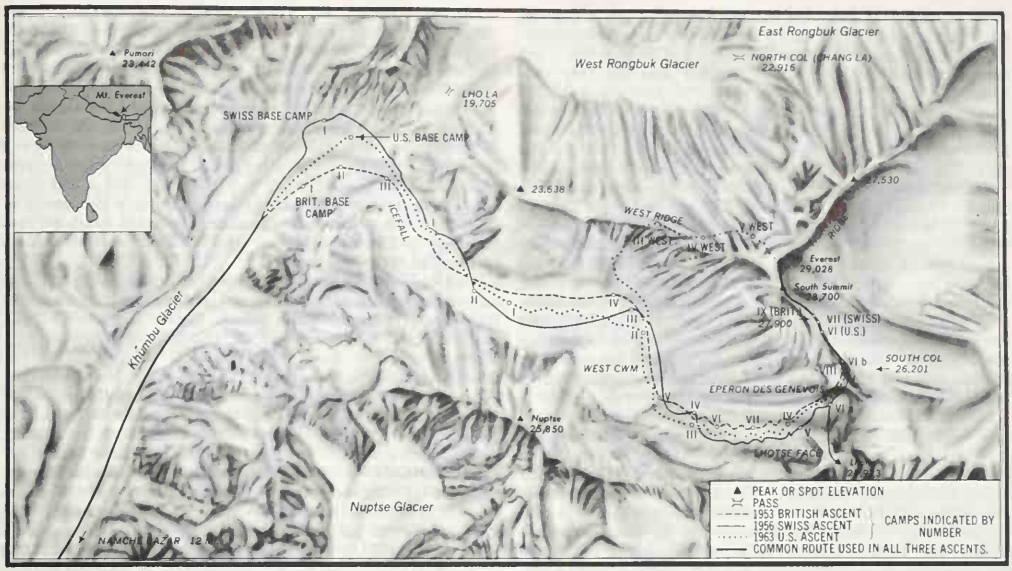
Reconnaissance of 1921.—In 1920 Tibetan permission was first given for a party under Lieut. Col. C. K. Howard-Bury to explore the whole range and find a route up Everest. The other members were G. H. Bullock, A. M. Kellas, G. H. L. Mallory, H. Raeburn, A. F. R. Wollaston, majors H. T. Morshead and O. E. Wheeler (surveyors) and A. M. Heron (geologist).

During the summer of 1921 the northern approaches to the mountain were thoroughly explored. On the approach march Kellas died of heart failure. Because Raeburn fell ill, the high exploration devolved almost entirely upon Mallory and Bullock. Neither had had Himalayan experience and they were faced with the problem of acclimatization besides the difficulties of terrain.

The first object was to explore the Rongbuk valley. The party ascended the main or West Rongbuk glacier, missing the narrower opening of the eastern branch. Though much exploration was done, they had missed the possible line up Everest. They returned eastward for a rest at Kharta Shekar. From there was discovered a pass, the Lhakpa, 22,200 ft., leading to the head of the East Rongbuk glacier. The saddle north of Everest, despite its forbidding appearance, was climbed on Sept. 24 by Mallory, Bullock and Wheeler and named the North col. A bitter wind prevented them from going higher, but Mallory had from there traced a route to the summit.

Attempt of 1922.—Members of the expedition were Brig. Gen. C. G. Bruce (leader), Capt. J. G. Bruce (transport), C. G. Crawford (transport), G. I. Finch, T. G. Longstaff, Mallory, Capt. C. J. Morris (transport), Major Morshead, Maj. E. F. Norton, T. H. Somervell, Col. E. L. Strutt, A. W. Wakefield and Capt. J. B. L. Noel (photographer). It was decided that the mountain must be attempted before the onset of the southwest monsoon, which deposits heavy snowfalls usually in early June. There is a brief calm period before the monsoon's arrival. In the spring, therefore, the baggage was carried across the high, windy plateau of Tibet. The men employed for high carrying by this and subsequent expeditions were chiefly Sherpas, originally inhabitants of the southern slopes of Everest, though many have migrated to Darjeeling.

Camps were carried from the base at 16,500 ft. to an advanced base at camp III. From there, on May 13, a camp was established on the North col. With great difficulty a higher camp, carried up by picked Sherpas, was set at 25,000 ft. on the sheltered side of the north ridge. Next morning, the 21st, Mallory, Norton and Somervell left Morshead, who was suffering from frostbite, and pushed on through trying windy conditions to 27,000 ft. near the crest of the northeast ridge. On May 25 Finch and Captain Bruce set out from camp III using oxygen. Finch, the protagonist of oxygen, was justified by the results. The party, with the Gurkha Tejbir Bura, established camp V at 25,500 ft. There they were



ROUTES OF THE THREE SUCCESSFUL EXPEDITIONS TO THE TOP OF MT. EVEREST

stormbound for a day and two nights, but the next morning Finch and Bruce reached 27,300 ft. and returned the same day to camp III. A third attempt during the early monsoon snow ended in disaster. On June 7 Mallory, Crawford and Somervell, with 14 Sherpas, were crossing the North col slopes. Nine Sherpas were swept by an avalanche over an ice cliff and seven were killed. Mallory's party was carried down 150 ft. but not injured.

Attempt of 1924.—Members of the expedition were Brigadier General Bruce (leader), Bentley Beetham, Captain Bruce, J. de V. Hazard, Maj. R. W. G. Hingston, A. C. Irvine, Mallory, Lieutenant Colonel Norton, N. E. Odell, E. O. Shebbeare (transport), Somervell and Captain Noel (photographer). Because of wintry conditions camp IV on the North col was established only on May 22, by a new and steeper though safer route; the party was then forced to descend. General Bruce had to return because of illness, and under Norton on June 1 camp IV was reestablished. At 25,000 ft., Mallory and Captain Bruce were stopped by exhaustion of their Sherpas. On June 4 Norton and Somervell, with three Sherpas, pitched camp VI at 26,800 ft.; next day they reached 28,000 ft. Norton went on to 28,100 ft., a height unsurpassed until 1953. Mallory and Irvine, using oxygen, set out from the North col on June 6. On June 8 they started for the summit. Odell, coming up that morning, believed that he saw them high up between the mists. Next morning he went up to search. On the 10th he revisited camp VI but found nothing. It was long supposed that they might have reached the top, but this seems highly improbable in view of the difficulties of the final pyramid.

Attempt of 1933.—Members of the expedition were Hugh Ruttledge (leader), Capt. E. St. J. Birnie, Lieut. Col. H. Boustead, T. A. Brocklebank, Crawford, C. R. Greene, P. Wyn Harris, J. L. Longland, W. W. McLean, Shebbeare (transport), E. E. Shipton, F. S. Smythe, L. R. Wager, G. Wood-Johnson and Lieut. W. R. Smijth-Windham and Lieut. E. C. Thompson (wireless).

Because of winds, great difficulty was experienced until May 15 in establishing the North col camp. Its occupants were cut off for several days. On the 22nd, however, camp V was placed at 25,700 ft.; again storm set in, retreat was ordered and it was not till the 28th that V was reoccupied. On the 29th Wyn Harris, Wager and Longland pitched camp VI at 27,400 ft. On the way down Longland's party, due to a blizzard, had great difficulty.

On May 30, while Smythe and Shipton came up to camp V, Wyn Harris and Wager set off from camp VI. Sixty feet below the crest of the northeast ridge they found Mallory's ice ax. They reckoned the second rock step impossible and were compelled to follow Norton's 1924 traverse to the great couloir (steep gorge) splitting the face below the summit. They crossed the couloir to a height about the same as Norton's but then had to return. Smythe and Shipton made their final attempt on June 1. Shipton, with severe digestive trouble, returned soon and regained camp V. Smythe pushed on alone, crossed the couloir and reached the same height as Wyn Harris and Wager. On his return the monsoon ended operations. During the summer of 1933 the Houston flight over Mt. Everest took place. In 1934 Maurice Wilson died above camp III on his futile attempt to climb Everest alone.

Reconnaissance of 1935.—In 1935 an expedition led by Shipton was sent to reconnoitre the mountain, to explore the western approaches and to discover more about monsoon conditions. Other members were: L. V. Bryant, E. G. H. Kempson, M. Spender (surveyor), H. W. Tilman, C. Warren and E. H. L. Wigram. In late July the party succeeded in putting a camp on the North col, but dangerous avalanche conditions turned them off the mountain, and they explored and climbed other peaks. One more visit was paid to the North col area in an attempt on the North peak. During the reconnaissance Wilson's body was found and buried; his diary was also recovered.

Attempt of 1936.—Members of the expedition were Ruttledge (leader), J. M. L. Gavin, Wyn Harris, G. N. Humphreys, Kempson, Morris (transport), P. R. Oliver, Shipton, Smijth-Windham (wireless), Smythe, Warren and Wigram. This expedition had the misfortune of an unusually early monsoon. The route up to the North col was finished on May 13, but the wind had dropped and almost immediately after the camp was established heavy falls of snow put the climbing of the upper part of the mountain out of the question. Several attempts were made to regain the col, but the weather had defeated the expedition.

Attempt of 1938.—Members of the expedition were Tilman (leader), P. Lloyd, Odell, Oliver, Shipton, Smythe and Warren. Unlike the two previous parties, some members of this expedition used oxygen. The party arrived early, in view of the experience of 1936, but this year they were too early and had to withdraw, meeting again at camp III on May 20. The North col camp was pitched under snowy conditions on May 24. Shortly after, be-

cause of dangerous snow, the route was changed and a new one made up the west side of the col. On June 6 camp V was established. On the 8th, in deep snow, Shipton and Smythe with seven Sherpas pitched camp VI, 27,200 ft., but next day they were stopped above it by deep powder snow. The same fate befell Tilman and Lloyd, who made their attempt on the 11th. Lloyd benefited from the open-circuit oxygen apparatus. (With the open-circuit equipment the climber breathes partly from the outside air.) The heavy snow and bad weather compelled a final retreat.

Reconnaissance of 1951.—Members of the expedition were Shipton (leader), T. D. Bourdillon, Edmund (later Sir Edmund Percival) Hillary (*q.v.*), W. H. Murray, H. E. Riddiford and M. P. Ward. After World War II political reasons precluded an approach through Tibet. In 1951 permission was received from the Nepalese for a reconnaissance of the mountain from the south. The party marched through the monsoon, reaching Namche Bazar, chief village of Sola Khumbu, on Sept. 22. At Khumbu glacier they found it possible to scale the great icefall seen by Mallory from the west. They were stopped at the top by a huge crevasse, but traced a possible line up the Western cwm (cirque) to the South col, the high saddle between Lhotse and Everest.

Spring Attempt of 1952.—Members of the expedition were E. Wyss Dunant (leader), J. J. Asper, R. Aubert, G. Chevalley, R. Dittert (leader of climbing party), L. Flory, E. Hofstetter, P. C. Bonnant, R. Lambert and A. Roch; A. Lombard (geologist) and A. Zimmermann (botanist). This strong Swiss party first set foot on the icefall on April 26. After considerable difficulty with the route, they overcame the final crevasse by means of a rope bridge. The 4,000-ft. face of Lhotse, which must be climbed to reach the South col, was attempted by a route running beside a long spur of rock christened the Éperon des Genevois. The first party, Lambert, Flory, Aubert and Tenzing Norgay (sirdar) with five Sherpas tried to reach the col in one day. They were compelled to bivouac quite a distance below it (May 25) and next day reached the summit of the Éperon, 26,300 ft., whence they descended to the col and pitched camp. On the 27th the party (less the five Sherpas) climbed up the southeast ridge. They reached approximately 27,200 ft., and there Lambert and Tenzing bivouacked. Next day they pushed on up the ridge and turned back at approximately 28,000 ft. On May 29 Asper, Chevalley, Dittert, Hofstetter and Roch reached the South col. They were prevented by wind conditions from going higher, and after three most uncomfortable nights descended to camp V, thence to the base.

Autumn Attempt of 1952.—Members of the expedition were Chevalley (leader), J. Buzio, G. Gross, Lambert, E. Reiss and A. Spöhel; and N. Dyhrenfurth (photographer). The party found the icefall easier than in the spring and had brought poles to bridge the great crevasse. Camp IV was occupied on Oct. 20. Higher up, however, they were constantly harassed by bitterly cold winds. On the ice slope below the Éperon one Sherpa was killed, and the party took to the glaciated face of Lhotse on the right. The South col was reached on Nov. 19, but the summit party climbed only 300 ft. higher before being forced to withdraw.

The Ascent of 1953.—Expedition members sponsored by the Royal Geographical society and the Alpine club were Col. John Hunt (leader; *see* HUNT, SIR [HENRY CECIL] JOHN), G. C. Band, Bourdillon, R. C. Evans, A. Gregory, Hillary, W. G. Lowe, C. W. F. Noyce, Ward, M. H. Westmacott, Maj. C. G. Wylie (transport), T. Stobart (cinematographer) and L. G. C. Pugh (physiologist). After three weeks' training on neighbouring mountains, a route was worked out up the icefall and it was possible to start ferrying loads to the Western cwm head. Two forms of oxygen apparatus, the closed- and open-circuit types, were tried. As a result of a reconnaissance of Lhotse in early May, Hunt decided that Bourdillon and Evans, experts on the closed-circuit, should make the first attempt from the South col. Hillary with Tenzing as sirdar were to follow, using open-circuit and a higher camp.

Lowe spent nine days, most of them with the Sherpa Ang Nyima, working at the lower section of the Lhotse face. On May 17 a camp was pitched on it at 24,000 ft. The route on the upper part of the face, over the top of the Éperon, was first made by Noyce and the Sherpa Annullu on May 21. Next day 13 Sherpas led by Wylie, with Hillary and Tenzing ahead, reached the col and dumped loads. The fine weather continued from May 14 but with high winds. On May 24 the first summit party, with Hunt and two Sherpas in support, reached the col. On the 26th Evans and Bourdillon climbed to the South summit of Everest, 28,700 ft., but were too late to go farther. Meanwhile Hunt and the Sherpa Da Namgyal left loads for the ridge camp at 27,350 ft. On the 28th the ridge camp was established at 27,900 ft. by Lowe, Gregory and Ang Nyima. Hillary and Tenzing passed the night there and next day reached the South summit at 9 A.M. Two and a half hours later they had traversed the final ridge and stood on the summit of Everest. They were met on the slopes above the South col that afternoon by Lowe and Noyce. By June 2 the whole expedition reassembled at the base camp. (C. W. F. N.; H. C. J. H.)

Everest-Lhotse, 1956.—The Swiss performed the remarkable feat of getting two ropes up Everest and one up Lhotse (27,923 ft.), using oxygen. Members were A. Eggler (leader), W. Diehl, H. Grimm, H. R. von Gunten, E. Leuthold; F. Luchsinger, J. Marmet, F. Müller, Reiss, A. Reist and E. Schmied. They followed roughly the British route up the icefall and the Lhotse face. From their camp VI Reiss and Luchsinger reached the summit of Lhotse on May 18. Camp VI was moved to the South col and the summit of Everest reached from a camp at 27,500 ft. by Marmet and Schmied (May 23) and Gunten and Reist (May 24).

Attempts of 1960.—In 1960 an Indian expedition with Sherpas, led by Brig. Gyan Singh, attempted Everest from the south. Camp IV was established in the Western cwm on April 19. Bad weather followed, but a party using oxygen reached the South col on May 9. On May 24 three members pitched a tent at 27,000 ft. on the southeast ridge but were turned back by wind and weather at about 28,300 ft. Continued bad weather prevented the second summit party's leaving the South col.

On May 25 a Chinese expedition led by Shih Chan-chun was said to have climbed Everest from the north. By their account they reached the North col in April and on May 24 Wang Fu-chou, Chu Yin-mua, Liu Lien-man and a Tibetan mountaineer, Konbu, climbed the slab by a human ladder reaching the top at 4:20 A.M. to deposit the Chinese flag and a bust of Mao Tse-tung. The credibility of their account was doubted. (C. W. F. N.)

The U.S. Ascent of 1963.—This, the first American expedition, was led by Norman G. Dyhrenfurth, who selected a team of 19 mountaineers and scientists from throughout the United States and 37 Sherpas for high altitude work. The purpose was twofold; to reach the summit of Mt. Everest, and to carry out scientific research programs in physiology, psychology, glaciology and meteorology. Of particular interest were the studies on how the climbers changed physiologically and psychologically under extreme stresses at high altitudes where oxygen starvation was unavoidable. These studies were related to the U.S. efforts in space, and among the 400 sponsors of the expedition were the National Geographic society, state department, National Science foundation, Office of Naval Research, National Aeronautics and Space administration, U.S. army quartermaster corps, Atomic Energy commission and U.S. air force division of scientific research.

On Feb. 20, the expedition left Katmandu for Mt. Everest 180 mi. away. More than 900 porters carried 53,000 lb. of food, clothing, equipment and scientific instruments. Base camp was established on March 20, at 17,800 ft. on the Khumbu glacier, one month earlier than any previous expedition. For the next five weeks the team selected a route toward the summit and established and stocked a series of camps up the mountain via the traditional South col route. They also explored the more difficult and untried West ridge route. On May 1, James W. Whittaker and Nawang Gombu, Sherpa climber and nephew of Tenzing Norgay (Norkay), reached the summit despite very high winds. On May 22, four other Americans reached the top. Two of them, William F. Unsoeld and Thomas F. Hornbein made mountaineering history by ascending the West ridge which had been considered "unclimbable." They descended by the traditional Southeast ridge toward the South col, thus making the first major mountain traverse in the Himalayas. On the descent, Unsoeld and Hornbein, along with Barry C. Bishop and Luther G. Jerstad, who also reached

the summit that day via the South col, were forced to bivouac in the open at 28,000 ft. All suffered frostbite during the night and Bishop and Unsoeld later lost their toes. On July 8, upon their return to the U.S., Dyhrenfurth and all members of the expedition were presented the National Geographic society's Hubbard medal by Pres. John F. Kennedy. *See* MOUNTAINEERING; *see* also references under "Everest, Mount" in the Index. (B. C. B.)

BIBLIOGRAPHY.—C. K. Howard-Bury, *Mt. Everest: the Reconnaissance, 1921* (1922); C. G. Bruce, *The Assault on Mount Everest, 1922* (1923); E. F. Norton *et al., The Fight for Everest: 1924* (1925); T. H. Somervell, *After Everest* (1936); Sir F. E. Younghusband, *The Epic of Mount Everest* (1926); H. Ruttledge, *Everest 1933* (1934), *Everest: the Unfinished Adventure* (1937); F. S. Smythe, *Camp Six* (1937); H. W. Tilman, *Mount Everest, 1938* (1948); E. Shipton, *The Mount Everest Reconnaissance Expedition, 1951* (1952); A. Roch *et al., Everest 1952* (1953); *Berge der Welt*, vol. 8 (1953); Sir John Hunt, *The Ascent of Everest* (1953; published in the U.S. as *The Conquest of Everest*, 1954); W. Noyce, *South Col* (1954); Sir Edmund Hillary, *High Adventure* (1955); A. Eggler, *The Everest-Lhotse Adventure* (1957); *National Geographic Magazine* (Oct. 1963); James Ramsey Ullman, *Americans On Everest* (1964).

EVERETT, ALEXANDER HILL (1790–1847), U.S. author, diplomat and editor of the *North American Review*. He was born in Boston, Mass., on March 19, 1790, the son of Oliver Everett (1753–1802), a Congregational minister in Boston, and the brother of Edward Everett. He graduated at Harvard in 1806, taking the highest honours of his year, though the youngest member of his class, and in 1807 began the study of law in the office of John Quincy Adams. In 1809 Adams was appointed minister to Russia, and Everett accompanied him as his private secretary, remaining attached to the American legation until 1811. From 1825 to 1829, during the presidency of John Quincy Adams, he was the U.S. minister to Spain. Everett was a member of the Massachusetts legislature in 1830–35, was president of Jefferson college in Louisiana in 1842–44, and was appointed commissioner to China in 1845. He died on May 29, 1847, at Canton, China.

Everett is known rather as a man of letters than as a diplomat. In addition to numerous articles, published chiefly in the *North American Review*, of which he was the editor from 1829 to 1835, he wrote: *Europe, or a General Survey of the Political Situation of the Principal Powers, With Conjectures on their Future Prospects* (1822), which was translated into German, French and Spanish; *New Ideas on Population* (1822); *America, or a General Survey of the Political Situation of the Several Powers of the Western Continent, With Conjectures on their Future Prospects* (1827); a volume of *Poems* (1845); and *Critical and Miscellaneous Essays* (first series, 1845; second series, 1847).

EVERETT, EDWARD (1794–1865), U.S. clergyman, educator, statesman and orator whose most famous oration was delivered at the dedication of the national cemetery at Gettysburg, Pa., on Nov. 19, 1863, immediately preceding Lincoln's Gettysburg address. He was born in Dorchester, Mass., on April 11, 1794. At 17 he graduated from Harvard college, taking first honours in his class. After studies in divinity he became, at the age of 19, the minister of the Brattle Street Unitarian church in Boston but resigned after little more than a year to accept a professorship of Greek literature in Harvard—a post he assumed in 1819 following several years of study abroad. From 1820 to 1824, while continuing to teach, he served as editor of the *North American Review*.

Elected to the national house of representatives in 1824, Everett resigned his professorship and thereafter his career was largely political. As congressman (1825–35) he supported generally the administration of John Quincy Adams but opposed that of Andrew Jackson.

In 1835 Everett was elected governor of Massachusetts, an office which he filled for four one-year terms before being defeated for re-election in 1839. Among the measures which received his support were the establishment of a state board of education and a system of normal schools. Everett subsequently served as United States minister to Great Britain (1841–45), president of Harvard college (1846–49), secretary of state (1852–53) and United States senator (1853–54). Confronted with bitter criticism for his failure to stand firm against the proslavery forces in the senate, he resigned his seat in May 1854. In 1860 he was the candidate of the Constitutional Union party for the vice-presidency on the ticket with John Bell, receiving 39 electoral votes.

Everett's chief distinction was not as a statesman but as a lecturer and orator, a field in which by 1820 he had established a reputation such as few men have enjoyed. To the end of his life he was constantly busy on the public platform. Careful preparation, an extraordinary memory and brilliance of style and delivery all combined to make him a superb speaker. He prepared an oration on the character of George Washington which he delivered in every part of the country, raising nearly $70,000 which he contributed to the purchase and preservation of Mount Vernon as a national shrine. During the Civil War he zealously supported the national government.

Everett died in Boston on Jan. 15, 1865.

See Foster Stearns, "Edward Everett" in *American Secretaries of State and Their Diplomacy*, vol. vi, pp. 117–141 (1927–29). (B. DY.)

EVERETT, a city of Middlesex county, Mass., U.S., just north of Boston, from which it is separated by the Mystic river. Originally a part of the town of Malden, Everett became a separate town in 1870 and was incorporated as a city in 1892. It was named in honour of the orator and statesman Edward Everett. It is a highly diversified industrial city, and its port facilities can accommodate ocean-going ships.

Manufacturing, retail trade and construction are important sources of employment. Its most important manufactures are pig iron, transportation equipment, chemicals and allied products, petroleum and coke, fabricated metals, shoes, machinery and paper products. Everett in the 1960s had the only oil refinery and the only blast furnace in New England. For comparative population figures *see* table in MASSACHUSETTS: *Population*. (L. G. H.)

EVERETT, a city of western Washington, U.S., 28 mi. N. of Seattle, on Puget sound, at the mouth of the Snohomish river, a port of entry and seat of Snohomish county. Facing the city is Whidbey Island, beyond which rise the Olympic mountains. The eastern horizon is bounded by the Cascades, with Mt. Baker and Mt. Rainier standing out to the north and south. The landlocked harbour has a depth of from 24 to 40 ft. at the wharves. Shipments consist mostly of lumber, logs, cedar poles, laths, shingles and paper. The city has many lumber, shingle and woodworking mills, large paper and pulp mills, several ironworks and steelworks, creosoting plants, factories making machinery for sawmills and shingle mills, and salmon, fruit and vegetable canneries. The city has a municipal junior college, established in 1941.

Just west of Everett is the spot where Capt. George Vancouver landed to claim New Georgia in 1792 while on his great voyage of exploration. Everett was settled in 1891 and incorporated in 1893, the year of completion of the Great Northern railroad, which first reached tidewater at the new city. It was named for the son of C. L. Colby, chief proprietor of the Wisconsin Central railroad and one of several eastern capitalists interested in the Everett Land Co., the townsite developer. For comparative population figures *see* table in WASHINGTON: *Population*. (R. E. BU.)

EVERGLADES, a marsh and prairielike region of southern Florida, U.S., comprises that portion of the state which almost encircles Lake Okeechobee (*q.v.*) in a narrow strip of land and extends 100 mi. S. and easterly from the shores of the lake at 15 ft. above sea level. As a shallow slough or inundated plain 40 mi. wide, bounded on the east by a coastal fringe of sand dunes and on the west by the Ocaloacoochee slough and the Big Cypress swamp, it stretches to the tidewater mangrove swamps at the southern tip of the state. It occupies an area of approximately 2,500,000 ac. without many trees and dominated by saw grass marshes that are nearly level and which are flooded or wet to their surface most of the year. The Everglades rests on a foundation of limestone rock, with a gradient of one-tenth foot per mile, upon which are layers of sand, alluvial deposits and decayed vegetable matter. The deposits have formed muck or peat to a thickness of 8 ft. in the vicinity of Lake Okeechobee, but the soil thins out in all directions toward the borders. These organic soils are the result of slow vegetative decay under the combination of low elevation, warm climate and heavy rainfall. The Ever-

glades forms a unit of a large watershed covering the interior of the state from Orlando to the Gulf of Mexico. Prior to the installation of drainage works, they were usually submerged from the tidal waves or overflow from Lake Okeechobee because of their level character, the rocky understrata and uneven rainfall.

Private attempts to drain the Everglades prior to 1900 in order to utilize the fertile organic soils for agriculture were unsuccessful. From 1905 to 1925, state and local efforts at drainage opened a few thousand acres to cultivation, but a devastating hurricane in 1928 brought over 1,800 deaths and much destruction of property. From 1930 to 1940, the federal government was engaged in building protective dikes and other works to prevent a second hurricane disaster. By 1947, over 100,000 acres in the upper Everglades or lakeshore region were in agricultural production, but the overall reclamation effort proved to be an enormous undertaking fraught with many difficulties. The gravity drainage of surplus waters neither relieved flooding in wet weather nor provided water for irrigation in periods of extended drought. Fires burned in the muck soils in 1939, 1943 and 1962, yet heavy rains did much damage in 1940 and 1947.

Beginning in 1947, the federal government and the state of Florida inaugurated a major co-operative engineering project, the Central and Southern Florida Flood Control district, to administer a program of flood control, drainage and water control. Under this program reclaimed lands became among the most fertile for the production of sugar cane and winter vegetables. (J. E. D.)

Everglades National Park.—As early as 1915, Florida set aside a part of the southern region as a refuge and retreat to preserve the flora and fauna of the area and in 1947 Everglades National park was established to preserve 1,499,428 ac. of exceptional biological interest. The park, lying partly within the tropical zone, has one of the most nearly level landscapes in the world. Southward and westward across its hummock-dotted plains, clear fresh-water drains through several mangrove-jungle rivers to empty into the Gulf of Mexico. Shoal Florida bay, with its islands, lies within the sanctuary, as do the Ten Thousand Islands on the park's west coast.

Eight species of palms range from the diminutive saw palmetto to the tall royal. Other notable trees are the gumbo limbo, strangler fig, bald cypress and Madeira mahogany, the limbs and trunks of which are adorned with ferns and air plants, including orchids. Alligators inhabit sloughs, and a large mammal, the manatee (*q.v.*), frequents salt-water coves and brackish rivers. Land mammals include the cougar, bobcat, black bear, otter, opossum and raccoon, while snakes include the king, tree, coral, water moccasin and rattlesnake.

The park's avian life is perhaps its best-known feature. Among the birds are herons, egrets, bitterns, ibises, frigate birds, gallinules, eagles, kites, anhingas and the superbly beautiful roseate spoonbill. Early in the 20th century, plume hunters catering to the millinery trade nearly wiped out the egrets and spoonbills. Egrets were slaughtered during their breeding season—the time of year when they carried their nuptial plumes, desired for trimming women's hats—resulting in the starvation and death of nestlings. Aroused public opinion stopped the sale of the feathers, and under national park service protection these birds began to come back. (Dx. B.)

EVERGREEN, a general term applied to plants that retain their leaves through the winter season and into the following summer or through several years; in contrast to the deciduous trees, which shed their leaves in the autumn and stand bare during the winter. In the tropics many broad-leaved angiosperms are evergreen, but in cold temperate and arctic areas the evergreens are chiefly coniferous shrubs or trees, such as pines, firs, junipers, hemlocks and spruces. The leaves of evergreens usually are more leathery and thicker than those of deciduous trees, and among the conifers are often needlelike or mere scales, thus exposing less surface to cold, dry winds. Each year some of the oldest leaves, those farthest back from the tips of the twigs, die and fall, while new ones are produced at the end of each branch. In some evergreens, such as the privet, a crop of leaves lasts only one year, those that have passed through a winter falling in the summer after new leaves appear. Others, such as some oaks and many conifers, retain each crop of leaves two to five years; and a few, *e.g.*, the bristle-cone pine (*Pinus aristata*), may retain functional leaves 10 to 18 years. Examples of broad-leaved, angiospermous evergreens are the live oak, English holly (*Ilex*), southern magnolia, madrone (*Arbutus menziesi*), California laurel (*Umbellularia californica*) and a number of shrubs. Evergreens are used extensively in landscape architecture (*q.v.*). *See* also ARBORICULTURE; CONIFERS; TREE; and articles on various evergreens.

See L. H. Bailey, *Standard Cyclopedia of Horticulture* (1943). (I. L. W.)

EVERLASTING, any plant that, because of its retention of form and colour when dried, finds use in dry bouquets and flower arrangements. Among the popular everlastings are several species of Compositae (*q.v.*), especially the true everlastings or immortelles, *Helichrysum* species. *Helichrysum* is native to north Africa, Crete and the parts of Asia bordering on the Mediterranean, and is cultivated in many parts of Europe. The immortelles possess a large involucre of dry, scalelike or scarious bracts that preserve their appearance when dried, provided the plant is gathered in proper condition. They require a light, sandy or stony soil, and are readily injured by rain or heavy dews. They can be propagated by offsets from the older stems. *Helichrysum* is tufted in its growth, each plant producing 60 or 70 stems, and each stem producing an average of 20 flowers. Partially expanded flowers are preferred; all fully expanded and immature blossoms are rejected. The colour of the bracts is deep yellow, but the flower heads may be dyed a variety of shades. One of the best-known everlastings is the Australian species *H. bracteatum*, the strawflower. *H. vestitum* (Cape of Good Hope) has white satiny heads.

Other plants grown as everlastings include species of *Helipterum* (west Australia and South Africa), *Ammobium* (Australia), *Achyrachaena* (western United States), *Antennaria* (extratropical except Africa), *Gnaphalium* (cosmopolitan) and *Xeranthemum* (south Europe). In North America, the pearly everlasting (*Anaphalis margaritacea*) is very widely distributed, occurring in dry soils from Newfoundland to Alaska and south to North Carolina and California. Several members of the family Amaranthaceae are considered everlastings: such are *Gomphrena globosa*, the globe amaranth, with oval heads of white, orange, rose or violet, and scarious bracts; and *Celosia argentea*, with its elegant, loose pyramidal inflorescences.

In addition to these, many grasses with showy panicles or spikes are classed as everlastings. The dried fruit of the ground cherry (*Physalis*) and honesty (*Lunaria*) are often used with everlastings in winter bouquets.

EVESHAM, a market town and municipal borough in Worcestershire, Eng., 15½ mi. S.E. of Worcester by road. Pop. (1961) 12,109. On the right (north) bank of the Avon, in the rich and beautiful Vale of Evesham, the district, mainly agricultural, is devoted to market gardening and orchards. Jam making, fruit and vegetable canning are allied industries and the making of jewel boxes is also carried on. A Benedictine house was founded at Evesham by St. Egwin in the 8th century. It became a wealthy abbey, but was almost wholly destroyed at the Dissolution. The churchyard is entered by a Norman gateway, and there survives an isolated bell tower (1533) of magnificent Perpendicular workmanship. A single Decorated arch is almost all that remains of the abbey walls, but near the tower are the parish churches of St. Lawrence (16th century) and All Saints, the latter containing the chapel of Abbot Lichfield, who built the tower. Other buildings include an Elizabethan town hall, the grammar school, founded by Abbot Lichfield, and the picturesque almonry. Evesham (Eof's Hamm or meadow) grew up around the abbey, and in 1055 Edward the Confessor gave it a market and the privileges of a commercial town. It is mentioned in Domesday Book but it is uncertain when the town first became a borough. Before 1482 the abbot practically had control of the town, but after the Dissolution the manor, markets, fairs and other privileges were granted to Sir Philip Hoby. In 1604 James 1 granted the burgesses their first charter and in the following year, by a second charter, he incorporated Evesham with the parish of Bengeworth; this remains the governing charter of the

borough. In 1933 the parish of Great and Little Hampton was added. Evesham returned two members to parliament in 1295 and again in 1337, after which date the privilege lapsed until 1604. Its two members were reduced to one in 1867 and the borough was disfranchised in 1885. (X.)

The Battle of Evesham (1265).—The battle of Evesham was fought on Aug. 4, 1265, between Simon de Montfort, earl of Leicester, and Henry III's son Edward (afterward Edward I). Montfort's victory at Lewes (*q.v.*) left him master of England, but the barons of the middle marches of Wales refused submission, and royalist garrisons held Pevensey, Bristol and other scattered castles. In 1264, with Welsh help, Montfort and his chief ally, Gilbert de Clare, earl of Gloucester, had twice forced the marcher lords to terms, but they resumed defiance when Montfort returned to England. Early in 1265 Clare, growing discontented, deserted Montfort and withdrew to Glamorgan. In April Montfort with a small army moved westward, taking Henry and Edward with him, hoping to conciliate Clare and then to crush the marchers, but Clare, secretly leagued with the marchers, rejected Montfort's summons. On May 28 Edward escaped at Hereford, joined Clare, Roger Mortimer and the other marchers, and with forces outnumbering Montfort's quickly seized the Severn crossings, cutting off Montfort's retreat. Meanwhile other royalists from France, under William de Valence, landed in Pembrokeshire and joined Edward. Montfort, belatedly realizing his danger, made a treaty with Llewelyn at Gruffydd, and urgently summoned his son Simon to leave the siege of Pevensey and bring all available forces to the rescue. Young Simon, however, moved too late and too slowly. Montfort pushed rapidly down the Wye, hoping to escape across the Severn estuary from Newport in ships brought from Bristol; but while Edward closed the road at Gloucester, Clare, using three royal galleys under his command, prevented the transports from sailing. Edward fiercely attacked Newport, but Montfort's army escaped westward under Welsh protection, returning exhausted to Hereford late in July after toilsome and hungry detours through the mountains. Young Simon's relieving army, making for Montfort's impregnable castle at Kenilworth, was drawing near, so Edward, adopting bold defensive strategy to keep the two armies apart, thrust himself between them at Worcester to guard the Severn crossing. Learning that young Simon's army, confident of numbers and distance, was encamped carelessly at Kenilworth, Edward left Worcester late on July 31 and, marching nearly 40 mi. by night, fell on young Simon at dawn, killing or capturing most of his unresisting army. Returning immediately to Worcester, Edward found that Montfort had already crossed the Severn at Kempsey and, ignorant of his son's defeat, was slowly moving toward Kenilworth. Late on Aug. 3 Edward left Worcester to intercept Montfort at Evesham next morning. Detaching Mortimer to block any eastward escape from Evesham across the Avon, Edward closed in from the north, displaying Montfortian banners captured at Kenilworth. Montfort, momentarily deceived into thinking it his son's army, soon realized that he was hopelessly trapped. Hastily drawing up his little force on the highest point of the ridge north of the town, he awaited Edward's attack. His Welsh auxiliaries fled, and he was quickly encircled and overwhelmed by far superior numbers. He and his friends died fighting. King Henry himself, held captive to the end, was wounded and nearly slain by the royalist forces, who disgraced themselves by foully mutilating Montfort's body. (R. F. T.)

ÉVIAN-LES-BAINS, a spa and tourist resort of eastern France, *département* of Haute-Savoie, lies on the southern shore of Lake Geneva, 81 km. (50 mi.) N.E. of Annecy by road. Pop. (1962) 4,562, usually augmented in summer by about 28,000 visitors. Facing Lausanne on the northern shore of the lake, Évian lies below the lowest Alpine spurs and has a mild climate suitable for the many people who come to take its waters, which contain carbon and calcium. The spa buildings, the new *hôtel de ville* (town hall) and the casino form the centre of the town. There is a lakeside promenade. The 13th-century parish church, the old *hôtel de ville* and the medieval fortress towers are in the old quarter. There are many big hotels and an imposing congress hall. The care of invalids and tourists is the inhabitants' main occupation. Called variously Aquianum, Vian, Les Vians and Eyvians, the town finally became Évian in the 19th century. First a lake dwelling, the present site has been inhabited since classical days. In the 11th century it became part of Savoy and later the residence of its rulers. Under Victor Amadeus II (1675–1730; duke of Savoy and later king of Sardinia) the waters of Évian began to achieve fame. The first spa buildings were opened in 1839. In 1860 Évian passed to France, and in World War II it was occupied by the Germans from Nov. 1942 until Aug. 1944.

(Y. C.)

EVICTION, the process of dispossessing a person of land, may be lawful or unlawful. Subject to any statutory provisions, it is lawful if the person evicted has a right to possession inferior to that of the person carrying out the eviction. The delivery of possession under order of the court is sometimes called eviction. *See* EJECTMENT; LANDLORD AND TENANT. (R. E. My.)

EVIDENCE, in a broad sense, consists of items of fact or assertions of fact which tend to persuade the mind that a given factual proposition is true or false. In courts of law and other tribunals the term is used more narrowly to designate facts which meet the requirements of the legal rules governing the proof of facts. It is with this body of rules, referred to as the law of evidence, that the present article is concerned. (For the nature of evidence in logic and science, *see* INDUCTION; SCIENTIFIC METHOD.) Those facts which meet the requirements of the law of evidence may consist of oral testimony of witnesses, of physical objects and of documents. They are also sometimes referred to as proof or proofs. The process of presenting the evidence to the tribunal is referred to as proving the case, although with individual items the phrase "to introduce evidence or proof" is more commonly used in the United States.

The present article is divided into the following main sections:

I. History
II. Presentment of Evidence
III. Witnesses
IV. Proof of Facts
V. Privilege and Public Policy
VI. Writings
VII. Burden of Proof and Presumptions: Quantum of Evidence
VIII. Application of Evidence Law

I. HISTORY

Prevailing methods of resolving disputed issues of fact in the Anglo-American world vary in detail but have one predominant characteristic in common: namely that disputed issues of fact are resolved by persuading some human agency that the fact asserted either does or does not, did or did not, will or will not exist. Historically, this was not the pattern. In pre-Norman England the object of litigation was to frame issues of fact in such a way that they could be submitted to the decision of a supernatural agency. One form of submission was ordeal. The person asserting the fact was required to endure submersion in water, burning by a hot iron or the like. If he survived these tests without more than an arbitrarily fixed harm, he won the dispute, for it was assumed that he would survive only if assisted by the Deity, who would aid only the cause of the just. Another form of ordeal was by battle between contestants, which, technically, survived into the 19th century. As a matter of fact, these methods went into decline about the time of the Conquest, although another less barbarous form of appeal to the supernatural continued in use through the 16th century. This was compurgation, or trial by wager of law. In it the man charged with asserting a fact swore to its existence and procured men called oath helpers, who reinforced his oath with their own solemn assertions that it was clean and true; they did not testify about the fact itself and, indeed, might have no personal knowledge concerning it. (*See* also ORDEAL.)

Beginning with the 13th century, the procedure of trial by jury became the common method of trial. This early jury was not the institution known today, where a requisite for selection is ignorance of the facts in issue. The characteristic of the early juror was that he had personal knowledge of the disputed facts and could pass upon them without receiving further evidence. As the

transition from informed juror to uninformed juror occurred, the procedure of submitting facts to jurors in the form of evidence or proof swallowed up the class of cases tried by compurgation. This development occurred by the middle of the 17th century. (*See* JURY.)

Although certain modern rules of evidence may be traced to earlier methods of resolving disputes of fact—primarily compurgation—the great body of evidence law evolved in response to the newer practice of submitting issues of fact to jurors who had no personal knowledge of the facts in dispute.

A companion influence on this evolution was the conception of the judging function which characterizes English law. Simply stated, it is that the role of the tribunal is primarily passive—to hear what the litigants present by way of evidence rather than to embark upon an investigation in which the initiative resides in the tribunal itself.

The relative contribution of these two sources of the rules of evidence is not easy to determine. Most scholars concede the operative force of each, yet in any given instance they may differ about which one explains the case in question. In recent years there has been less of such academic dispute, controversies centring upon what the rule should be rather than upon the forces which made it what it is. Still, the contribution of historical research is not to be discounted. Only the fact that it has been so well performed liberates the modern scholar to turn his attention to the concerns of the present.

The features which characterize the Anglo-American law of evidence are largely absent from the procedures of the rest of the world. Probably this is because the civil-law world, which draws its pattern largely from Roman law, makes far less use of the untrained fact finder, the juror, and employs a different conception of the role of the tribunal in deciding questions of fact. Unlike their Anglo-American counterparts, civil-law judges bear at least some responsibility to discover the facts and to bring out all those which bear upon the dispute.

Within the Anglo-American world, there is an amazing homogeneity of principal doctrine. More detailed rules vary from jurisdiction to jurisdiction, however. Within the United Kingdom (apart from Scotland, which mingles civil-law and common-law practices) and the commonwealth countries employing English law, the variances are slight indeed. Differences between rules in the United States and in the English countries are greater. In general, it may be said that many more cases turn upon points of evidence in the United States than in the English-law countries. How many more is difficult to assess, since more than 50 separate U.S. jurisdictions report decisions. Among factors possibly contributing to this greater concern with the subject in the U.S. are: (1) continued widespread use of juries in the United States as contrasted with the decline of the jury in civil cases in England and many of the commonwealth nations; (2) restrictions in the majority of U.S. jurisdictions upon the ancient right of the judge to express his views to the jurors on the meaning of the evidence received, for which the judges compensate by more readily creating rules to exclude evidence; and (3) the lack in the United States of a specialized profession of trial lawyers well versed in the rules governing proof of facts, compared with the retention of the highly-skilled barrister in England and in some other nations in the English-law group.

Despite these reasons and others which may exist, to any but the specialist the uniformity of Anglo-American law far exceeds the variations.

II. PRESENTMENT OF EVIDENCE

The principle that the finder of fact, be it judge or jury, has no or very little active role in obtaining information about the facts in dispute in a given case is productive of some elementary regulations at the outset. Finders of fact are not permitted to take independent steps, such as visiting the scene in question (except on an order of the court) or performing private experiments. Such misconduct, when shown, may void the results of a trial or warrant the setting aside of a verdict or judgment. As a consequence, virtually the entire factual picture involved in a case must in one way or another be produced to enlighten a finder of fact who is, ideally at least, wholly ignorant of the facts when he begins and powerless to inform himself. This has, in the past, been an expensive and time-consuming process which brought courts into a somewhat deserved disrepute. More modern methods of pleading and trial practice have diminished this abuse considerably. Increasingly, jurors are simply told that certain important facts may be accepted as established because the litigants have stipulated that the facts exist or that they are admitted to be true. This development is not properly a part of the law of evidence, but it has had a very substantial impact upon that law in actual operation.

1. Judicial Notice.—A doctrine which has had some effect in circumscribing the area comprehended by the law of evidence is that of judicial notice. This consists of the court's accepting certain facts essential to the case to be true and established and instructing the jury, if one is employed, that those facts are to be accepted by it. The kind of facts susceptible to such treatment were originally only those within the common knowledge of the community in which the trial was held. This had to be distinguished from knowledge held by the judge himself, which could not be employed unless it was shared by the community. Obviously, this category is not very large: events of recent history, major political and geographic boundaries, identity of public officers, the character of a particular locality and the effects of consumption of alcohol on a human are all matters which have commonly been the subject of judicial notice. Another standard of judicial notice has been evolving somewhat slowly. This would admit some facts which are not widely known in the community but which are nevertheless subject to verification in sources of undisputed authority. For example, it would be common knowledge and judicially noticeable that the sun rises each day and that it rises earlier at some times of the year than at others. The precise moment of sunrise would be widely known in few if any communities. But that information would be readily obtainable in standard reference works of unimpeachable reliability, and the evolving standard would permit the judge to ascertain such a fact and instruct the jury regarding its existence. It might be added that the class of facts comprehended by this second principle is constantly expanding in accord with the progress of modern science. But it also includes many propositions of history and geography that fit only uneasily into the older principle of common knowledge.

Although there is no formal doctrine of "jury notice," it is obvious that the function of jurors—evaluating evidence they receive in the light of reason and experience—is performed in the light of much general knowledge they bring to their tasks.

2. Real Evidence.—This is a term employed—more commonly in England than in the United States—to describe tangible physical objects introduced into evidence and displayed for examination by the judge and jury. Common forms of real evidence are the fatal weapon, clothing worn by the accused, samples of articles in question and the like. These are properly regarded as evidence in the case, and the jury may as certainly draw inferences from them as it may from testimony of witnesses. Because of this, they are received with caution. A pistol offered in evidence must be identified by a sworn witness who is able to testify of his own knowledge that it was the murder weapon or to facts which make it reasonably probable that it was. A characteristic of real evidence is that it comes into a courtroom mute; only when properly identified and received into evidence may it be employed as such.

A less clear case concerns physical objects which admittedly are not the fatal weapon, for example, but which resemble the type of instrument which may have inflicted the fatal blow, as deduced from the size and configuration of a wound. Working models of machinery and samples of property similar to that involved in the lawsuit are often employed for the same purpose, as are skeletons and models of the human figure. Such objects are commonly employed for the purpose of illustrating testimony on technical points, although they are seldom actually received into evidence. Standards for their employment are more lax than those applying to real evidence; substantial similarity with respect to important features is all that is required.

Photographs, both still and motion, are increasingly employed both as real evidence of the facts portrayed, when they are subject to the same careful rules about authentication as other real evidence, and as sources for illustrating or understanding oral testimony of witnesses, when they are subject to the lesser requirement of substantial similarity.

In both U.S. law and that of most English-law countries, two items have been the occasion of considerable dispute. In England and the United States it is proper for the jury or judge to be taken to the scene involved in the case or to inspect some object too large to be brought into the courtroom. The English courts and a majority of the U.S. courts insist that what is seen on this "view" is not evidence but merely to be used for illustration or understanding of testimony properly received. A few U.S. jurisdictions have finally abandoned this distinction. A subject of equal uncertainty is whether a child may be presented to the jury or judge for inspection to determine whether it bears resemblance to one alleged to be its father or other blood relative. A clear majority of English and U.S. rulings permit this practice, although some judges exhibit considerable uncertainty over its value and not infrequently assert that it is not "evidence."

3. Documents.—Documentary evidence is substantially similar to real evidence in that it consists of tangible physical articles. But an important difference has led to some highly specialized rules pertaining to documents. (Electronic voice recordings serve much the same purpose as documents in many cases and are, on the whole, subject to the same rules of admission and exclusion.) Unlike the physical objects discussed as real evidence, the document is not mute. Many of the major rules governing documentary evidence are designed to keep them silent until they have been properly sponsored. Thus it is imperative that a document tendered to prove its contents be identified in some way before receipt. The principal means employed is the production of a witness who executed it, who saw it being executed or who recognizes the writing or signature to be that of a certain person. In addition, U.S. law has evolved what is called the reply-letter doctrine, a letter received over the signature of Doe which is in response to a letter addressed to Doe may be assumed to be written by Doe although none of the foregoing means of authentication are available. Somewhat more stringent requirements of identification relate to certain instruments, notably wills, which are required by statute to be executed in the presence of witnesses who also sign or subscribe the document. English law and the majority of U.S. jurisdictions require that some effort be made to produce those subscribing witnesses to identify the document. All, however, permit alternate methods when the subscribing witnesses are dead or otherwise unavailable.

Additional means of identification are also available. It may be said with reasonable safety that in English practice any public document may be produced in court to authenticate itself. U.S. practice is more variable, being the product of an enormous range of statutory provisions which differ from jurisdiction to jurisdiction. Statutes, reports of judicial decisions and official registers are everywhere self-authenticating, however. Because many public documents cannot be removed from their official repository, it is commonly provided that their contents may be proved by copies certified by the legal custodian of the record to be accurate and "true."

The common law was obliged to recognize that the ordinary means of identifying private documents were apt to disappear with the lapse of time. For that reason, a particular rule relating to ancient documents—arbitrarily set at 30 years as the approximate length of a generation—permitted their receipt in evidence when the document was dated more than 30 years before the time of the trial, when it appeared to be regular or unsuspicious on its face and when it was found in a place where items of its character probably would be found. This rule prevails in the U.S. and has been altered in England only with respect to the period of time, which was reduced to 20 years by the Evidence act of 1938.

The above means of identifying documents serve only the preliminary purpose of authorizing their receipt into evidence; they do not finally establish that the document is genuine, an issue that may be contested at great length. When it is contested, the evidence may come from what may be called semiexperts, men whose profession involves repeated handling and judging of handwriting, such as bank tellers. It consists of comparing the instrument in question with established samples, known as exemplars, of the writing of the person who allegedly executed the instrument and sometimes of that of the alleged forger as well. An emerging class of specialists has taken over most of this service, however. Employing laboratories equipped for chemical analysis, for electronic inspection and for photographic enlargement, they have replaced the simpler expert in all but small cases. Similar services, perhaps less advanced in technique, may be performed where the instrument is wholly typed or printed.

It must be emphasized that even when fully identified the document is received only as what it purports to be. Statements or recitals of fact contained within it cannot be used to prove the existence of such fact unless the particular statement in question qualifies under some exception to the rule against hearsay evidence. For example, the written confession of the accused in a criminal case may, upon proper identification, be introduced to prove that he committed the offense; it comes within a standard exception to the hearsay rule. But an attempt by the accused to introduce the written confession of some other person might fail even if the writing were fully and properly identified because its contents are hearsay.

4. Testimonial Evidence.—By far the largest volume of evidence received during the trial of a modern lawsuit comes from witnesses on the stand under oath. Terminology differs in various jurisdictions. Some English writers distinguish between testimonial evidence, which is a direct assertion by a witness that a fact in issue did or did not occur, and indirect or circumstantial evidence, in which the witness testifies not directly to the fact in issue but to others which warrant the inference that the main fact did or did not exist. U.S. terminology regards both as testimonial, employing the distinction between direct and circumstantial evidence to express the difference described above. The differing nomenclature causes occasional verbal confusion but is not indicative of any significant differences in the rules of law involved.

Historically, the courts of law took all the testimonial evidence from witnesses sworn and in attendance at the proceeding. The courts of equity and some other judicial institutions took evidence in the form of written, sworn statements or affidavits. The widespread abolition of the separate systems of courts resulted in adoption of the law-court practice and the substantial exclusion of evidence in the form of affidavit. But during the present century there has been a tendency throughout the common-law world to accept again the written testimony of witnesses in order to spare the expense of their attendance in court. This form of testimony on commission or deposition is authorized by rules of the supreme court in England and by rules of procedure and statutes in approximately half of the U.S. jurisdictions. Further increase of the practice is both well-justified and certain to occur. It is a substantial improvement over the old equity affidavit because the testimony contained in the deposition is both under oath and subject to cross-examination by opposing parties at the time it is given. Perhaps reading it to the jurors is not equivalent to their hearing it firsthand, but the convenience and reduced expense more than balance that loss. *See* also OATH AND AFFIDAVIT.

III. WITNESSES

Because the great bulk of evidence received is in the form of verbal statements of witnesses at trial, a substantial body of law regulating this process has developed. Unlike continental terminology, where the term witness denotes only those having firsthand knowledge of the facts, Anglo-American usage applies it to anyone who gives testimony, including experts and those who identify real and documentary evidence.

1. Competency.—Historically, parties to the action were incompetent to serve as witnesses, as were their spouses and persons financially interested in the outcome. So, at different times, were convicted felons, Jesuits, non-Christians and those Christians whose scruples prevented them from swearing the oath. By statu-

tory evolution the old rules of incompetency have largely been swept away. Both English and U.S. law retain the perhaps indispensable grounds of immaturity and mental disability, but even these are flexible in operation; a child beneath statutory age may nevertheless testify if it is shown that he is able to perceive facts and if he acknowledges an obligation to relate them truly.

Nearly every United States jurisdiction adds a third ground of incompetence called the "dead man statute." Under it the surviving party to a transaction or conversation with a person since deceased (or, in many states, who has since become insane) is disqualified from testifying about the transaction or conversation when the survivor sues the estate of the deceased in an action based upon that transaction or conversation. The fear expressed by this is not unreal; the survivor has both motive and opportunity to fabricate or colour a tale in his interest. But the statute is as destructive of true evidence as of false and is inconsistent with the modern attitude that credibility is a jury question. For this reason, courts have applied the statute grudgingly.

The common-law disqualification of spouses of parties is not wholly abolished but has been vitally changed. Nearly all jurisdictions receive testimony which is given voluntarily by the spouse sworn as a witness and which is not objected to by the spouse who is a party to the litigation. A majority now permit the spouses to be called as witnesses both for and against each other in civil cases but still retain for criminal cases a right in the litigant spouse to prevent any testimony by the other. If justified at all, these rules now are more in the nature of privileges (treated below) than they are related to the historic rules of incompetency from which they derive.

2. Attendance and Procurement.—Witnesses who are within the jurisdiction of the court may be summoned by subpoena and compelled to attend and answer under threat of fine or imprisonment. Attendance will be excused only for such extraordinary cases as personal illness or family emergency. Witnesses not within the jurisdiction usually cannot be obliged to attend, although they may come voluntarily; and, increasingly, resort may be had to their depositions as a substitute for testimony in person. Those who do attend usually receive a daily allowance and a sum to compensate for travel expenses. They are given some protection from insult and abuse and are usually exempt from service of civil process while in attendance and from arrest if they voluntarily enter a jurisdiction to give testimony. For their statements on the witness stand they are given an immunity from liability for defamation, even if their testimony be proved intentionally false.

3. Oath.—Historically, all testimony was given under sanction of an oath: the witness called upon God to attest his testimony. This requirement has been universally relaxed. In the 19th century, those with religious scruples against swearing oaths were permitted to affirm under the pains and penalties of perjury instead. It is increasingly common practice to permit any person to affirm rather than swear without asking his reason for preferring affirmation. Even small children who have never heard of God and who would be immune from prosecution for perjury are permitted to testify upon a showing that they realize that lies lead to parental displeasure and possible punishment.

4. Examination.—Testimony is normally produced in response to questions directed to the witness by counsel. Direct examination, or, as it is called in England, examination-in-chief, is the questioning by the party who called the witness. The direct examiner is usually prohibited from asking leading questions—those which suggest the answer and require only confirmation by a co-operative witness. The rule is not absolute, and if the witness is hostile to the examiner or, because of language difficulty, senility or embarrassment, cannot be questioned without leading, it will be allowed. Another restraint requires that the witness' firsthand knowledge of the facts be given rather than his opinion or conclusion. Forming an opinion or drawing a conclusion is normally for the jury. Yet some forms of impression or opinion have always been allowed. The identity and age of persons, the speed of moving vehicles and other impressions incapable of detailed description are accepted. When sobriety, sanity and emotional state generally are in question, the witness must usually make some effort to state the facts which led him to his conclusion concerning the point, but he is also allowed to state his conclusion. Such exceptions to the rule against opinion are justified on the ground of necessity. A newer and growing invasion of the rule would permit opinion whenever it would appear helpful to the judge or jury in determining facts.

A category of witnesses called experts are not subject to the opinion rule. Experts are employed for two purposes: (1) to obtain firsthand knowledge that only specialists can perceive and (2) to express opinions on matters so technical that the judge or jury would have difficulty understanding them without the aid of expert knowledge. For example, a medical specialist might give his firsthand knowledge that a skull was fractured, although an ordinary witness could not detect it; he might add his supposition that the fracture caused concussion of the brain, although it is clearly a matter of opinion.

5. Cross-Examination.—Cross-examination is the process of directing questions to the witness by counsel other than the one who called the witness to the stand. Because the witness is normally hostile to the cross-examiner, the danger of accepting the suggestion of leading questions no longer exists, and leading is not only freely allowed but is the normal pattern.

Under the English and minority U.S. rule, the witness may be cross-examined on any subject which properly pertains to the case being tried. But the U.S. federal courts and a majority of state courts limit inquiry on cross-examination to subject matter that has been brought out on direct examination of that witness; the cross-examination must explain or modify or contradict the direct examination. This limitation on the scope of cross-examination has little to commend it. One case permitted a man charged with killing his wife by beating and poisoning to escape cross-examination about the beatings because, on direct examination, counsel had carefully restricted the testimony to a denial of the poisoning alone.

Cross-examination in U.S. jurisdictions may be followed by re-direct examination and it in turn by re-cross-examination. In England, cross-examination is followed by re-examination, but here, except by leave of court, the process ends. But since the universal rule prohibits expansion of these stages to matters not previously explored, examination of witnesses does not go on endlessly as the names would suggest.

Although cross-examination is not unknown in civil-law systems, it is accorded nowhere else the prominence it enjoys in Anglo-American procedure. A leading scholar has called it "the greatest legal engine ever invented for the discovery of truth" and "the great and permanent contribution of the Anglo-American system of law to improved methods of trial-procedure."

6. Impeachment.—Impeachment is the process of discrediting testimony by showing facts which tend to reflect upon the veracity of the witness and thus to authorize rejection in whole or in part of the testimony he has given. This can be done in many ways. The elicitation of testimony of contrary facts from other witnesses forces judges or jurors to choose which to believe. Again, doubting a witness' perceptive capacities—hearing, sight or his opportunity to perceive—may discredit his testimony. In addition, a body of formalized rules govern certain established forms of impeachment. There are four universally recognized forms: (1) showing that the witness has previously made statements inconsistent with those made on the stand (*i.e.*, in the box) under oath; (2) showing that the witness is biased either for or against one of the litigants; (3) showing general bad moral character of the witness; and (4) showing that his character for truth and veracity is bad. Two others are sometimes added. One is the introduction of evidence contradicting the witness on facts in the case. This is not subject to the specialized rules here treated. The other is virtually extinct; it consisted of questioning the witness about his belief in God, upon whom he had called in oath to vouch for his testimony.

Prior Inconsistent Statements.—Great latitude is allowed in this form, although it is everywhere required that the witness be allowed to explain or reconcile the prior statement, and affirmative proof of the inconsistency is allowed only when the subject matter of the statement is itself an issue in the case. An important limitation is that the prior statement is usually hearsay and hence cannot

be received to prove the facts recited in it but only to show that the witness has given differing accounts. For example, a witness testifies in court that he was not robbed on a certain date. For impeachment, it is proved that he has made a prior statement that he was robbed on that date. If nothing more occurs, the jury would be entitled to disbelieve his denial of the robbery but unauthorized to believe that he had been robbed; the earlier statement is hearsay.

Showing of Bias.—Sympathy for one party or animosity for another may colour testimony, wittingly or unwittingly. Hence family relationship, personal friendship and interest in the outcome of a case are open to inquiry. Hostile statements or actions by the witness in regard to one side may be shown. Greed and other personal interest may also affect testimony. One common example involves testimony in a criminal case from a witness who is himself in the hands of the law; wide inquiry is allowed to determine whether that testimony has been obtained or encouraged by suggestions of the prosecutor that leniency will follow cooperation in giving evidence. Payment received by an expert for testifying can be exposed because of its tendency to produce favourable testimony. Some courts have adopted the practice of obtaining an expert appointed by the court with the approval of the parties. The fee is paid ultimately by the losing party, but since payment no longer depends upon the import of his testimony the expert is freed of the urge of avarice.

Showing Bad Moral Character.—Almost everywhere, the fact that a witness has been convicted of a felony may be shown to the jury for the purpose of persuading it that he is to be disbelieved. Only about a dozen jurisdictions stop there, however; many permit the showing of misdemeanour convictions as well, and under the English rule and that followed in many U.S. states, even misconduct which has not resulted in conviction may be asked about. The only proofs of misconduct that will be accepted, however, are the witness' own admission or the record of his conviction.

Two criticisms often directed at the rules in this field are that no distinction is made between conduct that reflects upon veracity and that which reflects on other character traits, and that under the nearly uniform U.S. rule (although the contrary is true in England) the accused in a criminal case can be impeached as readily as any other witness despite the obviously increased prejudicial effect in his case of showing other crimes to the jury.

Character for Truth and Veracity.—The essential elements of this type of impeachment are ancient and now highly stylized. After a witness (usually crucial) has testified, the opposing side calls one or more witnesses who testify that they know the reputation of the crucial witness for truth and veracity in the community in which he resides, and that his reputation is bad. In England and many U.S. jurisdictions, the impeaching witness may add that he would not believe the crucial witness under oath. After, but only after, this has been done, the same process may be employed to show that the crucial witness has a good reputation for truth and veracity. Like all ritual, the process has occasionally suffered from a failure to remember its purposes.

One rule which applied historically to all four impeachment devices described was that a party could not impeach "his own" witness, and ownership was determined by who summoned the witness to the stand. But it is apparent that witnesses are sometimes treacherous, promising one thing and testifying to another. Less venally, they may forget or change their minds. Recognition of this has prompted a yielding of the rule in one area; a counsel who is genuinely surprised by a change of story will be permitted to call to the attention of the witness his prior statements on the same subject. He is still prohibited from maligning the character of the witness he calls.

An even broader exception is also pertinent. By widespread statutory provision in the United States, a party is permitted to call his opponent in the litigation and subject that opponent as a witness to the same manner and type of cross-examination and impeachment as could be employed against witnesses summoned by that opponent. In England a party could subpoena his opponent, but it would be most unusual to do so.

The consequence of successful impeachment is not the striking-out of the testimony of that witness; this would be merely to revive discarded common-law rules of incompetence. Instead, the judge or jury is at liberty to disregard the testimony or to accept part and reject part. Another consequence is that after impeachment has occurred the party producing the witness may make an effort to rehabilitate or bolster that witness. The witness may deny the impeaching evidence or reaffirm his oath to tell the truth in spite of it; and, in appropriate cases, his prior consistent statements may be shown as well.

Minor improvements might be made in the impeachment devices explored, and major advance may come someday through the agencies of modern science. A few cases have already permitted the use of psychological insights into veracity, but courts have unanimously refused to employ the results of the so-called lie detectors and truth serums, on the ground that they are not sufficiently perfected to be reliable.

7. Judge and Jury.—The summoning and examination of witnesses by counsel is the Anglo-American practice, as contrasted with the civil-law pattern of questioning by the judge. Although the right of the judge to question is generally conceded, it must be exercised cautiously; clarifying questions are proper, but extended questioning is not. Jurors are even more restricted, usually being permitted to address questions only through the judge. U.S. courts concede the propriety of a judge to summon witnesses not called by the parties; English and commonwealth practice usually forbids it without consent of the parties. Although these restraints are variously explained, it is evident that they originate fundamentally in the postulate of an adversary system of litigation employing a relatively passive finder of fact.

IV. PROOF OF FACTS

Two kinds of issues arise during the course of trial: those of law and those of fact. The former are the province of the judge. The latter are traditionally for the jury, if one exists, but there are a few exceptions to this standard. Whether a document has been identified sufficiently, whether the facts necessary to authorize application of a hearsay exception exist and whether the offered confession of an accused criminal was voluntarily given are the most frequent forms of questions of fact passed upon by judges prior to submission to the jury. Establishing such facts is frequently referred to in the U.S. as "laying foundation."

1. Relevancy and Materiality; Direct and Circumstantial Evidence.—The basic proposition of Anglo-American evidence law is that anything having a tendency in reason to prove the existence of a significant fact is admissible to prove that fact unless some particular rule of law precludes its use. The law of proof is thus, somewhat paradoxically, a contest between human reason on one hand and common sense on the other; its function is to balance the value of truth against the cost of attaining it.

In the United States a frequent form of objection to a fact is that it is "incompetent, irrelevant and immaterial." The objection of incompetence is too broad to be meaningful. Materiality does have meaning. A fact which must be proved in a legal proceeding—such as the fact of death in a murder prosecution—is a material fact. A relevant fact, in contrast, is one which is not material in itself but which may warrant a belief that the material fact does or does not exist. For example, a witness testifies that he saw the defendant fire a pistol bullet into the body of the deceased, who promptly died. This is a direct testimony of material facts in murder; the only question for the jury is whether the witness is telling the truth. If, however, the witness is able to testify only that he heard the shot and that he arrived on the scene seconds later to see the accused standing over the corpse with a smoking pistol in his hand, the question of relevance is raised. If human reason would regard this fact as making it more probable than it was before that the accused killed, the evidence is relevant. Even if believed by the jury, it is far from conclusive, however. The accused may have been a policeman shooting at the escaping killer; he may have been a bystander who picked up the weapon dropped by the true killer. There is more than the question of truth or falsity involved, and other evidence will serve to clarify the inference.

Another way of expressing the same distinction is that the testimony of the witness who saw the shooting is direct evidence, while that of the witness who arrived after the shooting is circumstantial. The not uncommon conception that one cannot be convicted on circumstantial evidence is wholly false. Most criminal convictions are based on circumstantial evidence; indeed, studies have indicated that fewer mistaken convictions are obtained on circumstantial evidence than are the product of direct but erroneous identification by eyewitnesses.

It should be recognized that direct evidence of material facts is not subject to an objection of irrelevance; only circumstantial evidence presents a relevance problem. The solution of that problem depends upon a fundamental test: whether the fact offered to be proved (if not itself a material fact) has a tendency in human reason to establish a material fact. But an affirmative answer to that question does not assure admissibility, for the cost of receiving it may be too great. Four judicially recognized categories of counterweights exist: (1) reception of the evidence will consume an amount of time disproportionate to its value; (2) the unwarranted prejudice to a party will outweigh the probative value; (3) the evidence will be confusing to the jurors or may create false issues in the case; and (4) parties or witnesses may be confronted with accusations which they are not prepared to meet and refute. The process of balancing may be said to be a new one in every case, and most U.S. courts and the house of lords recognize that considerable discretion must be exercised by the trial court in striking a balance. Still, many situations have so often recurred that something like rules have grown up about them. Several of these deserve mention.

Similar-Fact Evidence.—The law traditionally has regarded evidence of prior instances similar to the one under litigation with great suspicion. Previous dealings of the litigants with each other which resemble the case at hand may be received to indicate a certain course of conduct. If only one of the parties was involved in the prior instance, exclusion is the general rule. If neither party was involved, the evidence will rarely be received. The clause in a contract will be construed as if no prior contracts were made; the representations will be judged true or false without regard to prior false representations; the auto accident will be viewed as if the defendant had never been involved in prior accidents similar to this one. It may be said that more discrimination has been shown in recent years, but on the whole the rule is, and is likely to remain, in favour of exclusion.

Prior Offenses.—Perhaps only a specialized application of the rule against similar-fact evidence is the uniform practice of excluding in criminal prosecutions any evidence that the accused has previously been convicted of, or has committed prior offenses similar to the one now charged. It must be admitted that such evidence has some probative value; criminals do repeat. But the law holds that the possible harm to the accused from such evidence would exceed the legitimate inferences that might exist and that the limited value of the evidence does not justify the risk. What seems the best statement of the rule is this: evidence that a person has committed a crime on some prior occasion is inadmissible for the purpose of proving propensity or disposition to commit the crime with which he is presently charged. The reasoning that because he has committed crimes he is evil, and because he is evil he probably committed this crime, is everywhere disapproved. However, if the fact of prior crimes would serve some purpose other than showing propensity, it may be received. To show that a defendant did not administer arsenic in coffee under the delusion that it was sugar, it is proper to show that several of the defendant's husbands died of arsenic poisoning; to show that the defendant strangled and mutilated a child, it is proper to show that he had on other occasions strangled and mutilated other children under similar conditions. In the latter case the purpose is not merely to show that he is evil but that his crimes are so distinctive in form and execution that the method marks him as the culprit, while in the former the possible defense of mistake is rebutted by showing that the defendant is familiar with the properties of arsenic. The danger that the jury will, although not satisfied of guilt for the later crime, convict because of the former crime is realized but risked. Frequently it is said that prior crimes are admissible when the purpose is to prove absence of mistake or accident, motive, opportunity, intent, preparation, plan, knowledge or identity.

That the admissibility of prior offenses is essentially a matter of balancing risks is shown by an example of a rape case in which the defense is consent. It may not be shown that the defendant has previously raped girls to prove that he raped this one, although it may be shown that the prosecuting witness has previously consented to sex acts to show her propensity to consent to this one. Probative force is about equal in the two cases, but the potential for prejudice is greater against the defendant.

Character.—Character is sometimes a material fact in a lawsuit and is not then subject to the test of relevance. When it is stated that a man is a thief, his character in that respect is the very issue in an action for slander brought as a result of the statement, and it can then be proved freely. Apart from cases in which character is in issue, the law has a fairly inflexible rule that character evidence cannot be employed in civil cases. The only substantial body of exceptions is when the gist of the civil action is itself a criminal act, as when a man is sued for the harm he caused by deliberately burning a building; the similarity of the case to a criminal charge of arson often permits the criminal practice to apply. In accord with such practice, the accused has the privilege of putting his character in issue by producing witnesses who will testify that they know his reputation for the trait involved (such as honesty and integrity in a fraud charge) and that that reputation is good. From this the jury is to be asked to infer that because the reputation is good the character trait is probably good, and if the character trait is good the defendant probably did not commit the charged offense. This is an attenuated chain of inference, and it is not available to the prosecution unless and until the defense has employed it first.

When character is used circumstantially in this way, the evidence of character is universally restricted to reputation for the trait in question. Proof by specific conduct or by opinion of witnesses is regarded as too dangerous a method and too time-consuming for what is at best rather weak evidence. Indeed, in England the position has sometimes been taken that proof of character is not really evidence at all. A substantial number of U.S. jurisdictions, however, take the position that character evidence may of itself be sufficient to raise a reasonable doubt of guilt.

Habit.—Although the law has refused circumstantial use of character evidence in civil cases, it has been willing to employ habit. Perhaps the best distinction that can be drawn is that habit is narrower in scope. A general practice of being careful and cautious might be called a character trait, while the practice of checking the chamber of a gun for cartridges before storing it might be called a habit. Although the law avoids the use of habit evidence relating to skill or care, preferring to consider the question of negligence uncoloured by past practice, it nevertheless does admit habit evidence in other civil cases. For example, the habit of a stableboy to tease a certain horse was received for the purpose of showing that while teasing the horse he was kicked in the head and killed.

Insurance.—Motor-accident cases occupy much court time, and a question repeatedly raised in them in the U.S. is whether the fact that a motorist is insured against liability for negligence is admissible. The courts have as repeatedly insisted that it is not, when the only probative use of the evidence would be to suggest that the motorist exercised less care than required because of indifference induced by the protection to his pocketbook. What slight probability is contained in the idea is more than countered by a supposed tendency of jurors to be generous with insurance company money. But if the existence of insurance tends to prove something other than lack of due care, it can be received despite the danger; hence it is sometimes employed to show ownership of the vehicle when that is disputed.

2. Hearsay.—J. H. Wigmore called the rule excluding hearsay the "most characteristic rule" and "greatest contribution of that eminently practical legal system to the world's methods of procedure." No doubt it is the most distinctive and best-known rule of Anglo-American evidence law. But it is also one of the most misleading; the very name deludes and oversimplifies. From it,

the prohibition would seem to be upon testimony which consists of what the witness has heard others say. It is true that the following is a condition of suspicion: "I do not know myself, but I have heard others say that . . ."; but it is far from true that continuing the sentence would be bad in all cases, nor will avoiding that predicate avoid the rule in all cases.

U.S. and English courts are in agreement that the vices of hearsay are that it is testimony which depends for its value upon the truthfulness and accuracy of one who is neither under oath nor subject to cross-examination. The oath alone is no longer important, if it ever was. The affidavit, the classic statement under oath, is and always has been hearsay, for want of the opportunity to cross-examine. It is the latter upon which the rule depends. In addition to outright fabrication, hearsay presents the dangers of: (1) misuse or inaccurate reporting of language; (2) insincerity of the perceiving person; (3) lack of perception on the part of the perceiving person; and (4) faulty memory. The one whose language is (at least partly) in question, whose sincerity may be doubted, whose perception may have been faulty or obscured and whose memory may have failed is not the witness on the stand (who may be tested by cross-examination) but the person whose statements are being reported. This person is known as the declarant, and it is around the qualities of the hearsay declarant that all the controversy rages.

A brief statement of the hearsay rule, and a reasonably workable definition of it, is: An out-of-court statement offered to prove the truth of the matter stated is hearsay and inadmissible (unless it falls within one of the many exceptions). Testing the definition, it is clear that many human utterances are not hearsay, although they may indeed take the form of what someone heard said. "I accept your offer," "How far is it to Birmingham?" and the like are not within the definition because they are not assertions of fact. Because they are not, they cannot be "offered to prove the truth of the matter stated." Sometimes a statement is an assertion of fact but is offered not to prove the fact asserted but some other point. If not offered to prove the truth of the fact asserted, it does not violate the hearsay rule. For example, the defendant on a murder charge may prove that the deceased had asserted that he intended to kill the defendant. If offered to prove that the deceased did so intend (and was therefore the aggressor), the evidence would be hearsay; while if offered to prove that the defendant heard of this plan and was as a consequence fearful for his life, it is not offered to prove the matter asserted but only as grounds for a reasonable belief on the part of the defendant. A brief statement of this general principle is that when the mere fact that words were uttered (without regard to their truth or falsity) is itself relevant, proving the utterance does not violate the hearsay rule.

The rule may be violated by conduct and even by silence as well as by words. "I did not hear . . ." may violate as well as "I heard . . ." One example of conduct is an English case in which patrons of a public house left the beer they had ordered without drinking it. When the publican sued the brewer for furnishing unpalatable beer, the conduct of the customers was offered; it was valuable only in that it tended to show the opinion of the customers that the beer was unpotable. They may have been wrong, they may have been insincere, they may have been venal; cross-examination could test this. A United States case illustrates silence as hearsay. A feed dealer was sued for breach of warranty on corn he had sold as number 2 grade. To refute the complaint that the corn was of inferior quality, he produced evidence that he had sold quantities of the same corn to other buyers under the same description and that none of them had complained of the quality. Their failure to complain was valuable as evidence only insofar as it represented the views of those buyers that the corn in fact conformed to the representations, and they were not subject to cross-examination. This is the very frontier of hearsay.

In both of the cases, the evidence was received despite the hearsay dangers admittedly present. Highly respected scholars approve the decisions on the ground that the dangers of insincerity and the like are much less when the conduct offered is not intended as an assertion of fact, even though capable of being employed as one. Both English and U.S. courts remain uncertain of the proper rule, however.

It should be said, nevertheless, that despite such peripheral doubts the main body of the definition is well settled: Only occasionally will respected courts be led to solemn consideration of whether evidence that bloodhounds trailed a certain person from the scene of a crime by detecting his scent at the scene is hearsay, on the grounds that it represents merely the opinion of the hounds, who are neither under oath nor subject to cross-examination.

The definition given is so sweeping that it would be nearly impossible to try a modern lawsuit if the rule of exclusion were not riddled with exceptions. It is fair to say that more than 30 are recognized, in addition to which a large number of statutory provisions make allowance for proof by records and writings which contain hearsay. Examination of a few of the major exceptions will illustrate their character.

Prior Testimony and Depositions.—Testimony of witnesses recorded in prior trials as well as depositions taken for use in the trial in question are only doubtfully hearsay, for the testimony was both under oath and subject to cross-examination. Still, the present jury has not had opportunity to see the witness as he testifies, and these forms are universally regarded as hearsay and will be received only when the declarant who gave the testimony is unavailable for the trial because of death or absence or the like.

Admissions of Parties.—This is doubtless the most frequently invoked exception to the rule. Almost any prior statement or conduct of a litigant which is inconsistent with his present position at trial may be shown. The name admission implies that the statement is adverse to his interest, although it need not be conclusive in its terms nor need its inconsistency have been apparent at the time it was made. The concept of admission is responsible for one of the most common misunderstandings in evidence law; namely, that "self-serving" evidence cannot be received. But it is apparent that most testimony of the parties in a case supports their cause and is self-serving. What is excluded is the out-of-court self-serving declaration; the reason is that it is hearsay and not within the admissions exception to the rule.

Several common variants of the admission are employed. One of these is the so-called adoptive admission, when the party in response to an accusation or charge of misconduct made to him failed to deny or protest under circumstances which would provoke denial from an ordinary man. The accusation itself is and remains hearsay, but the conduct of the accused is thought to constitute a possible acknowledgment of its truth. Of course, not every accusation demands response; one of the major disputes is whether a man under police arrest should be considered to have accepted accusations by not denying them, or be thought merely to be wise to have refused to answer any questions at all.

Vicarious admissions are those in which the party did not himself make the statement or do the act but which are charged to him because the one who did was his agent authorized to speak or act on his behalf. The appropriate authority is not easily found, and the law is gradually moving in the direction of admitting the statement if it relates to some matter which is itself authorized even though the words spoken are not.

A very common form of admission of a party is the confession of the accused in a criminal case. It presents very few hearsay difficulties, however; the more compelling questions of privilege overshadow the hearsay involved.

Declarations Against Interest.—Although frequently confused with the admissions exception, declarations against one's interest differ markedly in character: the declarant is not a party and, indeed, the common-law and still prevailing rule requires that the declarant be dead at the time of trial before his statement can be received. No doubt his absence is the basis of the strict standard that the statement must be clearly against his interest, not merely inconsistent, and that it must have appeared so at the time it was made. The principal complaints about this exception are two: that the death requirement should be relaxed to permit use of the exception when the witness is unavailable for other reasons, and that by an indefensible limitation the interest impaired by the declaration must be pecuniary or proprietary rather than penal. A state-

ment which jeopardized the declarant's proprietary interest in a farm would suffice under the general rule; one which might convict him of murder and forfeit his life would not. To state the rule is to expose its deficiencies.

Despite grave differences between this and the admissions exception, the two have something in common: the justification for each is that statements adverse to one's interest are probably true, for a man seldom desires to fabricate them and usually will take pains to see that they are true and accurate.

States of Mind and Res Gestae.—Courts have been slow to admit that a man's out-of-court statements about his state of mind could constitute hearsay. Probably the reason for the reluctance has been a recognition of the necessity for having such evidence. To avoid this difficulty, and because analysis had not proceeded very far, state-of-mind utterances were put into a formless mass called *res gestae* that once included a large body of evidence rules. Increasing analysis of the problems involved have refined that mass until the term *res gestae* remains in United States practice only as a shorthand description of several related hearsay exceptions, although it retains greater meaning in English usage. The hesitation to admit that state-of-mind declarations were hearsay has disappeared, and it is generally conceded that they are. Those which relate to an existing state of mind are within this exception, as are similar statements of existing physical and emotional states. "My arm hurts," "I feel depressed" and "I hate John" all are receivable under this exception. Memories of past states do not qualify generally, although in the United States statements of past physical symptoms made to a physician for the purpose of obtaining medical treatment are received; the patient's desire to obtain proper treatment encourages him, it is thought, to be careful and accurate. Another aspect of U.S. development not accepted in England involves a statement of existing intention when offered to prove not only the intent but also that the declarant subsequently acted in accord with that intention; the *Hillmon* doctrine, as this is called, is accepted throughout the United States but employed with great caution.

Another doctrine singled out of the *res gestae* mass in the United States but viewed skeptically in England is the excited-utterance exception: declarations about an exciting event so closely coupled in point of time with that event that the excitement precludes calculation or deliberation—the hearsay dangers.

Dying Declarations.—More a historical curiosity than a major hearsay exception is the so-called deathbed declaration. On the assumption that one consciously aware of his impending death would not wish to meet his Maker with fresh lies charged to his account, the law receives declarations from such people when the statement is offered in a criminal prosecution for causing the death and relates to the cause of death.

Business and Hospital Records.—The hearsay rule was formulated at a time when all aspects of any individual business transaction would have been within the knowledge of some single person. In a more complex society, this assumption often fails. Large enterprise, modern accounting principles and electronic record-keeping machines all combine to create circumstances in which no one person knows or ever knew all the facts of any individual transaction. As a substitute for this personal knowledge, courts and legislatures in the U.S. have come to accept the probability that the necessity for accurate records for the conduct of commerce creates sufficient assurance of accuracy that those records may be received in suits arising from the business they reflect. Business-record statutes in the United States are a concession to this need.

Hospital and medical records are sometimes grouped with business records but stand basically on a different footing. Their assumed accuracy arises not from the impersonal, machinelike character of business records but from the very opposite: the personal and intense care which accompanies concern for human health and life. Slowly, that assurance is being accepted in the U.S. as a substitute for cross-examination and is creating a hearsay exception for medical records.

Public Records.—The records kept by public officials and the facts reported by them in the course of their duties fall within a broad and generous exception in English law. U.S. law, while employing many specific statutory exceptions, generally does not recognize any such formulation. It is true, however, that vital statistics as such are commonly admissible by statute: public records of birth and of death, of marriage and of divorce, are everywhere received.

Pedigree.—A moment's reflection indicates to the average man that the things he knows best are in fact only hearsay: his name, his age, the identity of his parents and family relationships in general are known to him only through the unsworn, uncross-examined repute regarding such matters in the family circle. And so, despite improved and expanded systems of record keeping, this ancient exception is probably no less necessary than it was two centuries ago.

The future of the hearsay rule is difficult to determine. Increasing recognition that hearsay does have probative force indicates that the ultimate disposition of the law will be to receive it with the caution its infirmities call for in each case. Greater trust in the jury as the finder of fact, and diminishing use of the jury, tend to make possible a case-by-case approach to hearsay evidence, as opposed to the traditional firm rule of exclusion accompanied by equally arbitrary exceptions.

V. PRIVILEGE AND PUBLIC POLICY

Much of the law of evidence is a product of balancing internal considerations with the object of obtaining the most, and the most reliable, truth. Privilege differs in that the value of truth is balanced against purposes extrinsic to evidence law and is sometimes sacrificed deliberately on the assumption that greater justice may flow from that choice.

1. Self-Incrimination.—One of the most significant and fundamental (and thus most controversial) of these privileges is that against self-incrimination. Used in a large sense, it includes both the right of a witness to decline to answer questions which will expose him to punishment for crime, and the right of the accused in a criminal prosecution to elect to give no evidence at all. The latter form is probably a mere remnant of what was originally not a right but a disability—the accused was not competent as a witness in his own behalf. But remnant or not, it is a valued right today, recognized throughout the Anglo-American world.

The narrower privilege against self-incrimination is equally widely recognized today. The language of the fifth amendment of the U.S. constitution protects one from being "compelled in any criminal case to be a witness against himself . . . ." Although this is a common formulation, it is inadequate to describe the scope of the privilege. It exists in any type of hearing—criminal or civil, judicial, administrative or legislative. And the only witness who will be denied it is the accused in a criminal case who voluntarily takes the stand to testify in his own behalf; with regard to any matter relevant to his case, he is said to waive the privilege not to give incriminating answers by electing to testify at all.

Incriminating testimony is that which would expose or tend to expose the witness to punishment for crime; that which would incur public scorn, economic loss or reprisal or even such governmental action as deportation does not qualify. And the possibility of punishment must be real, if not immediate. Apart from a vague doctrine which denies the privilege when the possibility of prosecution is only fanciful, more precise rulings make clear that if periods of limitation have expired, the answer can no longer incriminate and must be given. So also, a government sufficiently desirous of obtaining information may grant immunity from prosecution for any offenses disclosed; over bitter protests, statutes making such grants of immunity available have been upheld if the immunity is as broad as the offense disclosed rather than mere immunity from use of the testimony elicited. Another limitation upon the concept of incrimination denies it when the offense disclosed is not one against the law of the government seeking the answers. But although there are both English and U.S. rulings to that effect, the point has been doubted in England, and a growing body of U.S. law rejects the limitation as inappropriate between states of the U.S. and intolerable when applied between state and federal governments.

Important differences in the doctrine of waiver exist between U.S. and English practice. The general attitude in the U.S. is

that a witness who has disclosed any portion of an offense is no longer at liberty in that proceeding to rely upon the privilege, although in some subsequent proceeding he may do so. The adverse consequences of the doctrine are several. One is that it prompts early and even premature claims of the privilege. Another is that, despite the rule that the judge rather than the witness passes on the validity of the claim of privilege, the claim raised very early in questioning gives the judge little or no information on which to base his guess about whether the answer might incriminate. English practice avoids this waiver rule and, with it, the consequences described.

Some miscellaneous but important limitations on the privilege are that it does not protect artificial persons such as corporations and that it does not protect even natural persons from compulsory surrender of "public" records which contain incriminating evidence. Understandable as this may be when the record is truly public, it creates serious problems when the record is essentially private and is public only in the sense that the law requires that it be kept and made available.

Two observations which pertain to both forms of the privilege—remaining off the stand and declining to answer—should be stated. One is that the right seems to apply only to "testimonial utterances" and does not prevent visual inspection of the witness or defendant, compelling him to try on articles of clothing or even, perhaps, the nonviolent removal of body fluids and tissues for testing. The other is the scope of inference which may be permitted from the fact that the privilege is exercised. By the English rule and the substantial U.S. majority view, none at all is permitted in a criminal case; many U.S. courts do permit inferences of guilt or misbehaviour to be drawn by the jury when the privilege is invoked in a civil proceeding.

The common assertion in U.S. legal literature that the privilege against self-incrimination is unique to the common-law system is more false than true. On the continent generally, privileges for witnesses are provided by statute, and they do not differ significantly from the general Anglo-American right to refuse to give self-incriminating evidence. Unlike the criminal defendant under the common-law system, the European defendant may have questions addressed to him whether or not he elects to take the stand. He has a theoretical duty to testify, but no direct reprisal may be taken against him if he declines. And he retains the right that is now disappearing in England and is all but unknown in the United States to testify about the offense with which he is charged without becoming a witness or being sworn.

2. Confessions.—It would be of small value to an accused to have the right to remain silent during his trial if he could be compelled to speak of his guilt in advance of it. As a consequence, it is the universal rule—enforced in the United States as a part of constitutional law—that a confession obtained from an accused may not be used to prove his guilt unless it was voluntarily given. It is not necessary that inducement be avoided altogether by the police, nor do U.S. courts follow the British practice of requiring police to warn one arrested of his right to remain silent. But certainly torture or physical abuse are everywhere sufficient to require exclusion; duress, such as threats against family, and psychic force, such as prolonged questioning or trickery, may also suffice.

Two reasons are given for excluding involuntary confessions. Those obtained by force are thought to be inherently unreliable evidence, and even if the confession is shown to be true by independent evidence, the police should be discouraged from resorting to force and fear to obtain their evidence.

In the United States, the confession problem has taken on definite constitutional overtones. The right of the witness not to give testimony incriminating himself is often based specifically upon state and federal constitutional guarantees. It is apparent that the privilege against self-incrimination is not the only source of the problem, however, because the established doctrine, that the due-process clause of the 14th amendment does not incorporate the privilege against self-incrimination, does not prevent that same clause from avoiding convictions obtained by the employment of involuntary confessions.

3. Illegally Obtained Evidence.—Only one of the arguments for excluding involuntary confessions is available when evidence offered in a criminal case was illegally obtained, with the most common form being articles obtained without a proper search warrant. There is nothing inherently unreliable about such evidence, and if it was obtained without any invasion of the human person the common law received the evidence despite its taint. English law retains this rule. The U.S. federal courts abandoned it, holding that excluding evidence obtained in violation of constitutional strictures on searching and seizing by government officials was the best measure of assuring compliance with those constitutional limitations. An at first small but gradually increasing number of state jurisdictions followed the federal lead. Finally the U.S. supreme court ruled that state-court use of evidence obtained in violation of constitutional search and seizure limitations was in itself a violation of due process of law under the 14th amendment to the U.S. constitution. In the early 1960s it could not yet be said that every form of illegality in obtaining evidence would compel exclusion of that evidence in criminal prosecutions, although the U.S. federal and some state courts regularly excluded evidence obtained in violation of statutory prohibitions against unauthorized interception of telephonic messages. As yet there was no developed doctrine prohibiting the use of illegally obtained evidence in civil cases, however.

4. Spousal Privileges.—English and U.S. law allow some further privileges against compulsory evidence in court. Both follow the rule that one spouse cannot be compelled to give evidence against the other in criminal prosecutions, and a number of U.S. jurisdictions follow the rule in civil cases as well. All make some exception for cases in which the crime was an offense against the spouse or the children, and the civil statutes contain wide exceptions for disputes between the spouses themselves. In addition to this remnant of common-law rules of incompetence, both English and U.S. law recognize, by statute, some right of the spouses to be free from having what were confidential communications or secrets between them exposed to public view in a lawsuit. The statutes vary in detail, but the substance of the right is similar everywhere; during or after the marriage, one spouse may prevent the other from relating the secrets passed between them during marriage. Without apparent justification, the statutes vary widely on which spouse has the right—the one who told the secret, the one who heard it or both.

The stated justification for this obstruction to the finding of truth is that the marital harmony protected and promoted by it is of more value to society than the truth suppressed would be. The point is obviously conjectural.

5. Professional Confidences.—Similar grounds are advanced to justify the many privileges for confidences communicated in professional relationships in the United States and the few such which are recognized in England and the English-law nations. The rule which closed the mouth of the lawyer at the client's behest for any secret matter communicated to the lawyer during the professional relationship (other than matter aimed at commission of future crimes) was recognized at common law and continues universally. In many U.S. jurisdictions the relationship of physician and patient and clergyman and penitent are accorded substantially identical treatment. Claims of other emerging professions have received less respect; accountants, newspaper reporters, psychologists and marriage counselors are all recognized in certain places, however.

6. Governmental Privilege.—"Crown privilege" differs significantly from other forms of privilege in that the purpose of it is not to protect any individual but the public at large or the government. And the objects to be protected and the purposes of their protection vary greatly. One version or aspect of the topic is based upon supposed want of power in the courts to compel the executive or legislative branches to respond to commands to testify, an analysis foreign to privilege generally. Recognized to some extent in the United States, the privilege works both ways; judges there have refused to recognize the power of legislatures to compel the attendance of the former at the hearings of the latter.

But except at the very highest levels of government, the lack-of-power theory is replaced by positive justification of secrecy.

Protecting military security and diplomatic relationships has always been regarded as adequate ground for secrecy, and the recognition of those grounds tends to expand in times of international tension. Wholly internal considerations may also be found sufficient to justify suppression of truth. A well-recognized rule permits law-enforcement officials to keep secret the identities of their informers, at least when the informer plays no part in the offense being prosecuted, as a means of protecting and encouraging such sources. Grand and petit jurors have similar protection from compulsory disclosure of their deliberations to encourage fearless action on their part. Witnesses before grand juries are also thought to be encouraged by a widely recognized rule that their testimony is to be kept secret. And a final form of governmental interest is used to justify secrecy of tax returns, accident reports and the results of safety investigations; the interest of the government in obtaining full reports of incomes (and hence full taxes) and in getting information which will permit the avoidance of future accidents by taking safety precautions is often recognized by statutes providing secrecy for such things. Although courts do recognize and apply the statutes, they have been little disposed to enlarge this or any other field of privilege on their own initiative.

A significant aspect of these "crown privileges" is that the information may be suppressed at the behest of the government even when the possessor of the information has every wish to disclose it. Indeed, very often the judge will act on his own initiative to prevent disclosure. A much discussed but difficult to detect difference between U.S. and English practice is found in the idea that English judges respect a claim of privilege by the government minister without inquiring into its justification, while U.S. judges treat the question as one for the court alone.

7. Semiprivileges.—A few fields of semiprivilege may properly be included here. Certain conduct which the law wishes generally to encourage might be deterred if that conduct could later be proved against the actor to establish his liability. Specific and common examples are the making of offers to settle or compromise claims, the repairing or improving of premises following an accident or injury, and the providing of medical assistance to one injured in an accident. Action of each of those kinds at least suggests that the actor recognizes some fault or liability on his part. But the suggestion is of varying strength, and many courts have excluded such evidence by stoutly insisting that it proves nothing at all. More frank is the recognition that it does prove something at times, but that even when it does, it will be excluded for fear that admitting it now will deter others in the future from taking such favoured steps.

See also CRIMINAL LAW; *Criminal Procedure.*

VI. WRITINGS

Three slightly related legal doctrines applying to writings or written instruments have been the subject of confusion in the past and will doubtlessly cause trouble in the future also. One is a policy expressed in statutes of frauds that certain transactions —examples are certain transfers of real estate or contracts of long duration—are unenforceable unless reduced to writing and signed. While such statutes are prompted at least in part by fear of manufactured evidence, the laws do not function as rules of evidence and are regarded as substantive in character.

There was once a rule requiring litigants to produce the best evidence available on any given point. Though there is no longer such a rule, there is one called the "best-evidence rule," which requires that when the contents or terms of a writing are in issue the original document must be produced to prove those contents or terms. Because it relates in modern times only to writings, it would be better named a documentary-original rule. The evident purpose is to assure certain knowledge of the wording of the instrument, for vast differences in result may turn upon slight variations in phrasing or punctuation.

Applied literally, however, the rule would be needlessly obstructive, and certain practical limitations are almost a part of the rule itself. It would be unnecessary to produce an automobile in court to prove the digits written on its license plate. It is not necessary to produce originals when carbon copies of the original are offered; and, increasingly, photostatic copies are accepted as the equivalent of the original. Unavailability of the original may necessitate proof of the terms of the writing by what is called secondary evidence, such as testimony by persons who have read the originals, or by the production of copies. Courts accept such secondary evidence freely when the failure to produce the original is explained by showing its unintentional loss or destruction, or that diligent search has failed to disclose it, or that it is in the hands of persons who are unwilling to, and cannot be compelled to, surrender it.

The courts have on the whole applied this rule in a manner designed to accomplish its purpose—to discover with certainty the terms of the instrument. But there is some inclination to transfer to it at times the purposes of the Statute of Frauds (passed in England in 1677 and copied substantially in most U.S. jurisdictions) and to suggest that the best-evidence rule requires a writing, which it does not. And there is also inclination to confuse its purpose with that of the so-called parol-evidence rule.

A very common statement of the parol-evidence rule is that it precludes the showing of prior or contemporaneous agreements (whether "parol," *i.e.,* oral, or in writing) to vary, modify or enlarge the obligations of the parties set out in a written agreement or instrument. The widely accepted explanation of the doctrine is that parties who reduce their agreement to writing "integrate" in it the entire understanding and agreement between them, thereby superseding all previous agreements and negotiations relating to that subject matter. Evidence of prior or contemporaneous agreements is thus excluded not because of any deficiency of quality as evidence but because, even if proved, it would not change the legal relationship established by the writing. On this analysis, the rule is less one of evidence than of property and contract law. It does not preclude receipt of evidence of collateral agreements between the same parties with regard to other subject matter, nor does it exclude evidence of the relationship of the parties or of the setting in which the document was executed, for these facts often are necessary to interpret the language employed. However, the use of such evidence must be guarded to avoid making an agreement different from the one written. One important exception which frequently serves as a safety valve against injustice is the right to obtain reformation of a deed or contract when it is established that the written terms are the result of a mistake in reducing the oral agreement to writing—a "scrivener's error," from the name of the calling which specialized in preparing documents for others.

The three rules, then, bear this relationship to each other. The Statute of Frauds requires certain transactions to be written. The best-evidence rule provides that the original writing, whether or not one required by the Statute of Frauds, is the preferred evidence of the wording and terms it contains. The parol-evidence rule provides that, in cases within its terms, the writing is the only agreement between the parties. Hence, what the best-evidence rule makes the best evidence, the parol-evidence rule makes the only evidence.

VII. BURDEN OF PROOF AND PRESUMPTIONS: QUANTUM OF EVIDENCE

1. Burden of Proof.—Reconstruction of past events is nearly always incomplete, and the courts are constantly confronted with uncertainty of fact even after all the evidence is in; the evidence will seldom be conclusive and will occasionally be equally balanced for and against the proof of a particular fact. While the law never requires conclusive proof, it will not act unless the evidence is capable of disturbing an equal balance. The device for allocating the risk that the evidence will be insufficient to persuade is the burden of proof. The allocation is made for reasons of legal logic plus consideration of such things as ease of proof, accessibility to sources of evidence and the favour with which a particular result may be viewed. In general terms, the prosecution bears the burden on every issue in a criminal case, while in civil cases the plaintiff bears the burden on more issues than does the defendant. As an example, the doctrine of contributory negligence (that one injured through another's negligence cannot recover

if his own negligence contributed to his injury) is treated in two different ways. In a minority of U.S. states, a plaintiff must plead and prove his freedom from contributory negligence, while English law and a majority of U.S. jurisdictions make it an affirmative defense to be pleaded and proved by the defendant.

Unless the party who bears the burden produces enough evidence to permit a reasonable mind to find that the fact exists, he must fail for want of proof; if he produces barely enough to meet that burden, he normally will be entitled to have the issue submitted to the jury at the end of the case. They will be instructed that he bears the burden of persuading them of the existence of the essential fact, in the sense that in the light of all the evidence, by whomever produced, they must believe the fact to be true in order to return a verdict for that party. They will also be told what measure or degree of conviction they must hold. Three recognized degrees of persuasion are: (1) by a preponderance of the evidence, the usual standard for civil jury cases, and which means in substance that existence of the particular fact is more probable than not; (2) by clear and convincing evidence, applied often to issues involving charges of fraud or mistake, and which can be further defined only by saying that it is higher than a preponderance but less than beyond a reasonable doubt; and (3) beyond reasonable doubt, which applies almost exclusively in criminal cases and which is often said to be that degree of certainty which men insist upon attaining in the conduct of their most important affairs.

2. Presumptions.—Presumptions are legal devices sometimes employed to alter the usual operation of the rules of burden of proof here stated. They are assumptions of fact required by law to be drawn from the establishment of other facts. For example, one of the classic presumptions is that of death from seven years of unexplained absence. One charged with the burden of proving death might prove that the subject disappeared without explanation more than seven years before and had since been unheard of. Standing alone, this would discharge the burden of proof in the sense that reasonable men might find death from those facts alone. It would not compel a jury to so find. But when re-enforced by the presumption, those facts compel a finding of death if they are uncontradicted.

When evidence is produced which tends to explain the disappearance, the truth of these facts must go to the jury: it may choose to disbelieve the contradicting evidence, in which case the presumption remains operative and a finding of death follows from the unexplained absence. The real difficulty in the area of presumptions, and the source of an enormous amount of dispute, arises when there is direct testimony which contradicts not the basic facts but the presumed fact itself—as by producing a witness who testifies that he saw the subject only the day before in a nearby city. A substantial body of U.S. case law as well as scholarly authority take the position that the presumption has vanished entirely. And this supposedly is the British view as well. Fewer U.S. courts but equally respectable scholarly sources take the contrary view—that the presumption persists and that the jury should be told of it as an assistance to them in evaluating the evidence. Much of the dispute is purely verbal, but there are also important and difficult questions of policy involved. It is overly easy to dispose of the U.S. division of authority as just one of those many cases in which different rules obtain. That this is deceptive is shown by the English-law situation regarding presumptions; cases appear to be in diametric conflict, eminent scholars offer different solutions and analyses, and an outstanding authority describes British presumptions law as of "hopeless perplexity."

Whatever view be adopted about the continuance of presumptive force after contrary evidence is received, it is apparent that the disputants operate on a common assumption that presumptions give to the evidence affected by them a somewhat arbitrary probative force or weight. This is well illustrated by the standard U.S. view that presumptions may not be employed in a criminal prosecution to discharge the state's burden of proof unless the presumption involved is one of high "rational connection" between the basic facts and the presumed facts. The rule that guilt must be proven beyond reasonable doubt requires that many of the policy justifications for presumptions be inoperative in the criminal field.

Two common misuses of the term presumption deserve mention. One is the often-stated presumption of innocence; this is but an unfortunate way of saying that the state must prove guilt beyond reasonable doubt. The other is the "irrebuttable" presumption, which is only a masquerade for a rule of law and which is not subject to the rules discussed above.

Presumptions are of great variety. Their creation usually attends some recurring situation involving great difficulty of proof for one who bears the burden of proving a fact. They are justified, variously, by the high probability of the inference involved, convenience in administration of justice and the promotion of desired results. Often, all three justifications combine. Evidence scholars are adamant in the view that presumptions do not "shift" the burden of proof. It is obvious that they do much to lighten it, however.

Perhaps the great confusion regarding presumptions makes it fortunate that this is the only significant instance in which Anglo-American law has attempted to formulate rules about the probative force of evidence, preferring to require only that it have some minimal value to be received and leaving its weight to the finder of fact. Thus the direct evidence of a single witness is ordinarily sufficient to find a fact; if the evidence is circumstantial, the finding will be sustained if the inference involved is one that reasonable men could draw, even if other inferences may also be drawn. (An instruction sometimes given in criminal cases is that, in order to sustain conviction on circumstantial evidence, the inferences must be inconsistent with any finding but guilt.) The exceptions to this basic proposition are so few that they can be named. Treason must everywhere be proved by two witnesses; in a few places, perjury must also. In England and many U.S. states, the testimony of an accomplice is not alone sufficient to sustain a criminal conviction; while admissible, it must be corroborated. So, curiously, must the confession of the accused himself; the fact that an offense has been committed—the so-called corpus delicti—must find independent support in the evidence. To a lesser extent, the rule of corroboration is applied to complaining witnesses in cases involving rape or other sexual violations.

VIII. APPLICATION OF EVIDENCE LAW

Apart from the few specialized cases requiring corroboration of certain testimony in criminal proceedings, the rules are substantially the same in criminal and civil trials. Indeed, even these rules of corroboration are often denied the name of evidence law because they govern not admission or exclusion of the evidence but only the subsequent use to which it can be put. The vaunted special protection afforded the criminal accused in Anglo-American law is accomplished less through rules of evidence framed especially for his benefit than by giving him immunity from compulsory testimony and by imposing on the prosecution the burden of persuading the finder of fact on every issue and beyond reasonable doubt.

While the concurrence of criminal and civil practice is understandable, there is reason to doubt the wisdom of imposing the rules developed in judicial context upon every manner of fact-inquiry known to modern government. It is easily forgotten that it was only the law courts which functioned under these rules of exclusion; courts of equity and other official tribunals in English and early American history found no apparent need for such detailed restriction on their practice or for such elaborately formulated rules of exclusion. That equity's evidence rules differed from the rules of the law courts is far less significant than the fact that it functioned primarily without invoking any rules at all. That the legal rules expanded their grasp to more and more areas of dispute with the passage of time is often urged as a tribute to their intrinsic merit. It is more likely due to the dominance of the law-court judges in the merger of the courts and proceedings; it was not so much that the influence of legal rules expanded as it was that other types of proceedings disappeared. As the law courts absorbed these other jurisdictions, the rules of evidence enveloped them also.

The 20th century has seen a reversal of that trend. Many questions once judicial have been transferred to tribunals which proceed according to other forms, and many new questions created by a more active government and a more complex society have been entrusted to administrative agencies, commissions and other nonjudicial hearing bodies. Skeptics have suggested that one of the reasons for this bypass of established courts is a desire to avoid the ponderous and complex body of rules of evidence. Whether this desire is the chief motivation or not, legislative creation of such bodies almost invariably includes the provision that the agency is free of the "technical" rules of evidence. In the course of reviewing decisions of the agencies, the courts originally honoured such legislative provisions only grudgingly, but have become more indulgent with the passage of time.

This newer indulgence is well-founded. It seems based on increasing recognition that unwillingness to abide by the judicial rules does not necessarily imply condemnation of those rules in their context but rather an unwillingness to concede universal validity to what is at most a concession to circumstance. The prime circumstances shaping the exclusionary rules are the highly adversary character of the proceedings in law courts and the presence of the jury. The absence of the jury in the agency hearing means the withdrawal of a powerful force in the shaping of the judicial rules of evidence. And many agency proceedings are far less adversary in character than are judicial proceedings; the tribunal functions, as it should, as more than a mere umpire. In the absence of those shaping forces, the attention of the agency may properly be directed to considerations of the probative force and worth of evidence—the ultimate consideration in any proceeding, judicial or otherwise. (*See* also ADMINISTRATIVE LAW.)

The future of judicial evidence seems likely to be but an amplification of what has just been said of nonjudicial evidence. Improvement in the quality of jurors and the decreasing use of juries in civil cases give some promise that the great emphasis upon admission and exclusion which has been the concern of evidence law may yield to another question: What does this offered evidence prove, and how strongly does it prove it? The inability—even constitutional inability—of a judiciary subordinated to juries on questions of fact to pose that question at the end of the case has compelled them to ask it throughout.

BIBLIOGRAPHY.—*Sources:* National Conference of Commissioners on Uniform State Laws, *Uniform Rules of Evidence* (1953); American Law Institute, *Model Code of Evidence* (1942). *Treatises: U.S. Law:* B. Jones, *The Law of Evidence, Civil and Criminal*, ed. by S. A. Gard, 5th ed. (1958); C. T. McCormick, *Handbook of the Law of Evidence* (1954); J. M. Maguire, *Evidence: Common Sense and Common Law* (1947); E. M. Morgan, *Some Problems of Proof Under the Anglo-American System of Litigation* (1956); E. M. Morgan *et al.*, *The Law of Evidence: Some Proposals for Its Reform* (1927); J. B. Thayer, *Preliminary Treatise on Evidence at the Common Law* (1898); J. H. Wigmore, *Treatise on the Anglo-American System of Evidence in Trials at Common Law*, 3rd ed. (1940). *English Law:* R. W. Baker, *The Hearsay Rule* (1950); R. Cross, *Evidence* (1958); G. D. Nokes, *Introduction to Evidence*, 2nd ed. (1956); S. L. Phipson, *Law of Evidence*, ed. by R. Burrows, 9th ed. (1952).

Essays: U.S. Law: K. C. Davis, "Evidence Reform: the Administrative Process Leads the Way," 34 *Minnesota Law Review* 581 (1950); G. F. James, "Relevancy, Probability and the Law," 29 *California Law Review* 689 (1941); C. T. McCormick, "The Parol Evidence Rule as a Procedural Device for the Control of the Jury," 41 *Yale Law Journal* 365 (1932); B. Meltzer, "Required Records, and the Privilege Against Self-Incrimination," 18 *University of Chicago Law Review* 687 (1951); E. M. Morgan, "A Suggested Classification of Utterances Admissible as Res Gestae," 31 *Yale Law Journal* 229 (1922). *English Law:* Z. Cowen and P. Carter, *Essays on the Law of Evidence* (1956); Lord Denning, "Presumptions and Burdens," 61 *Law Quarterly Review* 382 (1945); H. A. Hammelman, "Hearsay Evidence, a Comparison," 67 *Law Quarterly Review* 67 (1951); H. Street, "State Secrets—a Comparative Study," 14 *Modern Law Review* 121 (1951). (R. E. DE.)

EVIL EYE. The ancient belief that certain persons can injure or kill with a glance remains a widespread phenomenon in folk culture as well as in primitive societies. The power was often supposed to be involuntary (*cf.* Deut. xxviii, 54); and a story was told of a Slav who, afflicted with the evil eye, blinded himself in order to avoid injuring his own children (K. W. Woyciki, *Polish Folklore*, trans. by F. H. Lewenstein, p. 25). In ancient Rome the "evil eye" was well recognized, and special laws were enacted against injury to crops by incantation, excantation or fascination.

Children and young animals were thought to be specially susceptible. Sacred writings or amulets were worn to deflect the evil eye or to attract the malignant glance onto themselves; the latter often were grossly obscene. Among the Greeks and Romans, spitting was a common antidote to the poison of the evil eye; and certain gestures were also regarded as prophylactics on meeting the dreaded individual. Because envy was believed to be the *raison d'être* of the evil eye, it was considered unlucky to have oneself or one's belongings praised, so that the use of some qualifying phrase such as, "as God will," or "God bless it" became common in medieval Europe.

Among some modern Turks and Arabs, horses and camels are believed subject to the evil eye. Texts are worn as amulets upon the person or hung upon camels and horses by Arabs, Chinese, Ethiopians and many other peoples. Among the Neapolitans, the fear of the *jettatore*, or owner of the evil eye, still seems to retain its potency. Protective amulets, most commonly in the shape of horns, frogs, moons and hands are worn; and formerly streets would clear at the approach of a *jettatore*.

Modern Egyptian mothers may excuse the sickly appearance of their babies by blaming it on the evil eye. In some Turkish villages passages from the Koran may be found painted on the outside of houses to preserve the inmates. Survivals of this widespread superstition are found in western Europe as well as other parts of the world. The evil eye seems to have little place in the new world among peoples of European descent; it may have been lost, like so many other old world folkloristic practices, in the ocean crossing. However, a similar belief was known among some American Indians.

It is interesting to note that although men are now more commonly attributed with the evil eye, in former times, women were the common possessors.

BIBLIOGRAPHY.—Johannes Christian Frommann, *Tractatus de fascinatione novus et singularis*, etc. (1675); F. T. Elworthy, *Evil Eye* (1895); R. C. Maclagan, *Evil Eye in the Western Highlands* (1902); E. Thurston, *Omens and Superstitions of S. India* (1912); Herklotz, *Islam in India*, ed. by W. Crooke (1921); R. E. Enthoven, *Folklore of Bombay* (1924); W. Crooke, *Religion and Folklore of Northern India* (1926); Edward S. Gifford, *The Evil Eye: Studies in the Folklore of Vision* (1958). (T. A. SB.; X.)

EVLIYA CHELEBI (1611–1683), Turkish author of a famous book of travels, was born in Istanbul, son of the chief court jeweler. Educated in the palace school and by private tutors, he later served as a member of the palace staff. Under the patronage of the grand vizier Melek Ahmed Pasha, his relative, and favoured by Sultan Murad IV, he traveled all over the Ottoman empire, and his elaborate account of his travels, *Seyahatname* (10 vol.; publ. 1898–1939), is a valuable source for its history and geography, its way of life and traditions.

The owner of a vivid imagination which sometimes mixed fact and fantasy, he was also a keen observer and his minute descriptions reveal both insatiable curiosity and a scholarly mind. Written in a simple and spontaneous style, for long ignored by the purists, his account is a good example of classical Turkish prose.

See A. A. Pallis, *In the Days of the Janissaries* (1951); R. F. Kreutel, "Ewlija Čelebis Bericht," in *Wiener Zeitschrift für die Kunde des Morgenlandes*, 51 (1950). (F. I.)

EVOLUTION, COSMIC: *see* COSMOGONY; EARTH: *Origin, Composition and Age;* GEOCHEMISTRY; STAR: *Stellar Evolution.*

EVOLUTION, MATHEMATICAL, the operation of finding a root (*q.v.*) of a number. The Greeks found the square root of a number by a method similar to the one given in elementary arithmetic and algebra books, and this method passed on to or was independently developed by the Hindus. The finding of higher roots attracted the attention of various writers of the 16th century, the method being based upon the expansion of a binomial. (*See* BINOMIAL THEOREM.)

The word "root" (Latin *radix*) is due to medieval translations from the Arabic, whence came the idea of "extracting the root," and thus Robert Recorde (1510?–58) speaks of "pulling out" a

root. The symbol for root is $\sqrt{\ }$, as in $\sqrt{2}$. For roots higher than the square root a superscript symbol called the index is written, as in $\sqrt[3]{5}$ and $\sqrt[4]{9}$. The index denotes the principal root. Every nonzero number has n nth roots, only one of which is indicated by the symbol $\sqrt[n]{\ }$. For example, the three cube roots of 1 are 1, $-\frac{1}{2} + \frac{1}{2}\sqrt{-3}$ and $-\frac{1}{2} - \frac{1}{2}\sqrt{-3}$, but its principal cube root is 1.

EVOLUTION, ORGANIC. The speculations of ancient Greece on the origin of man and animals foreshadows dimly the modern theory of organic evolution. In detail, however, these speculations seem almost as fantastic as the abrupt metamorphoses of men and animals that occupy an important place in the folklore of all parts of the world. Evolution, as we understand it, is a process of cumulative change. The aspect of relative persistence, however, is quite as important as that of change. In the development of a sound concept of the origin of forms of life, the first step probably had to be a firmly grounded belief in the autonomous reproduction and permanence of species.

Evolution is treated here as based on the properties of already existing life from the first self-reproducible substance, which itself emerged most remarkably in nature eons ago in primeval seas. For a discussion of this development of life from nonlife *see* BIOLOGY: *The Biosphere and Its Inhabitants*. For a discussion of the evolution of man, *see* MAN, EVOLUTION OF. The present article is divided into the following sections:

I. History
  1. Early Thought
  2. Darwin and Later
  3. The Species Concept
II. Evidence of Evolution
  1. Structure
  2. Development
  3. Function
  4. The Rock Record
  5. Geographical Distribution
III. Course of Evolution
  1. Relative Speed
  2. Trends
  3. Genetic Basis
IV. Theories of Evolution
  1. Orthogenesis
  2. Inheritance of Acquired Characters
  3. Mutation Theory
  4. Statistical Transformation of Populations
  5. Splitting of Species
V. Genetic Interpretation of Major Evolutionary Processes

I. HISTORY

1. Early Thought.—Among the Greeks, Aristotle and Theophrastus, in the 4th century B.C., began the patient accumulation of data on animal and plant life on which sound inductions might ultimately be based. But even Aristotle held that certain highly organized animals were generated spontaneously. Zoological writings of the middle ages were full of recipes for the generation of such forms as flies, bees and even mice from nonliving constituents. The experimental demonstration by F. Redi, in the 17th century, that maggots do not appear spontaneously in spoiling meat but come only from eggs of flies like those into which they themselves develop, was a landmark in biological thought. It was not until the 19th century, however, that carefully controlled experiments of Pasteur put an end to the belief in the spontaneous generations of observable species of microorganisms.

The modern conception of species began with John Ray (1627–1705) and reached a first culmination in the system of classification of Linnaeus. Linnaeus began with the belief that he was cataloguing creation in naming and defining the species of plants and animals and grouping them in genera. Later he questioned whether the genera might not be the separately created forms, an admission of some evolutionary change, but to most of his immediate followers the absoluteness of species was a basic dogma.

As noted, this was probably a necessary step in biological thought. There were, however, a few of the more philosophical of the biologists of the 18th and early 19th centuries who pointed to the implication of a continuity of descent in the "natural" system of classification that was being arrived at, and speculated on possible mechanisms of transformation. G. L. Buffon suggested a direct molding influence of the environment, perpetuated by heredity, and Erasmus Darwin and J. B. Lamarck both advocated evolution through the inheritance of acquired adaptations.

The great weight of Georges Cuvier, the founder of vertebrate paleontology, was thrown against such speculations. He discovered numerous bizarre prehistoric forms of life, but these did not suggest stages in the evolution of modern forms. They seemed rather to justify the concept of a succession of cataclysms, each followed by creation of a new fauna and flora. Étienne Geoffroy Saint-Hilaire tried to combine the ideas of continuity of descent with discontinuity of form by the hypothesis that new species and higher categories start from the occasional appearance of monsters capable of flourishing in an appropriate environment.

2. Darwin and Later.—Glimmers of the over-all pattern of evolution were caught by the naturalists Patrick Matthew and Edward Blyth in the 1830s. But it was reserved for Charles Darwin to win general acceptance of evolution, not only on the part of scientists but also the educated public, by showing in a most thorough and objective analysis of the data from all fields of biology how all of these fell into order under the organizing power of this concept. Moreover, Darwin, simultaneously with A. R. Wallace, put forward a simple, logical, possible explanation of the process, which was of much wider applicability and greater plausibility than those previously advanced. They pointed out the inevitability of evolutionary change in a world in which each species tends to produce more individuals than can find a place, provided there is sufficient variability, at least part of which is transmissible to the next generation, to give material for a natural selection within the species, comparable with the artificial selection practised by animal and plant breeders.

A great variety of evolutionary theories was formulated during the latter part of the 19th century. The importance of geographical isolation in the splitting of species was advanced from different viewpoints by M. Wagner and J. T. Gulick. The former laid most emphasis on the effects of different environments on isolated forms, while the latter pointed out the inevitability of divergence even without difference in environment. These factors may be considered as supplementary to Darwin's theory. Other authors proposed alternative theories. Among these was Karl Nägeli, who postulated that an inner directive force guides the course of evolution largely independent of the environment. T. Eimer also held that evolution tends to proceed along direct lines not necessarily adaptive, but attributed this "orthogenesis" to the direct molding influence of the environment. E. D. Cope led a revival of Lamarckian views based largely on paleontological observations. Finally may be mentioned the revival of the hypothesis of the abrupt origin of species by mutations by Hugo de Vries, based on experiments that will be referred to later.

The weakest link in all of these theories, including Darwin's, was the lack of detailed knowledge of the laws of heredity and variation. Thus the complaint was often made that the theory of natural selection was essentially negative: it accounted for the extinction of some forms, persistence of others by the "survival of the fittest," but threw no light on the "arrival of the fittest." Unknown to Darwin, the key that was to transform the subjects of heredity and variation from the most mysterious in biology to the most deeply understood was in process of discovery in the half dozen years after the publication of his *Origin of Species* in 1859. The results of the beautiful series of experiments with the garden pea conducted by Gregor Johann Mendel in the late 1850s in the gardens of the monastery at Brünn in Moravia were published somewhat obscurely and did not come to the general attention of biologists until 1900, when the Mendelian principles of heredity were rediscovered almost simultaneously by De Vries, K. E. Correns and E. Tschermak von Seysenegg. The rapid development of the science of genetics after that date, at first under the leadership especially of William Bateson and T. H. Morgan, made it possible to replace the speculations of the 19th century by deductions from securely established knowledge of heredity and variation. The earlier interpretations were largely along the lines

of De Vries' mutation theory, but later the analysis of the statistical consequences of the Mendelian mechanism led to a return to views more nearly like those of Darwin. (*See* DARWINISM.)

3. The Species Concept.—Under the hypothesis of separate creation the distinctions between true species should be absolute: members of the same species must all trace to the same originally created individuals, while members of different species can have no common ancestry whatever. Under the hypothesis of evolution no such absoluteness is to be expected. The definition and characteristics of species are thus crucial questions in arriving at conclusions on the reality of evolution and on its mode, if it is occurring.

The most obvious criterion for a species is a consistent difference in kind from all other species, but this immediately encounters difficulties. Offspring from the same parents may differ conspicuously from each other, not only among human beings, domestic animals and cultivated plants, but also in wild species. These differences can often be shown to be heritable and fixable by inbreeding.

It cannot be said that the difference in kind that marks a species is necessarily greater than that which marks a mere variant within a species. The difference in such characters as colour of body, colour and size of eyes, size of wings, size and distribution of bristles, number of tarsal segments, etc., which distinguish laboratory strains of the much-studied fruit fly or vinegar fly, *Drosophila melanogaster*, are in many cases much more conspicuous than the criteria used in distinguishing recognized species of *Drosophila*. It is rather typical for similar species to differ slightly in every character studied with sufficient care but for hereditary variants to be indistinguishable except in one conspicuous respect. Some mutations affect many parts of the body, and by combining mutations the number of differences between strains can be multiplied almost without limit. Moreover, subspecies of the same species differ in the same way as (though usually less than) species of the same genus.

Other possible criteria are given by physiological characters. Thus injection of proteinaceous substances, *e.g.*, blood cells or blood serum, from individuals of one species (donor) into those of another (recipient) leads to the development in the latter of antibodies that tend to react specifically with the substance of the donor species. But this is not an absolute criterion. Marked similarity may be shown in certain serological reactions between man and the anthropoid apes, which are admitted by all to belong to different species, whereas conspicuous differences may occur in other respects even between offspring of the same parents (as in the case of the human blood groups).

Since the discovery that heredity is carried by certain minute threadlike structures, the chromosomes, present in the nuclei of cells in characteristic number and size in each species, the possibility has arisen that these characteristics might be used to define the species. It has turned out, however, that conspicuous differences in the chromosome set can occur between strains of the same species and, indeed, between offspring of the same parents, whereas in other cases the set differs little if at all in two clearly defined species.

In blue-green algae and most bacteria and in sporadic forms among higher organisms in which reproduction is wholly by fission or parthenogenesis as far as known, there is no other criterion for recognition of species than degree of hereditary difference in characters of the sort referred to above. As persistent differences are continually arising among the descendants of single individuals, the naming of species in these cases is wholly a subjective matter. In forms with biparental reproduction, however, there is another criterion that must be considered, the ability to cross breed. Perhaps these forms can be put in different groups within each of which crossing is possible but between which crossing is impossible; and these groups may be the true species. Crosses have been made by H. H. Newman between a killifish (*Fundulus heteroclitus*) and a mackerel (*Scomber scombrus*), forms that are put in different orders and are as different as any of the modern bony fishes (teleosts). The hybrids die at an early stage but develop far enough to show characters of both parents. There is every gradation from such cases in which inviable, monstrous hybrids are formed to ones in which the hybrids are vigorous but sterile. The mules and zebroids from crosses among horses, donkeys and zebras are examples of the latter. There are a few apparently authentic reports of fertile mules, and examples can be given to illustrate every gradation from rare to complete fertility of hybrids. The attempt to use any grade to delimit the true species is arbitrary and frustrated by inconsistencies. It is common to find that the sexes of hybrids differ in viability or fertility. It is also common to obtain utterly different results from reciprocal crosses.

Especially disconcerting for the attempt to delimit the true species by such means are cases in which two forms that are clearly of different species by such a criterion would each have to be merged with the same third form by the same criterion. Thus several recognized species of *Datura* (*D. stramonium*, the Jimson weed, *D. ferox* and *D. quercifolia*) produce fertile hybrids among themselves, suggesting, perhaps, that they should be considered mere subspecies of the same true species. Other recognized species, *D. innoxia*, *D. metaloides* and *D. metel*, form a similar interfertile group, none of which produces viable embryos on crossing with the first group. But *D. leichardtii* is a form that (as the seed-producing parent) produces viable fertile hybrids with members of both groups. Many other examples have been described among plant hybrids. Among animals all of these complications have been found in intensive study of hybrids of *Drosophila* by J. T. Patterson and associates, and examples of some of these can be drawn from work on other groups.

Another complication appears in many cases in which forms that produce fertile hybrids in the laboratory live in the same territory in nature with very little hybridization. This is, for example, the case with the wood mouse, *Peromyscus leucopus*, and the cotton mouse, *P. gossypinus*, in the lower Mississippi valley. In other somewhat similar cases (*e.g.*, the blue-winged and golden-winged warblers, *Vermivora pinus* and *V. chrysoptera* respectively) breeding hybrids are found in nature but are so rare that the sharpness of the specific distinction is not materially blurred in the region in which the ranges overlap. In other cases, however, an intergrading but highly variable hybrid population occurs in the region of contact.

There is only a rough correlation between the degree of difference in characters and the degree of incompatibility in crosses. There are pairs of species such as *Drosophila melanogaster* and *D. simulans* which differ so little in any character that they were not distinguished until it was discovered that no fertile hybrids (and few hybrids of any sort) could be obtained from crosses between individuals from certain populations. On the other hand *Drosophila virilis* and *D. americana* differ in many conspicuous respects and in chromosome patterns but produce fertile hybrids. Similarly such forms as the bison, yak and domestic cattle, which are so different that they are put in different genera, nevertheless produce fertile hybrids. Such hybrids have also been obtained from matings of pigeon and dove, birds that are put in different genera.

In spite of these difficulties, those who are working with those groups of higher plants and animals in which knowledge of the forms in nature is approaching completeness are usually in substantial agreement as to the species.

A classification into distinct species of the forms in one locality usually offers little serious difficulty. If, however, collections of one of the species are made at increasing distances from this locality, it is usual to find that the averages of such characters as colour, size and proportions of parts of the body change more or less gradually. No line can be found at which one kind of organism stops and another begins; yet, if the range is sufficiently large and dispersion sufficiently slow, it is likely to be found that remote collections differ so consistently that there would have been no hesitation in recognizing two species if they had been found together. Such observations led in the latter part of the 19th century to the recognition of the subspecies or geographical race as a necessary category in classification.

The subspecies concept was at first used rather sparingly, but

as knowledge of such forms as the land vertebrates approached maturity it has come to be recognized that an enormous number of supposed species described from different regions show intergradation in intermediate regions and must be relegated to subspecific status. Thus at one time 35 species of otter (genus *Lutra*) were recognized. More complete knowledge, derived from studies by B. Rensch, has reduced these to about four, each with many subspecies.

According to this concept, the species names among contemporary biparentally reproducing organisms refer to populations or groups of populations within which there is continuity of interbreeding in some sense, rather than to the kinds of organisms in the qualitative sense. As the lives of species are to be measured in terms of hundreds of thousands or even millions of years, it is not desirable to give different specific names to two populations merely because they happen to be completely isolated from each other at the moment. Ideally, the species should consist of populations, or groups of populations, within which there is no genetically determined barrier to the establishment of a chain of intergrading populations between any two components, but between which and all other populations there is such a barrier. Unfortunately it is not usually practicable to determine whether or not such a barrier exists between forms that are separated by a geographical gap.

Even with geographical continuity there are borderline cases in which it is a matter of judgment whether two species should be recognized or merely two subspecies of a simple species. If fertile hybrids are found in the region of overlap, it must be decided whether they are sufficiently numerous to make the inhabitants of this region a single, if rather variable, population; or whether they are so rare that they do not obscure the existence of two intermingling distinct populations.

If two distinguishable populations of forms with exclusive biparental reproduction occupy the same territory with little or no hybridization, this is ordinarily considered enough to justify the recognition of two species. There are, however, a rapidly increasing number of cases in which two such apparent species have been found to be connected around a circuit by a chain of intergrading subspecies. Thus two sharply distinct kinds of herring gull, *Larus argentatus*, and a form, *graellsii*, are found in the British Isles, and each is represented by a slightly different form to the northeast (Murman coast). Collections across northern Europe, Asia and North America have established continuity of intergradation from *graellsii* to *argentatus*. Thus, according to B. Stegmann, only one species should be recognized.

Other cases are known in birds, mammals and reptiles. Perhaps the most remarkable, however, is that in the gall wasps (*Cynips*), in which, according to A. C. Kinsey, an apparently continuous chain of 76 forms returns on itself repeatedly. Consequently there are two or three apparently distinct species in some localities.

Where there is geographical isolation, some indication is given by breeding tests, but these are not in general conclusive. If fertile hybrids are produced by crosses between the two forms, it by no means follows, as already noted, that contact in nature would be followed by the production of an intergrading population. The failure to produce fertile hybrids in extensive tests is relatively conclusive evidence of specific difference but only after all possible connecting chains through other forms have been tested. In most cases it is, of course, hardly practicable to make the necessary extensive breeding tests, or any breeding tests at all, and the systematist must use other criteria. Two geographically separated forms are usually considered to be merely subspecies if there is so much overlapping of characters that some adult individuals of both sexes cannot be classified without reference to place of collection; on the other hand if all of at least one of the sexes can be so classified they are considered to be different species. There may be overlapping of all directly measurable characters, but if some function of the measurements, or some difference in typical appearance, recognizable by a trained observer, makes it possible to diagnose all individuals, two species are usually recognized.

It is obvious that the recognized species cannot be expected to correspond perfectly with the ideal species as defined above. There are undoubtedly cases in which good species (in the ideal sense), living in the same territory, overlap in characters or are even practically indistinguishable except by breeding tests (*e.g.*, *Drosophila pseudoobscura* and *D. persimilis*). This well may be the case for many supposed subspecies that are geographically isolated. On the other hand, the absence of overlapping in local populations that are recognized to be merely subspecies because of intergradation through a chain of connecting forms makes it likely that many geographically isolated supposed species would merge if brought together.

The shift to the population concept of species as the ideal has lagged among those working with invertebrates, but there is reason to believe that it is widely applicable from special cases that have been studied intensively. The situation in plants is complicated by an abundance of forms within species (ecotypes) each of which occurs more or less consistently wherever the same ecological niche is found. Such differences have been found to depend in part on direct environmental influences that complicate the expression of underlying genetic differentiation more than is usually found in animals.

In some groups of plants (*e.g.*, roses, hawksbeards, raspberries, etc.) uniparental (asexual) reproduction occurs predominantly among types with various numbers of sets of chromosomes; and occasional hybridization, followed by segregation of new types, has given rise to complexes of intermingled clones that present an almost insuperable problem to the classifier.

Although sexual species can be defined more objectively than any other category in the system of classification, it must be concluded that the difficulties of definition, whether by gross morphology, by chromosome sets, by physiology, by incompatibilities in crossing, or even by failure of intergradation with other species, are quite out of harmony with the absoluteness expected under separate creation. They all point toward an evolutionary interpretation of some sort.

An important question with respect to possible modes of evolution is the adaptiveness or lack of it in differences between subspecies of the same species, or species of the same genus. It long has been recognized that species of birds and mammals with a long north-south range tend to exhibit certain rather uniform trends. In the northern hemisphere the northern forms tend to be lighter coloured (Gloger's rule), larger (Bergmann's rule) and to have relatively smaller extremities—ears, feet, tail (Allen's rule)—than those to the south. These are probably to be interpreted as selective adaptations to systematic differences in exposure to the environment.

In some cases such trends are so uniform and gradual that any separation into subspecies is wholly arbitrary, and the situation that exists is called a cline. Usually, however, widely ranging species, whether or not they show a north-south trend, are divided into subspecies that are separated by rather narrow zones of rapid transition. At least some of the differences between such subspecies are rather clearly of the nature of minor specializations in relation to different ecological conditions. Among animals, differences in concealing coloration, in relation to the prevailing colour of the soil or other environmental features, are common. Thus desert forms and forms living on white sand beaches or dunes are characteristically pale in comparison with those of humid woodlands. The observation that rodents living in the desert on black lava are exceptionally dark is the kind of exception that proves the rule. Physiological adaptations to different ecological conditions are no doubt more important than adaptive coloration, but they are more obscure. (*See* ANIMAL COLORATION.)

The boundary between forms exhibiting what seem to be clearly adaptive differences may not coincide with any natural boundary, suggesting that differential population pressures may take precedence over close adaptation. In many cases indeed, differences in minor morphological characters seem to have no relation to adaptation. This may be due in part merely to lack of knowledge. Conspicuous differences may possibly be physiological correlates of intangible but adaptively important characters. It is difficult, however, to avoid the impression that a large portion of the differences merely happen to have arisen and have no adaptive significance.

Local collections within the central range of recognized subspecies usually show differences in the means of variable characters that are statistically significant but so slight that individuals could not possibly be classified by their own characters. Here especially the differences have the appearance of being almost wholly random in character.

The differences between species of the same genus are for the most part similar to those between subspecies, though usually greater in degree. Again there are differences that are obviously adaptive in relation to diverse ecological niches. For example, the termite *Amitermes excellens* of the tropical rain forest of British Guiana constructs on the trunks of trees nests that have remarkable rain-shedding devices. Another species, *Amitermes meridionalis*, in northern Australia, constructs a vertical blade-like nest with north-south orientation, probably as an adaptation to control of temperature and humidity. The morphological differences between the workers that construct the nests are slight. It is indeed difficult to cite strikingly adaptive morphological differences at the species level though not at the level of genus or family. Differences in protective coloration are, however, very common. Among the most remarkable are those found in butterflies, concerned with mimicry of noxious species (*see* MIMICRY).

There is a large class of species differences involving colour, form and behaviour which may be interpreted as recognition marks in which mere difference in otherwise rather similar species of the same region seems to be of value to the species in maintaining their integrity. Another class of rather striking differences is that in which there has been degeneration of an organ in one of the species. The cave fish *Anoptichthys jordani* shows almost complete loss of the eyes, with correlated changes in the skull and a compensating notable increase in the number of taste buds on the head and body as compared with the otherwise very similar normal river fish, *Astyanax mexicanus*, as noted by C. M. Breder, Jr. These are classified in different genera, but the existence of a cave population with intermediate conditions indicates that the relationship is very close, perhaps merely that of two subspecies.

The overwhelming majority of the morphological differences by which species of the same genus can be distinguished are ones to which it is difficult to assign an adaptive significance. The general conclusion that seems to be indicated is that the evolutionary process at the levels of the subspecies and the species has both adaptive and accidental aspects. *See* also SPECIES.

II. EVIDENCE OF EVOLUTION

1. Structure.—Even before the acceptance of evolution it was generally recognized by biologists that there is a single natural system of classification of organisms to be distinguished from the practically infinite number of possible arbitrary systems. The basis for this system was clarified by Sir Richard Owen (1804–1892), an opponent of evolution, who distinguished between two kinds of similarity: analogy and homology. *Analogous* parts of two organisms serve a similar function and show only such similarities as are explicable by functional adaptation, supplemented perhaps by coincidence. The wings, legs and eyes of a fly are, for example, analogues of the structurally very different parts of a bird that are given the same names. *Homologues* show a similarity in structure, mechanism, mode of development and relationship to other parts that goes beyond that explicable by functional adaptation and coincidence. The corresponding legs, wings and eyes of two kinds of insect meet all of the requirements for a judgment of homology. Corresponding statements can be made in regard to the legs, wings and eyes of two different species of bird.

Comparisons of the wing of a bird, the foreleg of a frog, the foreleg of a horse, the arm of a man and the flipper of a whale, parts which have the same relation to the rest of the body in these animals, bring out a similarity in the pattern of bones, muscles and other parts that is remarkable in view of the dissimilar functions. These must all be judged to be homologues.

Such similarities are especially noteworthy where the part in question is a functionless vestige in one of the forms compared. Whales have no hind limbs that protrude from the body, but in some of them there are groups of small bones in the appropriate position within the body that are obvious homologues of the hind leg bones of land mammals. The vestigial wing of flightless birds, the vestigial eyes of blind cave fishes, the vestigial appendix of man are homologous to functional organs of other forms.

Although homology is usually treated as an all or none matter, it is obvious that the implied similarity is a matter of degree. There are always considerable differences between homologues where function is markedly different. Homology between parts of two kinds of organism may only be traceable by comparison of both with another or through a chain of other organisms. Thus there is little direct basis for considering the incus and malleus of mammals, two minute bones that aid in conveying sound vibrations from the eardrum across the middle ear, as homologues of the powerful upper and lower jaws of sharks. Yet the mammalian ear ossicles are clearly homologous to the quadrate and articulare in the fossilized skulls of cynodont reptiles. These small bones, by which the jaws articulate, are obvious homologues of the larger articulating bones of other reptiles and amphibians and teleost fishes. A comparison of the teleost jaw with that of a shark indicates that, in the former, the jaw is a composite structure in which the articulating parts are homologues of the whole jaw of the latter, while the tooth-bearing elements of the former are comparable with dermal plates or scutes. The chain is further lengthened on recognizing homology of the jaws of the shark with anterior gill-supporting structures of the jawless cyclostomes (lampreys, hagfishes).

Even in cases in which the functions of homologues are the same, there may be great differences in the degrees of similarity of structure. The wings of birds and bats are constructed on such utterly different plans for carrying out their common function, that they can be considered as merely analogous as wings. Yet the basic patterns equally clearly show an underlying homology as fore limbs. Similarly the homologies of the single-hoofed legs of the horse and the cloven-hoofed legs of the antelope can be traced best by comparing both with mammalian legs that are not specialized for running.

The statement that homologous organs show similar relationships to the rest of the body carries the implication that it is possible to determine what are corresponding parts. This can be done without hesitation with two species of bird, with more uncertainty for certain parts of bird and fish, only in the roughest sense, if at all, in the case of bird and insect, and finally not at all in that of bird and plant (except for intracellular structures). Thus it is usual to find that if two organisms show the type of similarity in one part that indicates homology, they do in others, and relations among these parts are the same. Apparent exceptions tend to disappear on analysis. Thus the eye of the octopus is incomparably closer in plan to the eye of the bird than is the eye of an insect, although homologies in other organs are no more traceable in one case than the other. But this can be accounted for by coincidence. There are only two basic optical mechanisms for an image-forming eye. It happens that bird and octopus have eyes of the camera type, whereas the insect has a system of optically isolated radiating tubes. The detailed carrying out of the plan of the camera eye in the octopus is so different from that in the bird (and all other vertebrates) that the eyes must be considered as merely analogues by this criterion itself.

In cases in which homology can be traced throughout two organisms, particular homologous organs may show a degree of similarity that far exceeds that of other organs. Thus the horses and an extinct group of South American hoofed mammals show an extraordinary similarity in the structure of their legs and their general conformation, although not notably similar in other respects. It may be supposed that coincidence in the mode of adaptation to a similar function has here supplemented the basic similarity due to homology to bring about convergence.

In some cases particular organs of two forms appear to be homologous in structure but are located in different places in bodies that are otherwise homologous throughout. The locations of the kidney elements in cyclostomes, fishes and reptiles present such

a problem. Such cases bring up the relation between the phylogenetic homology, which is under consideration, and a different sort of homology, that of similar structures in the same organism, which will be discussed later.

Because of the prevailing parallelism in the degrees of similarity of homologous parts of organisms, any serious attempt at classification by morphology leads to a certain branching order, the natural system, referred to earlier. Species can be grouped into genera within which homologies can be traced in minute details. Within the higher categories, families, orders and classes, homologies can be traced for most organs but the similarities become less. The phyla are bound together as natural groups by only a few broad homologies; in the chordates, for example, largely a dorsal neural tube, an underlying supporting structure of a peculiar type (the notochord), and pharyngeal gill slits, each present, perhaps, in only a phase of the life cycle.

Difficulties of classification arise from departures from parallelism in the similarities of homologues, which may often be related to divergent or convergent adaptation, as in cases already discussed. The parts that are least subject to change of adaptation are generally recognized to be those on which most emphasis should be placed in classification.

The definition of homology, as a similarity greater than can be accounted for by common function or by coincidence, raises the question of its explanation. Only two explanations have been proposed. The first is that homology is an aspect of the design of creation. The alternative is that it is a consequence of relatively unmodified common heredity. The branching order of the natural system of classification is expected under evolution. It is perhaps unwarranted to speculate on what sort of order would be expected under separate creation. It may be pointed out, however, that creative activity as it is known in man is not restricted to a branching order in its products. Ideas are transferable from one field to others, with the consequence that human creations are related in a multidimensional network. At best the concept of design is one that is more plausible as a presumptive hypothesis rather than as an interpretation of concrete cases such as the sporadic occurrence of vestiges that are homologous to functional organs of other forms.

If common heredity is the explanation of homology, this sort of similarity should be traceable at all stages of development. It is indeed often the case that homologies are more easily recognized in embryos and young than in adults. For example, mammals have a heart that consists of two muscular tubes that do not communicate directly. The right side receives the venous blood from the body and pumps it to the lungs, whence after aeration, it is carried to the left side of the heart to be pumped to the tissues through branches of a single great artery, the aorta. There is little apparent similarity with the circulatory system of a fish, in which the heart is a single muscular tube that pumps the venous blood through several paired aortas to gills in the walls of a series of slits in the pharynx through which water is kept moving. If, however, the embryos of a fish and a mammal are compared, there is no difficulty in establishing homologies. In both there is a simple tubular heart, paired aortic arches and gill slits. The rudiment of the mammalian lung is seen to be homologous to an organ that in some cases functions as an accessory respiratory organ, the swim bladder of the fish. The adult condition of the fish is arrived at rather directly, whereas that in the mammal is reached only by suppression of all of the gill slits except the first (the ear opening), atrophy of most of the aortic arches and the growth of a partition through the heart in such a way that all of the blood must pass through the lungs at each circuit. The mammal thus seems to pass through a fish stage. As there are other reasons for believing that all land vertebrates are descended from aquatic forms, this seems to justify the aphorism "ontogeny recapitulates phylogeny." This expression is, however, seriously misleading.

2. Development.—There is an appearance of recapitulation wherever a progressive form that has elaborated the later developmental stages is compared with a conservative form that develops relatively directly. It happens, however, that in the case of mammal and fish the progressive form has also complicated the earliest stages by producing a membrane about the embryo, the amnion, which cannot be interpreted as a recapitulation. Again evolution also may proceed by dropping later stages. The tunicates are saclike marine animals that spend their adult lives attached to rocks, filtering food particles from a current of water drawn through the mouth and passed out through slits in the wall of the pharynx. There are no characters of the adult tunicate and the adult frog to suggest homology. Yet the tunicate develops from a tadpole with much more than superficial resemblance to that of a frog. The recapitulation theory suggests a tadpolelike common ancestor, but there is considerable reason for a different interpretation: an originally simple but free larval phase of the sessile tunicate may have been elaborated to give more efficient dispersal. The new way of life thus evolved then may have taken precedence over the fixed life. The vertebrates thus may have had their origin in the development of sexual maturity in an elaborated tunicate tadpole and the suppression of the later fixed stages. (*See* EMBRYOLOGY.)

Among invertebrates there are many examples which seem to require elaboration of early stages or suppression of either early or late stages as well as ones that indicate the addition to late stages, postulated by the recapitulation theory. The homologies between adult annelid worms and arthropods leave little doubt of the derivation of the latter from the former. Yet nothing could be more different than the ciliated trochophore larvae of the marine annelids and the characteristic nauplius larvae of the marine crustaceans with their jointed appendages and absence of cilia. There is, moreover, little resemblance between the modes of development of the terrestrial and marine members of either phylum. On the other hand it hardly seems likely that the detailed similarity between marine annelids and most mollusks in the spiral cleavage of the egg, in the fate of the various early formative cells and in the similar trochophore larvae can be wholly coincidence or common adaptation. Ultimate derivation of these otherwise very different phyla from the same type of unsegmented worm is suggested. The similarity of the nauplius larvae of the barnacles to those of the lower crustaceans leaves no doubt of the ancestry of barnacles. *Sacculina*, a formless parasite of the crab, has nothing in its adult phase to suggest arthropod affinities, but its nauplius larva leaves no doubt of its relation to the barnacles.

The allocation of forms to their proper position in the natural system of classification requires consideration of possible homologies of all parts at all stages of development, with due weight to the possibilities of adaptive elaboration or simplification at any stage.

3. Function.—Classification by morphological homologies tends to carry with it a grouping by physiological similarity. It has already been noted that the proteins, or compound proteins, of species show a high degree of specificity, as revealed by serological reaction, but that this is not absolute. Attempts have been made on a large scale in both animals and plants to make classifications based on the similarities indicated by such reactions. These agree remarkably closely with those arrived at on morphological grounds. Thus antihuman serum gives a very striking precipitin reaction with serum from anthropoid apes (Pongidae), a somewhat weaker reaction with serum from old-world monkeys (Cercopithecidae), still weaker with serum from South American monkeys (Cebidae, Callithricidae) and no reaction (at the titre considered) with serum from lemurs (Lemuridae), in agreement with relationships deduced on morphological grounds.

The epidermal cells of vertebrates characteristically produce keratin, whereas chitin is the characteristic product in arthropods. Arginine phosphate plays a role in the energy transformation of crustacean muscles analogous to that played by creatine phosphate in vertebrates as far as is known. Oxygen transport in vertebrates is always based on the properties of hemoglobin. This is not the case among most invertebrates, in some of which a copper-containing protein, hemocyanin, plays a similar role. Hemoglobin does however occur in some mollusks, annelids and arthropods.

In general, however, classification primarily by physiology gives less consistent results than classification by morphology, except probably in the case of protein specificity. The reason seems to be that the basic physiological processes are characteristic of all living matter and furnish perhaps the strongest evidence that all life as it is now known has a common origin. The erratic appearance of hemoglobin as an oxygen carrier throughout the animal kingdom becomes more intelligible when it is learned that its peculiar prosthetic group, based on porphyrin, is found in oxidixing enzymes of almost universal occurrence (cytochrome, cytochrome oxidase, catalase) and also in the chlorophyll of green plants.

4. The Rock Record.—The most direct evidence for evolution comes from the actual remains of past organisms buried in sediments, the relative ages of which can be determined. As noted earlier, Cuvier was a vigorous opponent of evolution. In his day the estimates of past time allowed only a few thousand years since creation. It was obvious from old inscriptions that man and many animals had not changed appreciably in 3,000 or 4,000 years, and it was ridiculous to suppose that they could have evolved from common ancestors in a few more thousand years. It required a drastic revision of the earth's time-scale before evolution could seem a reasonable hypothesis. But such a revision was forced on the attention of science by the studies of rock formations. The examination of thousands of feet of sandstone, shale and limestone strata revealed characteristics that could only be interpreted as those of slowly accumulating deposition, interrupted from time to time by erosion. Estimates based on probable rates of deposition were in terms of many millions of years.

Up to the end of the 19th century, the minimum estimates of geologists were far in excess of the maximum that the physicists would allow for the age of the solar system on the basis of known sources of the energy radiated by the sun. When the enormously greater energy from conversion of mass became known, there was no difficulty in reconciling the estimates. It became possible indeed to determine the age of certain deposits from knowledge of the rate of transformation of uranium into lead with much more confidence than from estimates of rates of deposition. With the Cenozoic era estimated to have begun 70,000,000 years ago, the Mesozoic 200,000,000 years ago and the Paleozoic 500,000,000 years ago, the lack of apparent evolutionary change in the last 6,000 years is seen to be of no significance.

The sequence in which types appear in the strata is quite in accordance with the deductions from morphology and other lines of evidence. The first vertebrates, for example, were jawless forms in the Ordovician period; jawed fishes appeared in the Silurian, amphibians in the Devonian, reptiles in the Pennsylvanian, mammals in the Jurassic and man only in the Pleistocene—after 99.9% of this history had passed! (*See* PALEONTOLOGY: *Evolutionary Evidence.*)

The most complete evidence comes from the transformations of some of the more abundant marine invertebrates observed in thick continuous deposits (*e.g.*, the ammonite *Kosmoceras*, the brachiopod *Spirifer* and the echinoderm *Micraster*). At a given level, fossils of such a species show as wide a range of variability as is seen in a living species. In passing up through deposits that represent hundreds of thousands or even millions of years of almost continuous deposition, a gradual shifting of the mean can be observed until at length the array no longer overlaps that in the lowest stratum. Forms far apart in such series and living contemporaneously would unquestionably be considered as different species. (*See* also FOSSIL.)

In an earlier section was discussed the delimitation of the contemporary species as based in part on the absence of intergrading populations. This criterion breaks down in successive species, in which there is the same sort of continuity as in a geographical cline. But while arbitrary stages in a geographical cline need merely be considered as subspecies, a temporal cline must be divided arbitrarily into species.

5. Geographical Distribution.—Significant events in evolution would be expected to be localized at particular places as well as at particular times if dispersion over the earth has been as severely restricted as it is today for most organisms. The geographical distribution of species interpreted as more or less closely related should provide evidence on evolution. It seems, indeed, to have been evidence of this sort that impressed Darwin and Wallace more than any other. Darwin noted characteristic differences between mammalian faunas of South America and the old world for which climate could offer no explanation. South America contained many species and genera, living and extinct, of armadillos, sloths and anteaters. The whole order to which these belong is not found elsewhere (except for a few migrants into southern North America). If species (and higher categories) of armadillos were separately created, why created only for South America? Darwin saw that evolution from an ancestral armadillo under conditions of isolation from the rest of the world offered the simplest explanation.

The results of paleontological as well as neological study after Darwin's time fall into a consistent pattern on the assumption that South America was isolated from North America near the beginning of the Cenozoic and until relatively recently (Pliocene) received only a few stray forms from other continents in addition to an initial very primitive mammal population. The effects of easier exchange of North and South American forms since the Pliocene can be traced in the fossils and living forms of both continents. The even more peculiar mammalian fauna of Australia is readily interpreted by a longer and more complete isolation. The fossil record of North America and Eurasia on the other hand indicates an alternation of long periods of isolation with ones of relatively easy communication. (*See* ANIMALS, DISTRIBUTION OF; ZOOGEOGRAPHY.)

Darwin was especially impressed by the peculiar situation which he found in the Galapagos Islands. There were numerous species of ground finches found nowhere else in the world. Separate creation gave no explanation of the fact that these species were more like each other than like anything elsewhere but also more like forms from the West Indies than from the old world. Separate creation could also give no explanation of the highly unbalanced character of the fauna, which suggested instead derivation from a few accidental immigrants belonging to types that might conceivably have been carried across a large body of salt water under exceptionally favourable conditions. Darwin was irresistibly drawn to the evolutionary interpretation. Intensive study of the faunas and floras of other oceanic islands has given abundant confirmation.

III. COURSE OF EVOLUTION

Accepting evolution as a fact, it is of interest to examine the conclusions that have been reached on the course of evolution in various groups to see what generalizations are indicated on the nature of the process. It appears immediately that there has been the utmost diversity in every respect, including rate, directness of trend, progress in complexity and amount and nature of branching of phyletic lines.

1. Relative Speed.—There are existing genera for which evolution seems virtually to have ceased hundreds of millions of years ago. The longest record for stability is probably that of the brachiopod *Lingula*, which has persisted from the Ordovician to the present, with only slight change from *Lingulella* of the Lower Cambrian, some 500,000,000 years ago. There are other brachiopods (*Crania, Discrania*) and bryozoans (*Stomatopora, Berenicea*) and a bivalve (*Nucula*) which have also lived since Ordovician.

The oldest living genera of vertebrates are several of the sharks and a skate, all from the Jurassic. There are, however, the lungfish *Ceratodus* and the coelocanth, *Latimeria*, which have changed little from their ancestral genera of the Devonian and thus far antedate the great diversification of teleost fishes in the Late Mesozoic. *Sphenodon*, a reptile, has changed little since the Lower Triassic, a period within which the dinosaurs had their beginning, their spectacular diversification and their extinction. The opossum has seen the origin and differentiation of the orders of placental mammals from the Cretaceous. Among plants the horsetails (*Equisetum*) and club mosses (*Lycopodium*) are relics of the dominant groups of land plants of the Mid-Paleozoic, and certain gymnosperms, notably *Araucaria* and *Ginkgo*, represent

types of flowering plants of the Late Paleozoic and Early Mesozoic that far antedate the appearance in the Cretaceous of the present dominant flowering plants, the angiosperms.

Leaving out of account such cases of extreme stability, there are characteristic differences between the rates found in different groups. Thus G. G. Simpson estimates the average survival of the extinct genera of bivalves to have been 78,000,000 years in comparison with 6,500,000 years of extinct genera of carnivores.

Rapid as has been the evolution of mammals during the Cenozoic, it seems to have been enormously more rapid in all lines during a brief period just preceding this. The only placental mammals known in the Upper Cretaceous were insectivores. Well-differentiated carnivores, primates, edentates, rodents and lagomorphs, as well as several extinct orders of ungulates, appeared in the Paleocene, and most of the remaining orders including such divergent forms as bats and whales were present in the Eocene. Similar periods of extremely rapid evolution are indicated for fishes in the Devonian, amphibians in the Mississippian, reptiles in the Permian. The turtles, for example, seem to have evolved from generalized reptiles at some time in the Permian. By the middle of the Triassic their peculiar structure was essentially as it is today. It is characteristic of these periods of rapid transition that fossils are exceedingly rare, indicating small or sparse populations.

Much of the paleontological record consists of gradual transformations along direct courses. The transformation of Lower Eocene "dawn horse" or eohippus into the modern horse through a succession of several genera is perhaps the most complete history of this sort. (*See* EQUIDAE.)

In the long run, however, evolution has been far from a direct process. The tetrapod vertebrates hardly had more than emancipated themselves from the water as primitive reptiles than some of them, the mesosaurs, ichthyosaurs and plesiosaurs, proceeded to reverse this direction by readaptations to marine life. Similar reversals occurred in the whales, sirenians and seals among the mammals. The resulting forms were, however, very different from fishes. The peculiarities of the starfishes and other mobile echinoderms would hardly have been arrived at except by a succession of abrupt changes in direction of evolution. The same is true of the salps, Larvaceae and perhaps the ostracoderms as free derivatives of the fixed tunicates. The history of the horse from ostracoderm or even tunicate to eohippus was certainly a most tortuous process.

Direction and rate are related. Typically a lineage exhibits a zigzag course in which periods of slow advance of constant direction are separated by periods of very rapid remolding toward a new direction.

2. Trends.—A comparison of Early Paleozoic and recent forms of either animals or plants no doubt indicates increased complexity of organization as a characteristic evolutionary trend. But here also there is no absolute rule. Under certain conditions, especially parasitism, the trend is toward simplication.

Other characteristics in which the evolutionary process differs in different cases are the amount and nature of branching. The very limited amount of branching in the history of the Equidae (horse, ass, zebra, etc.) since the Eocene contrasts with the rich branching of the Bovidae (cattle, sheep, goat, etc.) in the shorter period since the Miocene. The Muridae (the so-called modern mice) are another family of mammals notable for multiplication of genera and species without much significant change of type. This is also the case with the perching birds, many groups of teleost fishes, many families of insects and large groups of angiosperms.

In further contrast is the more extreme type of adaptive radiation that leads to widely diverse types, each adapted to an utterly different way of life. This is illustrated by the rapid evolution of many types of ostracoderms in the Silurian, tracing presumably from an unarmoured lower chordate; of fishes in the Devonian, from an ostracoderm; of amphibians in the Mississippian, from a crossopterygian fish; of reptiles in the Permian, from a labyrinthodont amphibian; of archosaurs in the Triassic, from a thecodont reptile; of birds in the Upper Cretaceous, tracing through reptile-like birds to a Jurassic archosaur; of placental mammals in the Paleocene, tracing to Triassic therapsid reptiles.

The best examples of abrupt change of direction and also of extremely rapid evolution are drawn from such cases in which many changes of direction occur simultaneously. The basis is probably in all cases the achievement of an adaptation or complex of adaptations by the parent form that gives an advantage of a general sort over the forms already occupying many ecological niches. Adaptive radiation in the cases cited above seems to have been based on armour in the ostracoderms; jaws in the fishes; the leg and five-toed foot in the amphibians; the amnion and other devices giving emancipation from the water in the reptiles; effective locomotion on land in the archosaurs; wings, feathers and temperature regulation in the birds; effective limbs, hair, temperature regulation and mammary glands in mammals. The primary adaptive complex may or may not have been a rapid achievement itself.

A corollary of adaptive radiation is the extinction of outmoded types. The character of the animal and plant life of the world has been revolutionized by the extinction or near extinction of such once dominant groups as the trilobites, eurypterids, ammonites, ostracoderms, placoderms, palaeoniscids, labyrinthodonts and archosaurs among animals; Equisetales, Lycopodales and pteridosperms among plants; and the expansion of the dominant groups of today from apparently insignificant beginnings. Only rarely does a once dominant group that has been supplanted by another, as the Permian and early Triassic therapsids were by the archosaurs, leave a specialized line that later supersedes its supplanters as the mammals did after an interval of 100,000,000 years.

Summing up, the characteristic evolutionary process may be described as the emergence of a complex of adaptations of general significance, the rapid exploitation of this in diverse ways of life by adaptive radiations at successively lower levels, leading ultimately to gradual orthogenetic advance along each line, accompanied in some cases by extensive diversification of genera and species with jointly nonadaptive and minor adaptive aspects. On rare occasions a new, relatively general, adaptive complex may emerge at any stage in the process, initiating a new cycle. The broad course of evolution has the appearance of being guided by selective expansion and elimination among the higher categories.

However, in spite of the relative abruptness of changes in direction and the rapidity of evolution in certain cases, the nature of the changes in morphology indicates essential continuity. The possibility of adaptations to widely different ways of life at different stages of the life cycle (tunicate, amphibian), or in an alternation of generations (alga, bryophyte, hydromedusan jellyfish), provides a basis for the accomplishment of violent changes in the direction of evolution in a continuous manner by a mere shift in emphasis.

The reorganization of the circulatory system of the fishes, required in land vertebrates by the shift from gills to swim bladder as the primary respiratory organ, did not occur by an abrupt change in plan. It required hundreds of millions of years to reach the degree of adaptation to land life comparable with that which existed for marine life in fishes. All the changes in the heart and blood vessels were of the nature of differential growth or suppression, actually repeated in the individual careers of higher vertebrates. Similarly a comparison of the reptilian jaws, articulated by a chain of bones, and the mammalian jaw, with direct articulation at a different point, at first sight seems to require change within a single generation. Yet the conditions found in the jaws of the early mammallike reptiles show that a continuous series of intermediate conditions actually occurred.

New organs characteristically make their appearance as simple outpocketings or outgrowths (*e.g.*, swim bladder, amnion, horns, etc.) or as readaptations of old organs (*e.g.*, teeth, lungs, legs, feathers, wings, etc.). It is very common for an indefinite number of simple, undifferentiated elements to become transformed into a definite number of complex, individualized structures. The gill arches appear to be very simple and variable in number in lower chordates (tunicates, amphioxus, the jawless fish). The number seven becomes standardized in the fishes. The first arch becomes differentiated as a jaw and later, as already noted, becomes the articulation for the jaw and finally (in mammals) reduced in size

and changed in function to become a chain of ear ossicles. The second (hyoid) arch also becomes somewhat specialized, while the posterior arches largely disappear with the loss of the gill apparatus. The dermal bony armour of some of the jawless fishes gives rise not only to the bony scales of higher fishes but also to the highly individualized dermal bones of the skull and to the teeth. The latter are simple, undifferentiated and highly variable in number and location in lower vertebrates. Beginning in therapsids and carried on by the mammals there is individualization of incisors, canines, premolars and molars and a tendency toward fixation of number. The acanthodians among the earliest jawed fishes show various numbers of paired ventral fins (up to seven) as well as unpaired dorsal and anal fins. Two pairs of appendages become standardized in the higher vertebrates. The five-toed foot of the amphibian emerges from a lobe fin with an unstandardized number of branching and radiating lines of small bones. The same principles can readily be illustrated from invertebrates. The indefinite number of similar segments of the annelid, apparently arising by transformation of a method of asexual reproduction into one for producing a compound single organism, becomes in the arthropods a more or less definite number of individualized segments, with more or less individualized appendages.

Occasionally the reverse process occurs. The dental pattern of three incisors, one canine, four premolars and three molars seems to have become standardized in the ancestral placental mammals and thereafter evolved largely by differentiation and loss of elements. In the toothed whales, however, dedifferentiation has occurred and a return to an indefinite number of simple undifferentiated teeth.

Structures of wholly different origin, but which happen to be close together, may become merged in a new complex. There is little similarity between the cartilaginous cranium that protects the brain and sense organs of the lowest fishes and the bony scutes in the skin. In the higher vertebrates, the former becomes ossified and the latter become associated with it to form the compact cranium in which cartilage and dermal bones can be distinguished only by their modes of development. The complex mammalian ear integrates a number of relatively unrelated structures of lower vertebrates.

3. Genetic Basis.—As evolution implies change of the hereditary characteristics of a species, an adequate explanation of an evolutionary process must be expressed in terms of mechanisms of heredity. The basis for existing views on this subject was laid by the discovery at about the beginning of the 19th century that all higher plants and animals are composed of essentially similar entities, the cells, each of which has the essential properties of an organism on its own account. Experimental evidence of the most direct sort has made it clear that the essential persisting organization of cells (at least in all forms above the bacteria and blue-green algae) resides in the threadlike chromosomes of the cell nucleus. Before each ordinary division (mitosis) each of the chromosomes duplicates; thus each daughter cell receives a set of chromosomes identical with that in the parent cell. As the egg and sperm both contain the set characteristic of the species, the fertilized egg has a double set. This is maintained by mitosis in all cells of the body (animals, plant sporophytes). However, in the formation of reproductive cells in animals and of spores in plants, there are typically two peculiar divisions, constituting the process of meiosis, in which the normal number of chromosomes is halved (the haploid condition). These cells are the germ cells in animals. In plants there is more or less multiplication in this phase (gametophyte generation) before the germ cells are formed by mitosis. The union of two germ cells restores the double (diploid) set and completes the cycle.

While the mechanisms by which the set of chromosomes is maintained intact through cell division and cell fusion are extraordinarily precise, irregularities or mutations occur from time to time, resulting in genetic change and giving a possible basis for evolution. These mutations, which are induced by various agencies such as certain chemicals and certain radiations, may affect not only the normal diploid number of chromosomes (and necessarily likewise increase the number of the genetic determiners or genes strung along the chromosomes) but also the quality. Chromosomes may be broken apart, rearranged, united to other chromosomes, etc., in seemingly infinite variations. Equally important, the genes themselves may be completely refashioned by mutations, so that their influence on body structure and function is more or less altered. (*See* HEREDITY; GENETICS.)

The relation between gene and character is important in considering the mechanism of evolution. The concept of homology seems to imply a separate heredity for each part of the anatomy that can be homologized in related species. Thus one of the objections formerly raised against the theory of evolution by natural selection was the difficulty in conceiving of adequate simultaneous selection of all parts of the body. It was thought, for example, that the right and left eyes of vertebrates must have been perfected independently. The complicated adaptive architecture of each bone was thought to have its own separate evolutionary history. This implies the old doctrine of preformation. The evidence from genetics indicates, on the contrary, that every cell in the body has the same array of genes, but that it is an array that gives the cells a wide repertoire of possible responses in relation to different local conditions.

The closest approach to a one-to-one relation between gene and character is in the determination of elementary metabolic processes of cells. In the mold *Neurospora*, for example, each of seven successive steps in the synthesis of the amino acid arginine is under the control of a separate gene. There is some reason to suppose that the only direct action of genes apart from self-duplication is in regulating the pattern of macro-molecules, *e.g.*, protein, some of which as enzymes are the primary regulators of metabolism.

The relation of genes to gross morphological characters with which the evolutionist is largely concerned must be very indirect. The localization of gene action must depend not on localization of the genes but of the conditions under which their actions are invoked. Action may be restricted to a particular tissue, *e.g.*, epidermis, pigment cell, blood cell. In some cases, it may be invoked wherever a given tissue occurs but more often only in parts of the body where some condition in the tissue in question is above (or below) a certain threshold. Localized direct effects may have secondary consequences that ramify from one tissue to others producing extensive changes throughout an organ or the whole body.

No distinction in kind can be made between the indirect effects of a single gene replacement and those of an altered environmental factor on the pattern of a developing embryo. The normal pattern as a whole, however, must be considered to be the consequence of differential action of the system of genes in a mass of cells in which there are from the beginning physiological gradients.

The action of any gene usually depends on the prior action of other genes, as illustrated in the simple case of the synthesis of arginine. The genes may accordingly be grouped into reaction systems that, however, may overlap. Experiments indicate that these systems are almost wholly independent of the spatial relations of the genes in the chromosomes. The calling into action of one whole system or of another may depend on which first step is invoked by local conditions. In cases of homoeosis in which a single gene may, as in the fruit fly, *Drosophila melanogaster*, replace antennae by legs, certain mouth parts by legs, balancers by wings, etc., the gene is to be looked upon not as a germinal representative of the whole complex structure but as a switch that alters conditions so as to set going a long-established reaction system in a strange location.

The term homology is used not only for corresponding structures in different species but for similar structures in different parts of the same organism. There are bilateral homologues such as the eyes; serial homologues such as the hands and feet; and other replicative homologues such as hairs, feathers, etc. The basis of such ontogenetic homology (on an individual organism) is clearly the calling into action of the same or similar reaction systems by conditions that recur in development. The interpretation to be put on phylogenetic homology (between members of different groups) is closely related. Both depend on the calling into play of more or less similar reaction systems by more or less similar local conditions. In ontogenetic homology the total array of genes

is identical, although the conditions may differ, while in phylogenetic homology every gene of the ancestor may be replaced by a somewhat different acting allele without altering the pattern of development enough to prevent closely similar conditions and responses from appearing in corresponding places. In both cases homology is properly a matter of degree rather than an absolute.

The action of the system of genes responsible for the general structure of bone, including the property of deposition of bone in adaptive relation to prevailing stresses, is doubtless common to all bones and thus gives a certain degree of homology to all. The action of the system of genes responsible for the form of a particular bone is more related to the local conditions. There may be almost perfect homology, as between the bones of right and left leg, or less perfect homology, as between bones of arm and leg, or merely the low degree of homology of all the bones referred to above.

Turning to phylogeny, the frequent transformation of an indefinite number of undifferentiated replicated homologues into a definite number of differentiated partial homologues, or the reverse, to which earlier reference has been made, is readily understandable. Elements that develop under virtually identical reaction systems in the ancestor may hardly be homologues at all in the descendant from the standpoint of common gene action. Conversely, structures such as chondrocranium and dermal scutes, which depend initially on almost wholly different reaction systems, may, in this sense, acquire a degree of ontogenetic homology (cartilage and dermal bones). A reaction system that is called into play only in certain anterior somites in a primitive vertebrate, to produce pronephric tubules, may be called into play in modified form in posterior somites of descendants, to produce partially homologous mesonephric or metanephric tubules. *See* also ANIMALS, PHYLOGENY OF.

IV. THEORIES OF EVOLUTION

Theories of evolution may be classified according to the roles assigned to chance variation and to influence of the environment. Both are denied in the Aristotelian conception of evolution as analogous to individual development, the realization of an innate potentiality irrespective of environment (except for provision of the necessary conditions for life). In modern times, this idea appears in the orthogenesis of Nägeli and as the core of H. F. Osborn's aristogenesis. According to another group of theories in which chance plays no role, evolution is a continuation of individual physiology through the inheritance of characters acquired either by a molding influence of the environment (Buffon, Eimer) or by active adaptation to the environment (Lamarck). Under the mutation theory, on the contrary, a single chance mutation is the crucial event in the origin of species. In the more extreme forms of this theory, it is held that the higher categories even up to the phyla arise only from mutations of appropriate magnitude, almost a denial of evolution as a cumulative process (Saint-Hilaire, J. C. Willis, Richard Goldschmidt). In the mutation theory of De Vries, species arise by single mutations, but the higher categories are cumulative products. Finally come those theories under which evolution is a statistical process of transformation of populations. Random variability furnishes the raw material, but the course, at least in the long run, is indirectly determined by the relation to the environment through natural selection (Darwin, J. B. S. Haldane, R. A. Fisher, S. Wright).

1. Orthogenesis.—Orthogenesis leaves unexplained the branching of the tree of life into millions of species, each marvelously adapted to some niche in nature. In modern times, indeed, it has usually been applied only to those cases in which there has seemed to be a uniform linear trend for a considerable period in a direction difficult to account for by adaptation (*e.g.*, elaboration of the sutures of ammonites, the increase in size of horns of titanotheres and of the Irish elk to unwieldy proportions, etc.) Most authors, however, believe that these cases are open to alternative interpretations in terms more in harmony with our present knowledge of heredity. A certain element of truth in this principle would be, of course, generally recognized in the principle that possible directions of "random" genetic change are restricted, sometimes severely, by the organization already attained. (*See* ORTHOGENESIS.)

2. Inheritance of Acquired Characters.—The theory of the inheritance of acquired characters, whose foremost proponent was Lamarck, seems at first sight to be the simplest interpretation of certain kinds of adaptations. Thus the evolution of complicated patterns of instinctive behaviour suggests heredity to be a sort of racial memory that is continually being modified by individual experience. Unfortunately for this view the most complicated instincts are probably those exhibited by the sterile workers, but not by the reproductive castes, among the social insects. The universality of such a caste system and also of nest building among the termites indicates that both of these characteristics trace back to the origin of the order. The differences in the nests even among species of the same genus, as noted earlier in the cases of *Amitermes excellens* and *A. meridionalis*, indicate a continuing evolution in the behaviour of the workers which cannot be due to direct transmission of changes in heredity based on their individual experiences.

The whole theory of the transmission of changes in anatomical characters or in behaviour, acquired by parents under special conditions, to descendants that have not been subject to these conditions runs counter to the principles of heredity and development established by experiment. The basic property of genes is the persistence of their properties, irrespective of the characters which develop from interactions with other genes or environment. The occasional mutational changes which occur are due to direct modifications of an individual gene or chromosome on absorption of thermal, chemical or radiant energy and are unrelated to the characters of the organism as a whole. The theory, moreover, implies a one-to-one relation between such characters and elements of heredity. Such a relation indeed seems to hold for protein specificity, but for anatomical characters as well as for behaviour any such preformation in the germ cells is quite out of harmony with the observed epigenetic course of development.

It may be added that there have been no repeatable experiments indicating that such a character acquired by the parent under special conditions tends to develop in the offspring in the absence of these conditions. This class of theories therefore is unacceptable. (*See* LAMARCKISM.)

3. Mutation Theory.—The mutation theory has the merit of being supported by experiments in which the origins of recognized species have been duplicated in the laboratory by inducing mutations. A mutant resulting from doubling of the entire set of chromosomes (tetraploidy) may be fully fertile by itself while producing a sterile triploid hybrid on crossing back to either parent species. In the case of autotetraploids (arising within a species) the character change is so slight that a specific difference is likely not to be recognized unless by chromosome count. If of sufficiently ancient origin, however, the reproductive isolation from the parent species will probably have led to a drifting apart from the latter by processes discussed later. An amphidiploid hybrid (arising from the union of two different species) on the other hand has all of the essential properties of a new species from the first. There is no doubt that many species of flowering plants have arisen in this way. Thus plants which are virtually indistinguishable from a natural species, *Galeopsis tetrahit* with 32 chromosomes, have been produced from hybrids of *G. pubescens* crossed with *G. speciosa*, each with 16 chromosomes. A considerable number of other cases are indicated.

The origin of a new species in nature has been observed in the case of *Spartina townsendii*, a plant which first was found in 1870 in largely unoccupied mud flats in southern England. After 1900 it multiplied enormously. Its characters and its chromosome count (126) indicate that it is an amphidiploid hybrid between European *S. stricta* with 56 chromosomes and *S. alterniflora*, an American species with 70 chromosomes, the seed of which probably was carried to Europe accidentally. Its preadaptation to an almost empty niche in nature is noteworthy.

There are many cases in which forms that have all the characteristics of new species have been produced in the laboratory, among which one of the most remarkable is the generic amphidiploid *Raphanobrassica* (36 chromosomes) produced by G. D. Kar-

petchenko from the radish *Raphanus sativus* (18 chromosomes) and the cabbage *Brassica oleracea* (18 chromosomes). *Raphanobrassica* is a luxuriant plant which is very different from either parent.

There are probably cases of autotetraploid and amphidiploid species among hermaphroditic and parthenogenetic animals, judging from chromosome numbers. The relative prevalence of exclusive biparental reproduction in higher animals is probably the reason for the much lower frequency of all forms of polyploidy among them than among the higher plants as a means of multiplication of species. There is little chance of establishment in nature unless there is the possibility of extensive multiplication by uniparental reproduction.

Types of chromosome change (other than the multiplication of entire sets of chromosomes, polyploidy) and gene mutations, whatever the magnitude of their effects on the organism, merely produce segregating variations within the populations in which they appear. This was the case, for example, with a number of forms of evening primrose (*Oenothera*) which De Vries found growing wild in Holland. These differed from the predominant form, *Oenothera lamarckiana*, in every respect, suggesting that they were different species. Since *Oenothera lamarckiana* was an American form, escaped from cultivation, it seemed to De Vries that he was observing here the abrupt origin of new species. His mutation theory had its origin primarily in these forms. It has turned out that most of these have the chromosomal complement of *Oenothera lamarckiana* (14 chromosomes) plus an extra chromosome. The latter is transmitted only in ovules, since pollen containing it is abortive. These types thus can exist only as segregants of a *lamarckiana* population and cannot be considered as new species. In other cases, especially including inversions, translocations and small duplications, a new chromosomal type can be produced in true breeding form but only by statistical processes within the population or an isolated portion of it.

Comparisons of the chromosomes of closely related species often show that an accumulation of many changes must have occurred in one or both after their separation. Crosses between species, however, also indicate multiple gene differences. Occasionally it is possible to demonstrate the segregation of alleles (alternative conditions of a gene on any particular point, or locus, of a chromosome) with major effects on one or more characters. Most of the differences between characters, however, show the type of segregation that indicates differences in a large number of genes with individually slight effects.

Genetic analysis of the variability actually found within local populations reveals differences of the same nature as those found to distinguish species. Differences in chromosome pattern are found occasionally. Especially common are differences in order within chromosomes. Thus in *Drosophila pseudoobscura* 15 different orders are found within the third chromosome alone. These are distributed so that three or four are usually segregating in each locality. Extensive breeding tests of wild populations always reveal the presence in low frequencies of alleles of many genes. In man and in many species of other mammals and in birds, fishes, grouse, locusts, snails, butterflies, etc., there is conspicuous polymorphism (variation in form) because of the presence of two or more alleles at high frequencies. In addition to alleles with such conspicuous effects that genetic analysis can readily be made, there is always considerable genetic variability of all measurable characters—including general size, relative proportions of all parts, intensity of colour, etc.

It is clear that the genetic differences between closely related species are not such as could have arisen at a single step and that the raw material for the accumulation of the differences is present in the genetic variability existent in all species. It is necessary, therefore, to consider the processes that may bring about a gradual transformation in the genetic composition of a population.

4. Statistical Transformation of Populations.—Darwin viewed natural selection of better fitted variants as the chief molding force of evolution, but of course he had no conception of the complex genetic mechanisms involved in the production of such variants (*see* SELECTION; VARIATION). It remained for the neo-Darwinian school, having benefit of the knowledge that proceeded from the rediscovery of Mendel's work, to systematize the genetic basis of evolution through statistical analysis of gene pools in populations. (*See* DARWINISM.)

Under the mutation theory of evolution, a species is treated as if characterized by a single typical array of genes, organized in a single particular pattern in the chromosomes. Under the statistical theory, the genetic constitution of a species can be adequately represented only in terms of sets of gene and chromosome frequencies. It is recognized that observed mutation rates are such that any reasonably large population may be expected to carry more than one allele, usually many, at many if not all loci. Because of the symmetry of the Mendelian mechanism, the frequency of each allele in the population tends to remain constant in the absence of disturbing factors. The proper representation of the genetic constitution of a random breeding population of diploids is of the type $(p_1A_1 + p_2A_2 \ldots + p_mA_m)^2(q_1B_1 + q_2B_2 \ldots + q_nB_n)^2 \ldots$. Here the A's represent one set of alleles and the p's their relative frequencies; the B's represent another set, the q's their frequencies and so on. The arrays of each set are squared to take account of the postulated random unions of gametes. The arrays for the different sets are represented as multiplied in accordance with the fact that the combinations within each set may be expected to be combined at random with those in other sets, in the long run, including even those in the same chromosome.

This mode or representation may appear rather complicated. Really, however, there is an enormous simplification in attending to the stock of genes carried by the population instead of to the array of genotypes of the actual individuals. Assume 4 alleles (10 combinations) in each of 1,000 sets. The formula for the species would require symbols for 4,000 genes, each with a symbol for its frequency. This is a rather large number but it is incomparably less than the potential number of genetically different kinds of individuals, 10^{1000}.

The system of gene frequencies may be expected to remain substantially constant for many generations, but the array of actual individual genotypes (genetic constitutions) cannot be expected to have a single element in common in any two generations. Consideration of the gene frequencies shows how certain statistical properties of the population such as the average, and the amount of variability in each respect, may be expected to persist. It is to be noted that variability itself is a property of major importance in adapting a population to its environmental vicissitudes.

The elementary evolutionary process from this viewpoint is simply change of frequency of a gene or of a chromosome pattern. The immediate processes may be grouped according to the degree of determinacy. First are three types of systematic change: (a) mutation pressure, due to recurrent change of a given sort in the hereditary material; (b) immigration pressure, due to introduction of different heredity from without; (c) selection pressure, due to any systematic cause by which the gene or chromosomal arrangement in question tends to increase or decrease in frequency without either mutation or immigration. Differential mortality, differential rate of attainment of maturity, differential mating, differential fecundity and differential emigration are such causes. Next are random changes in gene frequency due to fluctuations in the systematic processes and to accidents of sampling, determinate only in variance. In the final group are indeterminate unique events (*see* GENETICS OF POPULATIONS).

Recurrent mutation, immigration and selection all tend to have systematic effects. The alleles that are present in a population are, in general, ones for which these systematic pressures tend to oppose each other in one way or another. Thus the tendency to elimination of a given gene by selection may be balanced at a certain frequency by its continued reintroduction by mutation or immigration. Selection favouring a heterozygote (A_1A_2) over both of the corresponding homozygotes ($A_1A_1A_2A_2$) may maintain both alleles at certain frequencies. Presumably all mutations that are likely to arise at one or two steps from the more abundant genes present in the population have been tried out by natural selection and found wanting, and thus are found at negligibly low frequencies, if at all. There may be very valuable mutations that

could only arise through a succession of unfavourable ones, but these will have very little chance of occurring. The result is a system of equilibrium frequencies, any departure from which tends to be followed by restoration. Thus there is a tendency toward stability of subspecies and of species in spite of continuing mutation, continual interbreeding at a low rate with neighbouring populations and continuing selection.

Selection relates to the organism as a whole or to the social group, not to single genes, except as a net resultant. Genes that are favourable in one combination may be unfavourable in another. This indeed should be the rule for the class of genes carried at intermediate frequencies in the population. Suppose, for example, that there are many sets of alleles that affect the relative lengths of legs and body in an animal that depends on speed. Maximum speed requires a certain proportionality. Genes that increase the relative length of the legs would be favoured or opposed by selection depending on whether they are associated with an array of other genes that tend to make the legs too short or too long. If merely the combination of two pairs of alleles are considered, there may be two different "peak" combinations with respect to adaptive value, *e.g.*, *AAbb* and *aaBB* in contrast with both *aabb* and *AABB* A species may be held to approximate fixation of one of these even though the other is somewhat more favourable because the passage through populations of lower average adaptive value, *e.g.* $[(\frac{1}{2})A + (\frac{1}{2})a]^2[(\frac{1}{2})B + (\frac{1}{2})b]^2$, may be too improbable. In the system of adaptive values of the combinations of all sets of alleles, thousands in number, there may be a virtually infinite number of distinct peaks, each corresponding to a different harmonious combination of traits. The direct pressure of selection tends to carry the system of gene frequencies to the neighbourhood of the peak on the slope of which it already lies, but this is not likely to be the highest peak. The genotypes present in the population when equilibrium is reached about this peak will all be at relatively few removes from the peak genotype, although no two of them may be expected to be exactly the same.

Increase in the severity of selection will bring about a higher level of adaptation, but only by forcing a denser clustering of genotypes about the peak type; thus variability tends to be reduced. Increase in the general rate of mutation has the opposite effect, a wider dispersal of genotypes from the peak type, giving a lower average of adaptation but more variability. There is an increased chance of production of individuals highly adapted in a different way, *i.e.*, with a genotype in the neighbourhood of another and perhaps higher peak combination; but, as the genotype is broken up in the formation of germ cells, this does not create an evolutionary trend in the new direction. The problem of evolution from this viewpoint is to find a process by which a species may pass from control of one peak system to control by a higher one. The most obvious possibility here is the occurrence of a novel mutation. There are, however, other processes that are probably more frequent. So far it has been assumed that there exists a constancy of the conditions of selection. But conditions are constantly changing as a result of changes in the abundance of other species, the evolution of these and of secular changes in the nonliving environment. The adaptive values of peak genotypes may be depressed while those of other genotypes may be elevated. Selection under these conditions keeps the species continually on the move in the direction, whatever it may be, in which the pressure is momentarily greatest. In the long run the species tends to move into the predominantly higher portions of its general field of potentialities, an important evolutionary process.

It has been assumed that there is random interbreeding within the range of the species. If the species occupies a large range in which the conditions of selection are not the same in all regions, the local selection pressures may prevail over the tendency toward homogeneity due to migration, provided that dispersal is sufficiently restricted by geographic or ecologic barriers or if dispersal is slight in comparison with the range of the species. The result is the splitting up of the species into subspecies characterized by adaptive differences. The species as a whole has a more extensive field of genetic potentialities than under random mating through increase in the number of alleles carried at each locus. This also is true of the separate subspecies because of interbreeding with each other at the boundaries of their ranges. Where ranges overlap there is a greater chance that one of the subspecies, in increasing its adaptation to the local conditions, may hit upon an adaptation of general rather than merely local significance, which thereupon may spread through the whole species.

Under uniparental reproduction each mutation gives rise to a group, in this case a clone, of individuals that are genetically identical until split into diverse types by a new mutation. Because of this, selection would be enormously more effective than under biparental reproduction, if the number of genotypes on which it could operate were comparable. As noted earlier, however, the field of potential genotypes is actually less than under biparental reproduction. The advantages of the two systems are combined if there is predominant uniparental reproduction but sufficient crossing to provide an unlimited variety of potential kinds of clones. Such a balance between the methods of reproduction provides a trial-and-error method by which the species may pass from control by one peak combination to another. It is especially effective in organisms whose numbers increase enormously under favourable conditions but are forced to contract on return of unfavourable conditions. The maintenance of a relatively stable population requires more labile local adaptations than provided by this system.

No consideration yet has been given to the random processes as a cause of change of gene frequency. The resultant of the continual tendency of gene frequencies to move toward certain equilibrium values under the systematic pressures that have been discussed and the tendency to drift at random due to accidents of sampling is a certain distribution of probabilities that should be exemplified in the course of time. With given systematic pressures, the mean deviation from equilibrium due to this cause is slight in large populations, but large in small ones. In very small populations, in fact, the cumulative effect of accidents of sampling dominates the situation so completely that one allele or another tends to drift into fixation at each locus. Being random, the resulting genotype virtually is certain to be poorly adaptive. The population suffers the well-known degenerative effects of too close inbreeding, with extinction as the probable result.

With a certain balance between the effects of the random processes and the pressures of mutation and selection, loss of variability and average lowering of fitness may be slight, even though all gene frequencies are drifting rather widely about their equilibrium values. The constitution of the population keeps moving in a largely random fashion in the field of possible gene frequencies but under loose control of an adaptive peak. It may, however, wander across a "valley" to come under control of another peak, corresponding to the emergence of a new adaptive type. There is a trial-and-error process here, but one that is too slow to be of much evolutionary significance.

A similar process may, however, be of great evolutionary significance when occurring in each of a large number of small, largely but not completely, isolated local races. In this case, the constitution of each race drifts separately under loose control of the peak type. The rate, based here on the balance between the effects of the random processes and immigration pressure, may be enormously more rapid than in the preceding case (in which mutation rate is a limiting factor).

With continually shifting, only partially adaptive, differentiation among numerous local populations, there is an excellent chance that the constitution of one will cross a valley to come under control of a higher peak and become the principal source of immigrants to other populations, with the consequence that these will be drawn toward it in composition. In the end the whole species may come under the control of this higher peak.

This trial-and-error process, which may continue indefinitely, may be described as one of intergroup selection. The selection depends on the genetic system of each local race as a whole, instead of merely on the net effects of single genes (all that is possible in a homogeneous population). Characters may be fixed that are favourable to the group even though disadvantageous in individual competition. The consequences of this system are ob-

viously somewhat similar to those of the balance between uniparental and biparental reproduction discussed earlier. The properties of the local race are, however, more favourable to the maintenance of reasonably stable population numbers than are those of the clone.

It is to be noted that the stock of potential but largely unexpressed variability carried in the species as a whole is so great, if there is extensive subdivision into partially isolated local races, that evolution is not likely to be limited by the rate of occurrence of mutations. The limitation is in the difficulty of passage from domination by one adaptive type to domination by another, against the conservative pressure of selection.

5. Splitting of Species.—As frequently noted, a mutation, however great its effect, cannot be considered as a new species if it persists only as a segregant in the population in which it appeared. The splitting of a species implies the formation of two populations that do not exchange genes. A daughter species may arise from individuals that by a rare accident reach a geographically isolated region. The mere isolation is, of course, not enough by itself. Few biologists would be willing to accept the conclusion that individuals born as members of one species become members of another merely by isolation, or that their undifferentiated descendants may be considered as of a different species. In the long run, however, the processes that lead to transformation of single species would usually insure that two isolated groups would become sufficiently different, even without any differences in the conditions of selection, to become recognizable species by the usual criteria.

They would not be separate species by the ideal definition, however, until the appearance of differences that would prevent interbreeding on contact. Among the conditions that can clinch more or less permanently a species difference are psychic or morphologic differences that interfere with cross mating; physiological inhibition of fertilization; abnormal development of hybrids; hybrid sterility, whether due to imperfect development of gonads or to failure of gametogenesis from accumulation of chromosomal differences; and if all else fails, prevailing abnormal development in F_2 (second filial generation) segregants.

For the most part, such differences can obviously come about only by the accumulation of steps each one of which has no appreciable isolating effect from that immediately preceding.

These three processes—primary isolation, differentiation in characters and the permanent clinching by genetic barriers to interbreeding—may occur in various relations to each other. The sequence—geographical isolation followed by a secondary drifting apart in morphology and in conditions that prevent possible fusion—has probably been of primary importance. On the other hand, two populations that retain an intergrading connection may become as different as typical species and then graduate from subspecific to specific status by the practical definition, without further change, merely by extinction of the connecting population. They even may be unable to interbreed on restoration of contact and thus attain the rank of distinct species by the ideal definition as shown by the species rings discussed earlier.

All aspects of speciation may proceed coincidentally in two adjacent subspecies that are becoming adapted to different ecological conditions. Here an intergrading population may be progressively reduced in number and ultimately eliminated by selective favouring of bars to the production of hybrids poorly adapted to both conditions.

The splitting of a species within a single territory is probably very unusual in biparentally reproducing forms, but geographical isolation need not be so great, as otherwise to permit speciation, if correlated with marked ecological differences. Moreover, in forms with very weak means of dispersal, species may arise in relatively small, adjacent regions, especially if these are ecologically diverse. In special cases a single mutation may permit speciation within the territory of the parent form by its isolating action. The cases of autopolyploidy and amphidiploidy have already been discussed. A mutation may bring about a change in the time of flowering in a plant so great that no interbreeding with the parent form is possible.

Directly isolating mutations, however, are unlikely to become established within a population that reproduces wholly by means of separate sexes. Prevailing uniparental reproduction enormously increases the chance of success of such mutations but also, as noted, reduces the definiteness of the species concept itself.

V. GENETIC INTERPRETATION OF MAJOR EVOLUTIONARY PROCESSES

The prevalence of apparently directed evolutionary trends has given paleontologists a strong bias toward the conception of an innate perfecting principle or of an organic momentum. There seems no overwhelming difficulty, however, of interpretation by statistical genetics. A form that has been fitted for life in one ecological niche, in a region in which all niches are occupied, is likely to be in a better position to modify its organization toward more perfect adaptation to its niche, or toward minor changes that occur progressively in this niche, than any other species, but is correspondingly in a poorer position than others to make over its organization for adaptation to other niches. General control by natural selection seems an adequate explanation of such cases as the somewhat orthogenetic sequence of genera of the horses from the Eocene to the present. This is not incompatible with considerable divergence of species and subspecies of jointly adaptive and nonadaptive character. Trial-and-error processes at these levels indeed may furnish the basis for apparent orthogenetic advance at the higher level. This was essentially W. D. Mathew's interpretation of the paleontological record of the horses. In terms of the system of adaptive values of gene combination such forms may be said to be ones for which there is only a single available line of increasing peak values.

More difficulty perhaps has been felt in cases in which there is an apparent trend toward the development of an injurious character leading to ultimate extinction. A trend preceding extinction may mean merely that the form has become committed by its past history to a mode of adaptation that is inherently less effective than that of a new rival form, forcing it into a line of ever narrower specialization, which eventually becomes a blind alley. In other cases, as J. S. Huxley has emphasized, selection between individuals of the species in respects that contribute to success in mating may direct evolution along a course opposed to the welfare of the species. Horns, tusks and total size may become unwieldy through such a process. Again, as Huxley also has pointed out, strong selection in one respect, especially toward increased size, may alter the proportions of parts in a way that suggests organic momentum because of the correlated effects of a system of differential growth rates of the parts.

One sort of trend that requires special consideration is that toward degeneration of parts that have ceased to be useful. A certain amount of reduction of course may be accounted for by direct selection against useless encumbrances, but evolution often carries degeneration beyond the point at which this explanation seems plausible. Another possibility is direct mutation pressure. The most general characteristic of mutations, due to their accidental nature, is a degenerative effect, as is well-illustrated by a random array of *Drosophila* mutations in which defective eyes, wings, bristles and pigmentation are the rule. Mutation alone may be adequate especially in small isolated populations. There are, however, indirect effects of selection that may be more important. Genes may have effects on seemingly unrelated characters. In a homogeneous population, alleles are selected on the basis of their net effects. Certain genes, maintained at high frequencies because of roles in the development of an important organ, may be less favourable than certain of their alleles in other respects. If this organ ceases to be useful, selection in the other respects takes precedence, with consequent continued degeneration of the useless organ. On the other hand, a useless organ may be maintained in spite of a slight encumbering effect, if the genes involved in its development have effects that are tightly interwoven into the general developmental system. In general, these indirect effects of selection may be expected to speed up the elimination of useless organs in a species that is being rapidly made over, but are likely to bring about their retention as vestiges in a conservative form.

The long-continued cessation of evolution in some forms is another problem. The apparently obvious explanation that mutation rate is exceptionally low is probably not the primary one. It is more probable that such a form has reached the limit of adaptation along its line of specialization. In terms of the system of adaptive values of gene combinations, they are held to the neighbourhood of the only peak immediately available. However, as mutation rate itself is known to be subject to gene action, it is possible that mutation rate may be secondarily lowered by selection in such a case.

In the opposite direction is the problem of the apparently abrupt emergence of a new higher category and its rapid branching by adaptive radiation. Here the most obvious explanations are either exceptionally rapid mutation or the occurrence of individual mutations with exceptionally great morphological effects. But rapid mutation merely increases the number of freaks, and the greater the morphological effect the less the chance of a harmonious, adaptive result. The explanation is to be sought rather in the magnitude of an opportunity. The field of gene combination is so extensive that, even with complete cessation of mutation, explosively rapid evolution is possible for a considerable period under favourable conditions, particularly ones in which the conservative pressure of selection is relaxed.

Favourable conditions fall into two major categories. First is introduction into unoccupied territory. A large portion of the land birds of the Hawaiian Islands belong to one family, the Drepanididae, found nowhere else in the world, and consisting of several highly diversified genera and over 20 species. Presumably this family traces to the immigration of a few stray individuals of an American species. This species, whatever it was, had no such extraordinary outburst of evolutionary activity in its original home in which all ecological niches were occupied. The system of adaptive values of gene combinations in Hawaii (in contrast with that in America) doubtless included a very large number of peaks, easily available by trial-and-error processes. Species with different adaptations might become established on different islands, even though there were no important differences in conditions on these. Stray migrants between islands after speciation had been clinched therefore would find unoccupied ecologic niches until all niches were occupied on all islands.

The rapid adaptive radiation of the few primitive types of mammals that reached South America between the Paleocene and Pliocene, and presumably that of the marsupials in Australia, are similar examples on larger stages.

A comparable opportunity is presented to a species that attains by any means a general adaptation that opens up a new way of life or an advantage in many ways of life. There may be nothing especially unusual or rapid in the first steps before a critical threshold is passed in the attainment of the adaptation. Thus, whereas the amphibian leg and foot seem to have emerged very rapidly from the crossopterygian fin, the complex of adaptations that made the mammals a successful class required most of the Mesozoic for its development before the threshold for widespread adaptive radiation was reached at the beginning of the Cenozoic.

See also references under "Evolution, Organic" in the Index volume.

BIBLIOGRAPHY.—Charles Darwin, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (1859), *The Descent of Man* (1871) (several reprints of each published); E. O. Dodson, *A Textbook of Evolution* (1952), *Evolution: Process and Product* (1960); J. C. Greene, *Darwin and the Modern World View* (1961); H. H. Ross, *A Synthesis of Evolutionary Theory* (1962); Sol Tax (ed.), *Evolution After Darwin*, 3 vol. (1960); E. Mayr, *Systematics and the Origin of Species* (1942); J. S. Huxley, *Evolution, the Modern Synthesis* (1942); T. Dobzhansky, *Genetics and the Origin of Species*, 3rd ed. (1951), *Evolution, Genetics, and Man* (1955); G. G. Simpson, *The Major Features of Evolution* (1953).

(S. WT.)

EVONYMUS: *see* EUONYMUS.

ÉVORA, a surname of a number of members of Portuguese religious orders who were born in the city of Évora.

JOÃO DE ÉVORA, of the Order of the Trinity, studied philosophy and theology and was instrumental in building the first residence of that order on Mt. Sintra, where, by authority of John I, in 1400 a convent was constructed. He became confessor to John I and in 1414 was appointed bishop of Viseu. He died in Lisbon in 1426.

JOÃO DE ÉVORA, a 15th-century canon of the see of Évora, was director of the tribunal formed by Afonso V in 1461 to free prisoners of the infidels. Another man of the same name was a 16th-century reformer of the Order of St. Jerome who helped establish various dioceses.

EUSÉBIO DE ÉVORA, a 16th-century friar, was a teacher and supposed writer of the Order of St. Jerome. He was confessor to John III, king of Portugal, and handled all religious matters concerning the royal family. He was appointed governor of Algarve by John III and later became bishop of that city.

ÉVORA, the capital of an administrative district of the same name in Alentejo province, Portugal, lies 154 km. (96 mi.) E.S.E. of Lisbon by road. Pop. (1960) 51,968 (mun.). The town occupies a fertile valley enclosed by low hills. It is surrounded by ramparts flanked with towers, and has two forts. The cathedral, originally a Romanesque building erected 1186–1204, was restored in Gothic style about 1400; its richly decorated chancel was added in 1761. The church of São Francisco (1507–25) is a good example of the blended Moorish and Gothic architecture known as Manoellian. The regional museum, formerly the archbishop's palace, contains a Roman collection, and Portuguese and Flemish paintings. The Estaus palace has been almost demolished to make room for the market square, in which a large fair is held at midsummer. A small Roman temple is usually known as the Temple of Diana, a name for which no valid authority exists. The former university, founded in 1559 to succeed the College of the Holy Spirit (or of the Jesuits) founded in 1551, was suppressed in 1759. A state secondary school, a commercial school and an almshouse now occupy the buildings. Évora is of little commercial importance, except as an agricultural centre, but Évora district (area 2,855 sq.mi.) is famous for its mules and abounds in corkwoods; there are also mines of iron, copper and asbestos, and marble quarries. Under its original name of Ebora, the city was from 80 to 72 B.C. the headquarters of Quintus Sertorius, and it long remained an important Roman military station. It was called Liberalitas Julia because of certain municipal privileges bestowed on it by Julius Caesar. Its bishopric, founded in the 5th century, was raised to an archbishopric in the 16th. In *c.* 712 Évora was conquered by the Moors, who named it Jabura; it was retaken in 1166. From 1663 to 1665 it was held by the Spaniards. In 1832 Dom Miguel, retreating before Dom Pedro, took refuge in Évora; and there was signed the convention of Évora, by which he was banished. *See* PORTUGAL.

EVPATORIA: *see* YEVPATORIYA.

ÉVREUX, a city of northwestern France, capital of the *département* of Eure, lies on the slopes of the Iton valley, 100 km. (62 mi.) W.N.W. of Paris by road. Pop. (1954) 20,441. Although heavily damaged during World War II, the town has been largely rebuilt and retains its former character. Besides its public park and boulevards on the site of the former ramparts, Évreux possesses many ancient monuments. The cathedral of Notre Dame (12th, 13th and 16th centuries) has fine windows and woodcarvings. Other features of interest are the 15th-century belfry; the church of St. Taurin (11th and 15th centuries), a former Benedictine abbey in which is preserved the shrine of St. Taurin, a magnificent 13th-century gold carving; the museum, rich in Gallo-Roman exhibits, which is housed in the former bishops' palace (15th century); and also several educational establishments. Halfway between Paris and the coast, Évreux is on the main railway from Paris to Cherbourg. It is the centre of a rich agricultural district, but is also of industrial importance, with foundries and manufactures of electronic equipment, medicines, metal goods and chemicals.

Known to the Romans as Mediolanum Aulercorum (site of Le Vieil-Évreux), it was the capital of the Gallic tribe of the Aulerci Eburovices and a flourishing city during the Gallo-Roman period. Its bishopric dates from the 4th century. The first known family of the counts of Évreux was descended from an illegitimate son of Richard I, duke of Normandy, and became extinct in the male line with the death of Count William in 1118. The countship passed

successively to the lords of Montfort-Lamaury, to King Philip II Augustus in 1200, to the brother of Philip the Fair in 1307, to Charles the Bad, then to Charles the Noble who ceded it to King Charles VI in 1404. In 1427 Charles VII gave it to Sir John Stuart of Darnley. At his death the countship reverted to the crown. In 1651 it was finally made over to Frédéric Maurice de la Tour d'Auvergne, duc de Bouillon. In German hands for a short time in 1870–71 during the Franco-German war, in World War II Évreux was held by them from June 1940 to Aug. 1944.

(L. He.)

EVZONES, members of elite mountain infantry units in the Greek army, analogous to Scottish Highlanders. They are distinguished by their picturesque white jackets, wide skirts and Albanian-type slippers with turned-up tufted toes. Organized units of evzones, who originated in Epirus, have existed since the Greek War of Independence. They won fame for their bravery and tenacity in the Balkan Wars and in World War II. In Nov. 1940 an evzone regiment turned the tide of the Italian invasion by trapping and destroying a crack *alpini* division at Metsovo pass, using their traditional tactic of sweeping down from the ridges screaming their battle cry "*Aera! Aera!*"

After World War II the regiments were disbanded and only a royal guard was retained at Athens, but evzones remained the idols of Greek youth. (P. N. T.)

EWALD, HEINRICH GEORG AUGUST VON (1803–1875), one of the foremost German orientalists and theologians of the 19th century, was born on Nov. 16, 1803, at Göttingen. Educated at the university of his native town, he was made professor there in 1827. Having in 1837, along with six of his colleagues, signed a formal protest against the abolition of the liberal Hanoverian constitution of 1833, he was expelled from the university. In 1838 he received a call to Tübingen, where he taught for ten years. To this period belong some of his most important works, and also the commencement of his controversy with F. C. Baur (*q.v.*) and the Tübingen school. In 1847 he returned to Göttingen, the liberal constitution having been restored.

The chief crisis in his life arose out of the political events of 1866. His loyalty to the king of Hanover would not permit him to take the oath of allegiance to the king of Prussia, who had deposed him and annexed the little kingdom, and he was therefore retired from the faculty in 1867. The violent tone of some of his printed manifestoes led to his being deprived (1868) of the right to lecture, ordinarily held by retired professors, and to an indictment, which, however, resulted in his acquittal (1869). Then, and on two subsequent occasions, he was returned by the city of Hanover as a member of the North German and German parliaments. In June 1874 he was found guilty of a libel on Prince Bismarck. He died on May 4, 1875.

As a teacher Ewald had a great power of kindling enthusiasm. His works on the poetic and prophetic books of the Old Testament *Die poetischen Bücher des alten Bundes*, 1835–37, and *Die Propheten des alten Bundes* (1840–41, Eng. trans. 1875–81) reveal this. His *Hebrew Grammar* (1827; Eng. trans. 1836) marked new progress in biblical philology; and as an exegete and biblical critic no less than as a grammarian he left an abiding mark. His *Geschichte des Volkes Israel* (*History of Israel*, five volumes, 1843–55, two further volumes added to 2nd ed., 1858–59; Eng. trans. 1865–86), the result of 30 years' labour, was epoch-making in that branch of research. His other chief works are *Alterthümer Israels* (1848; Eng. trans. 1876), and *Die Lehre der Bibel von Gott oder Theologie des alten und neuen Bundes* (four volumes, 1873–76).

See T. Witton Davies, *H. Ewald* (1903); T. K. Cheyne, *Founders of Old Testament Criticism* (1893). (E. M. Kr.)

EWALD, JOHANNES (1743–1781), who inaugurated a new period of Danish literature, was born in Copenhagen, Nov. 18, 1743. On the death of his father, a poorhouse chaplain, he was sent to school at Slesvig (Schleswig) where the reading of *Tom Jones* and *Robinson Crusoe* aroused his spirit of adventure. In 1758 he went to Copenhagen to study theology, but fell in love and, in search of quickly gained glory, ran away to the Seven Years' War. He returned with neither glory nor money to find that the beloved Arendse, whom he immortalized as his muse, had married another. He passed his final examination when he was 19 and was already becoming known as a writer of prose and occasional poetry. When finishing *Adam og Eva* (1769), a dramatic poem in the style of French tragedy, he met Klopstock, and at about the same time he read Shakespeare's plays and James Macpherson's *Ossian*. Their influence resulted in the historical drama *Rolf Krage* (1770), taken from an old Danish legend of Saxo Grammaticus.

Ewald's life now began to show signs of serious disorder, especially an addiction to alcohol. In the spring of 1773 his mother and the pietistic pastor J. C. Schönheyder secured his removal from Copenhagen to Rungsted. There he produced his first mature works: *Rungsteds Lyksaligheder* ("The Joys of Rungsted"), a lyric in the elevated new style of the ode; *Balders Død* (Eng. trans. by G. Borrow, 1899), a lyric drama on a subject from Saxo and old Norse mythology; and the first chapters of his memoirs, *Levnet og Meninger* ("Life and Opinions"), explaining his enthusiasm for the adventurous and fantastic. In 1775 he was transferred to a still more solitary place near Elsinore, where he went through a religious crisis—a struggle between the pietistic idea of self-denial and his own proud independence (*cf.* his "Ode to the Soul"). In 1777 he was allowed to return to Copenhagen. His poetic genius was recognized and his life became calmer despite increasingly severe illness. On his deathbed he wrote the heroic-pietist hymn, "Udrust dig, helt fra Golgatha" ("Gird thyself, Hero of Golgatha"). He died in Copenhagen, March 17, 1781.

Ewald renewed Danish poetry in all its genres. Of his dramatic works only *Fiskerne* ("The Fishermen"; 1779), an operetta, is performed today. His greatest work in prose is his posthumously published memoirs, in which lyrically pathetic chapters about Arendse intermingle with humorous passages. He is known best as a lyric poet, especially for his great personal odes, and for songs, such as "Kong Kristian stod ved højen Mast" (trans. by Longfellow as "King Christian Stood by the Lofty Mast"), which is used as a national anthem, and "Lille Gunver," the first Danish "romance." His poetry was influenced by all the tendencies of his time. In form rooted in the classic tradition, it heralds Adam Öhlenschläger and the romantic revival by its enthusiasm and use of themes drawn from old Norse literature.

Bibliography.—Ewald's *Samlede Skrifter* appeared in 6 vol. with a bibliography in vol. 6 (1914–24). *See* also M. J. Hammerich, *Ewalds Levnet* (1860); A. D. Jørgensen, *Johannes Evald* (1880); and monographs by H. Brix (1913) and E. Frandsen (1939). (O. A. F.)

EWART, WILLIAM (1798–1869), English politician and opponent of capital punishment, was born in Liverpool on May 1, 1798, and died at Broadleas, near Devizes, on Jan. 23, 1869. Ewart carried through parliament in 1834 a bill for the abolition of hanging in chains, and in 1837 he secured the passage of an act abolishing capital punishment for cattle stealing and other offenses. In 1850 he carried a bill for establishing free libraries supported out of the rates, and in 1864 for legalizing the use of the metric system of weights and measures. He advocated the abolition of capital punishment and on his motion in 1864 a select committee was appointed to consider the subject.

EWE, a Negro people of the Guinea coast who inhabit the southeast of Ghana and the south of Togo. They fall into three main groups: the Anglo, on what was formerly the Gold Coast to the west of the lower Volta, 300,000; the Ewe proper, in the south of the former British Togoland (integrated with Ghana in 1957) and southwest Togo, 520,000; and the Watyi, in southeast Togo, 150,000; altogether they total 970,000 (1960 estimate). To these may be added the Mina, who live on the southeast coast of Togo (Anecho) and number 75,000.

According to tradition, the starting point of the Ewe tribes before they were dispersed was Nuatja, in southern Togo. The Ewe were divided in 1899 between the British Gold Coast and German Togoland, and in 1919 the latter was split into a British and a French mandate; but they have always laid claim to unity, in particular since 1945. They comprise the largest political group in the government of Togo.

The Ewe language belongs to the Kwa group, and there are

several dialects. The people's staple diet is composed of yams, maize (corn) and fish; crafts (weaving, pottery, ironwork) and trade are highly developed.

The family forms the basic social group; the Anglo have clans, more loose in structure and which restrict marriages to the clan. Formerly, tribes (more than 100) were the political unit; their chief was elected from the royal family and was assisted by a council.

Ewe religion is organized round a creator god (*Mawu*) and a large number of secondary gods (*trowo*)—sky, earth, mountains, etc.—which have priests and special forms of worship. The spirits of ancestors are connected with sacred stools, and divination is very widely practised. Many Ewe have become Christians.

See M. Manoukian, "The Ewe-speaking People of Togoland and the Gold Coast," *Ethnographic Survey of Africa: Western Africa,* pt. vi (1953). (Hu. De.)

EWING, SIR (JAMES) ALFRED (1855–1935), Scottish physicist and specialist in magnetism, was born at Dundee on March 27, 1855. He studied at Edinburgh and held successively the following appointments: professor of mechanical engineering, University of Tokyo (1878–83); professor of engineering, University college, Dundee (1883–90); professor of mechanism and applied mechanics, King's college, Cambridge (1890–1903); director of naval education to the British admiralty (1903–16); and principal and vice-chancellor of the University of Edinburgh from 1916. Ewing's work on the magnetic properties of iron, steel and other metals is noteworthy. He modified Weber's theory of induced magnetism, and constructed a magnetic model which behaved in accordance with his theory, he also observed the phenomenon which he named hysteresis (*q.v.; see* also Magnetism). The Royal society awarded him the royal medal for this work in 1895.

Ewing was the author of a number of papers on thermoelectric properties of metals, on the effects of stress and magnetization on the thermoelectric properties of iron, on the crystalline structure of metals (in conjunction with Rosenhain) and on seismology.

He invented an extensometer, a hysteresis tester, a permeability bridge and other apparatus for magnetic testing.

Ewing was a member of the explosives committee and of the Ordnance Research board; from 1914 to 1917 he was in charge of the department of the admiralty dealing with cipher. He was a member of a number of learned societies and was awarded medals and honorary degrees. In 1911 he was knighted.

Ewing wrote *A Treatise on Earthquake Measurements* (1883), *Magnetic Induction in Iron and Other Metals* (1891, etc.), *The Steam Engine and Other Heat Engines* (1894; 4th ed., 1926), *The Strength of Materials* (1899, etc.), *The Mechanical Production of Cold* (1908), *Thermodynamics for Engineers* (1920), *An Engineer's Outlook* (1933).

EWING, JULIANA HORATIA (1841–1885), English writer of books for children, a daughter of Margaret Gatty (*q.v.*), was born, Aug. 3, 1841, at Ecclesfield, Yorkshire, where her father was vicar. She was the original "Aunt Judy" of her mother's book, with a gift for entertaining six younger brothers and sisters, and she was later a frequent contributor of short stories and serials to *Aunt Judy's Magazine.* She married in 1867 Alexander Ewing, whose work as a soldier took them to New Brunswick, and then (1869–77) to Aldershot, Eng. Ill-health troubled Mrs. Ewing all her life and she could not again go abroad with her husband; toward the end of her life they were separated for four years. In 1883 he was stationed at Taunton, where her last months were happily spent in their own home. She died at Bath, May 13, 1885.

Her best stories are drawn from her own experience, in particular her childhood: children, "characters," dogs, plants and country scenes were her favourite subjects. Randolph Caldecott, who illustrated some of her work, said its "force and finish" frightened him: these are the qualities of her many successful short stories (Ruskin liked *Our Field,* 1877), where she sketches her picture with great economy, compelling belief and engaging sympathy. In her earlier long stories, she fell into the temptation of imparting the whole wealth of imagined detail, but *Jackanapes* (1883), *Daddy Darwin's Dovecot* (1884), *Laetus Sorte Mea* (1885) and the last, sunlit *Mary's Meadow* (1886), were almost as severely pruned as her best short tales; the rather slight plots are managed truthfully and with humour, though death comes in too often.

Mrs. Ewing rarely wrote without didactic purpose; her many themes, treated with nursery morality, include sketching, sanitation and girls' education. Her perception is fine and her touch light: but her best gift is the happy reflection in all her writing of an unusual delight in "everything to be seen and heard."

(M. C. Cr.)

EWING, THOMAS (1789–1871), U.S. lawyer and statesman who as first secretary of the interior organized that governmental department, was born near West Liberty, Va. (now West Virginia), Dec. 28, 1789. He was graduated from Ohio university at Athens in 1815 and was admitted to the bar at Lancaster, O., in 1816. He early took an interest in politics, supporting the Whig party. He was United States senator from Ohio, 1831–37; secretary of the treasury from March 1841 to September of that year, resigning with the other members of the cabinet in the crisis following the death of Pres. William Henry Harrison; and secretary of the newly established department of the interior from March 1849 to July 1850, when he was again deprived of office by the death of Pres. Zachary Taylor. His final public office was another brief term in the United States senate (1850–51), completing the unexpired term of Thomas Corwin. Throughout the Civil War Ewing was a loyal supporter of the Union cause and two of his sons, Hugh Boyle Ewing (1826–1905) and Thomas Ewing (1829–96), rose to the rank of brigadier general, with the brevet of major general, in the Union army. Gen. William T. Sherman was both his adopted son and his son-in-law. Ewing died at Lancaster on Oct. 26, 1871.

(B. Dy.)

EXAMINATIONS. This article deals with the nature, purpose and development of different types of educational examination used in various parts of the world. Special attention is paid to examination practices in the common school or single-track system of the United States, the specialized or multitrack systems of England and of Europe as exemplified by France, and the Communist party systems of the U.S.S.R. and Communist China. For information on particular certifying or qualifying examinations *see* Civil Service; Legal Education, etc. For additional information on various types of tests, their uses, construction and administration, *see* Psychological Tests and Measurements. *See* also Teaching, Methods of.

The main divisions of this article are as follows:

I. General Historical Background
II. Uses of Examinations
  1. Evaluation of Progress
  2. Evaluation of Effectiveness of Instruction
  3. Guidance
  4. Selection
  5. Certification
III. Types of Examinations
  1. Written, Oral and Practical
  2. Essay-Type and Objective Examinations
  3. Internal and External Examinations
IV. Construction and Administration of Examinations
  1. Construction
  2. Administration and Scoring
  3. Reporting Results
  4. Examination Standards
V. Examinations in Various Countries
  A. United States
  B. Great Britain
    1. England
    2. Modifications in Wales, Scotland and Northern Ireland
  C. France
  D. Union of Soviet Socialist Republics
  E. China
    1. Nationalist China
    2. People's Republic of China

I. GENERAL HISTORICAL BACKGROUND

An educational examination may be defined as the assessment of a person's performance when confronted with a series of questions, problems or tasks set him in order to ascertain the amount of knowledge that he has acquired, the extent to which he is able to utilize it or the quality and effectiveness of the skills he has developed. Formal examinations, involving written answers to a

series of prepared questions, were used by the Chinese in the 2nd century B.C. to select recruits for the civil service. Candidates were required to pass written examinations in literature and history and also to demonstrate their proficiency in writing poetry, in music and in archery. Those who were successful had to take the examinations afresh every nine years to retain their posts and the highest official, the prime minister, was expected to prove that he was entitled to his superior rank by competing regularly and winning the highest place.

In the universities of medieval Europe, students were examined at the end of their course by means of a disputation in which they had publicly to expound and debate points of logic and theology with their teachers and fellow students. The proceedings were observed and the individual student assessed by an examiner who sat on a three-legged stool, or tripos. The examinations for honours degrees in the University of Cambridge are still referred to as tripos examinations. In some modern Scandinavian universities, also, candidates for higher degrees are required to discuss the contents of their thesis or dissertation in a formal public debate and in the presence of examiners.

The Jesuits introduced written examinations into their schools in the 16th century. The definitive *Ratio atque institutio studiorum* of 1599, which was not revised until 1832, contains a code of rules for the conduct of school examinations which, apart from the fact that it is written in Latin, could be used today. These examinations were held annually and determined whether or not children were promoted to a higher class or grade. Competitions for prizes also were based on the results. The Jesuit system later came in for considerable criticism because of the emphasis it placed on rote learning and competition.

During the 19th century formal written examinations became regular in universities, schools and other educational institutions. Examinations were also increasingly employed for the selection of recruits to the civil service and the professions and to posts in industry and commerce. This development resulted first from increased public expenditure on education, which led to a greater concern for its efficient administration, and secondly from the demand for equality of opportunity for all citizens and consequently for appointments to be made according to ability and suitability rather than through influence or nepotism.

II. USES OF EXAMINATIONS

1. Evaluation of Progress.—Examinations devised by teachers to assess their pupils' progress must be as old as formal education itself, for they are an integral part of the process of teaching. Although their main and obviously essential purpose is to ascertain the amount of knowledge a pupil has been able to accumulate and retain, they have other important pedagogic uses. It is generally recognized that a pupil cannot be successfully educated if he is required to assume a wholly passive role. He must be given the opportunity to practise his skills if they are to be effectively developed and to apply his knowledge if it is to become systematically organized. Skilfully designed examinations not only enable a teacher to determine whether a satisfactory rate of progress is being maintained but also give the pupils themselves an opportunity to measure their own progress and to review their knowledge, to apply it to the solution of problems and, in the process, to recognize for themselves its limitations.

These assessments of progress also serve a number of subsidiary but important purposes. Rarely can a teacher give his pupils incentive to work hard over a long period without resorting to the use of tests or examinations. For the majority of pupils—and this is true also of relatively mature students—an examination of some kind is a spur to sustained effort. Whether or not this end is achieved depends largely on the way in which these assessments are designed and utilized. If in constructing an examination more emphasis is placed on discovering the pupils' shortcomings than on affording them the opportunity to show their knowledge, there is a grave risk of engendering apathy and resignation. Similarly if too much stress is laid on competition and the results are regularly expressed as orders of merit, the effect on some pupils may well be discouraging. It has been shown that, for this purpose, the most effective examination is one difficult enough to offer a challenge but not too difficult to allow most pupils to experience some success, and one which affords a comparison not so much between pupils as between each individual's performance and his own previously achieved standards.

Regular assessments also enable teachers to furnish meaningful reports on the progress of pupils to the head of the school, to the head of another school to which the pupil may transfer, to a prospective employer or to the pupil's parents.

Another useful by-product of the progress assessment is that it focuses the attention of teachers and pupils alike on the objectives they are jointly seeking to attain. If a teacher regularly sets examinations to test his pupils' progress he can scarcely avoid considering the ends for which his course of instruction has been designed. The examination results will indicate how far these objectives are likely to be attained and therefore the extent to which they can be regarded as appropriate goals for the pupils.

2. Evaluation of Effectiveness of Instruction.—An examination that measures the performance of pupils may also be used to evaluate the effectiveness of the teacher and of the methods he employs. A poor performance does not necessarily imply that a pupil is lacking in ability or application. He may have been inadequately instructed, supplied with insufficient or inappropriate textbooks or placed in circumstances not conducive to his scholastic progress. Examinations are increasingly used to investigate the conditions in which successful learning takes place, to compare the effectiveness of different methods of teaching, to examine the influence on pupils' progress of various forms of school organization, etc. The Educational Testing service in Princeton, N.J., the National Foundation for Educational Research in England and Wales and comparable institutions in other parts of the world devote a large part of their resources to producing tests and examinations for such purposes.

Diagnosis of Pupils' Needs.—Examinations designed to assess levels of attainment in a particular subject may reveal that certain pupils require some remedial attention. This, however, is only the first stage in diagnosis and it is probably true that the diagnostic possibilities of testing procedures are much in advance of the teachers' skill in using these instruments. Many tests are available, however, especially in the English-speaking countries, that can enable the skilful teacher to determine the pupil's special strengths and weaknesses in, for example, reading, arithmetic and language skills. Other tests can reveal the individual pupil's varying competence in the different subject fields at the elementary, secondary and higher education levels. Some tests are available to appraise the student's interests in different school subjects; his reading interests; his attitudes toward education, the school and the teacher; and his study habits and methods of work. These tests can be summarized in such a way as to make it possible to compare a student with other students of the same age, grade or level of general ability and to identify individual difficulties and their causes. The tests may also be used to determine whether particular remedial procedures have been helpful.

The effectiveness of remedial reading instruction, as an example, depends to a marked degree on the use of careful diagnostic testing procedures followed by specific remedial procedures adapted to the particular needs of the individual student. This selection of specific remedial procedures to meet the specific needs of each student is the basis of such programs.

3. Guidance.—In most modern schools examinations serve a wider purpose than that of progress assessment alone. Two major developments during the 20th century have combined to transform the organization of educational systems, particularly within highly industrialized countries: first, the considerable accumulation of knowledge about individual differences in pupils' ability and aptitude; secondly, the technological advances that have increased the complexity of the processes involved in the production and exchange of goods to such an extent that society demands specialized training, knowledge and adaptability. In these circumstances schools, particularly secondary schools, find it necessary to provide a wide variety of courses, both to cater for pupils who develop differing aptitudes and interests as they grow older and

also to prepare entrants to the steadily increasing number of new occupations. The allocation of pupils to appropriate educational courses is an important task, and a considerable amount of research is devoted to discovering suitable techniques for it. Tests and examinations designed to reveal the pattern of each pupil's abilities, aptitudes, attainments and interests play an important part in these procedures.

4. Selection.—A complementary process to that of guidance is selection of individuals for a particular educational institution or type of employment. While guidance involves determining the kind of education suited to a pupil's needs and the type of subsequent employment which will give scope to his aptitudes, selection is used to determine which individuals will be admitted to particular educational institutions or particular employment, and in many countries examinations are regularly used for this purpose. For example, more than 90% of the education authorities in England and Wales use examinations in English and arithmetic as part of their procedure for selecting entrants to secondary grammar (academic) schools.

Admission to courses of higher education in universities and other institutions is frequently determined, at least in part, by examination results. Selection procedures which include examinations are also used to fill posts in the civil service, in the various local government services and in many branches of industry and commerce.

5. Certification.—Another important purpose for examinations is to provide evidence that a person has achieved a specified standard of attainment. Examinations at the end of a secondary school or university course, for example, certify that an individual has completed the course successfully. Examinations of this kind are also set by professional institutions to determine whether or not a person should be regarded as a qualified member.

III. TYPES OF EXAMINATIONS

1. Written, Oral and Practical.—The form that an examination takes will depend on the purpose for which it is intended, the nature of the subject matter to be tested and the age and status of the examinees. A written examination may be used primarily to determine whether or not a candidate has remembered a specified body of knowledge or to ascertain how far he is able to organize and apply his knowledge. In some examinations of the latter kind, candidates are permitted to use notes and reference books.

Unless a written examination is of inordinate length it is unlikely either to enable a candidate to exhibit his knowledge and abilities in their entirety or to disclose all his deficiencies. An oral examination, conducted with skill, can do much to remedy these defects. It is sometimes possible for a candidate who has fared badly in a written examination to convince the examiners in an oral examination that he possesses ability and knowledge that his written answers did not make apparent. On the other hand many a student who has performed creditably in a written examination has had his inadequacies exposed by an oral examination.

Oral examinations are also used to measure skills which cannot be assessed by written examinations. In examining a modern language, for example, it is important to assess not only a person's ability to write and to translate the language but also his ability to use it in conversation.

Practical examinations involve the evaluation of an individual's performance when he is confronted with a concrete task. A person may well be able to demonstrate in a written examination that he understands, for example, the principles of radio transmission. Whether he also is capable of locating faults in a receiver and of repairing it efficiently can be adequately assessed only by a practical examination.

Examinations in some subjects may involve the use of all three methods. In the natural sciences, for example, written, oral and practical examinations are all appropriate. The form of examination most suitable for a particular purpose can best be determined by empirical methods. The technique of follow-up study has been used to gauge the efficiency of a number of examinations, by the recording of their results and by observations of the subsequent careers of the candidates.

2. Essay-Type and Objective Examinations.—*Essay Examinations.*—The most familiar kind of written examination is the so-called essay type. This consists of a relatively small number of questions to which, particularly from advanced students, answers of a fairly substantial length, written in the form of essays, are required. In assessing these the examiner inevitably exercises a certain amount of subjective judgment. Answers of this kind cannot be assessed as wholly right or wrong. They are usually awarded marks according to the extent to which the examiner considers that the candidate has introduced all the relevant facts or arguments or a sufficient number of them. The candidate's ability to write with clarity and precision, or even with what the examiner regards as a pleasing style, may also affect the result. Critics of the essay-type examination have discerned a number of inadequacies and limitations in it. They maintain that the results of these examinations can be shown to be unreliable in that they depend to a large extent on the subjective judgment of the examiners. It has been demonstrated that the marks awarded to the same papers by different examiners vary significantly; and that a given examiner, when invited to assess the same paper on two different occasions, is unlikely to award it the same marks. It can be shown, too, that the performance of the candidates is uneven. If two examinations, judged to be similar in content and in level of difficulty, are given to the same group of pupils, even with a very short intervening time interval, it is found that the pupils' performances and the relative gradings differ considerably. This is partly due to an error of sampling. Examinations of this type contain only a small number of questions, representing a tiny fraction of the total number that could be formulated to cover every aspect of the field of knowledge with which the paper is concerned. Thus the selection made on a given occasion is likely to be found more or less congenial by some candidates than by others, and similarly with the selection made on a different occasion. It follows that the larger the number of questions an examination contains the more adequately is it likely to sample the total area of knowledge under review and the fairer will be its estimate of each candidate's performance.

Critics of the essay examination also maintain that there is a tendency for marks to be awarded for skills other than those which are being ostensibly assessed. They hold it unfair that, for example, a candidate in geography who has a high degree of verbal ability will score better than one who is unable to express himself in writing with grace and lucidity, if both be supposed to have mastered the subject matter equally well.

Objective Examinations.—Largely for these reasons essay-type examinations have been found less suitable for a number of purposes than examinations constructed on more scientific lines and known as "objective" tests, usually involving some combination of multiple-choice problems. Such tests comprise a large number of questions, thus sampling more effectively than essay examinations the areas of knowledge with which they are concerned, and these questions are devised so that they yield only one acceptable response.

A marking key is usually provided by means of which any examiner can assess any script unambiguously without having to rely on subjective judgment. Some of these tests are standardized through administration to large samples of individuals, and thus it is possible to relate the performance of one person to that of a large number of others of the same age and status.

In the United States many examinations in schools and at college and university level are of the objective kind. In England these latter have largely replaced the traditional variety in the procedures employed for grammar school selection, although they have not as yet found favour in the grammar schools themselves or in universities.

3. Internal and External Examinations.—Examinations may be either internal, *i.e.*, set within a school or other educational establishment by its own teachers, or external, *i.e.*, constructed and administered by outside bodies. Internal examinations enable both teachers and pupils to estimate the progress that has been made and the effectiveness of the instruction that has been given, but the value of the results is limited by the narrowness of the

basis of comparison. The more advanced a student becomes the more desirable it is that his attainments should be assessed against a wider background than his own class or establishment, especially in view of the requirements of employers, heads of institutions for further education and those responsible for admitting entrants to courses of professional training; and external examinations provide such standards.

The disadvantages of the external examination system are that it serves to dictate the curriculum to be followed and to some extent the methods of teaching to be employed in the schools or other institutions which prepare the candidates. Thus, it is argued, teachers are robbed of their freedom to provide the kind of education which they consider best suited to their pupils' needs. It is also argued that external examiners cannot take special circumstances into account in making their assessments, while a teacher who knows his pupils intimately is able to judge whether or not a particular examination performance is representative of a candidate's real abilities. Nevertheless, external examinations are highly valued because of their apparent objectivity and freedom from personal bias.

To counter the objections mentioned above a type of examination is sometimes employed which is internally set and marked but is subject to external assessment, the external assessor being empowered to modify both the paper and the marking. The charters granted by the privy council to the modern English universities require them to appoint an external examiner to each panel of examiners for degrees. Similarly the examinations for the ordinary National certificate in England, which is the first of a series of qualifying stages through which aspiring technologists have to pass, are conducted by the individual technical colleges but assessed by external examiners approved by the ministry of education. Although the examinations for the general certificate of education given to senior pupils in secondary schools are external, teachers serve on the examining boards, advise and assist those who prescribe the syllabus and set the questions and share in the marking of the papers.

IV. CONSTRUCTION AND ADMINISTRATION OF EXAMINATIONS

1. Construction.—The development and widespread use of standardized, objective tests have influenced the attitudes of teachers and others concerned with devising examinations.

It is important first of all clearly to define what an examination is to measure. It is insufficient merely to outline the ideas or topics to be included. A detailed list must be prepared of the specific items of knowledge, the skills and aptitudes which are to be assessed. The specifications should also be drawn up in terms of the objectives of the study course. In a science examination for senior pupils or for university students, for example, these may include, as well as knowledge of specific facts, terms and concepts, the ability to apply principles to new problems, to comprehend research reports, to design experiments for testing hypotheses and to set up and use laboratory equipment efficiently. In planning the examination it is important to ensure that each of these separate skills is tested adequately.

Questions appropriate to each area of knowledge and type of skill must then be chosen. Each type of question or problem to be used must be selected because it is effective in sampling the kinds of thinking, skill or behaviour included in the specifications. A great deal of research has been done, especially since about 1925, to identify the kinds of tasks which most effectively measure different types of educational objectives. This has revealed the place of essay questions, multiple-choice questions, open-book examinations, observation of the student and his work and many other ways of collecting evidence in the development of good examinations. Each technique has been found to have a place.

The questions chosen must meet two requirements. Firstly, in their scope and variety they should form a representative sample of the hypothetical total list of questions that could be posed. Secondly, each should be scrutinized to ensure that it is unambiguously worded and that it is of the appropriate level of difficulty. If the questions are not tested empirically by being administered to a group of students or pupils comparable to those for whom they are intended, they should be carefully appraised by a group of experienced teachers.

2. Administration and Scoring.—The special conditions under which the examination is to be given must be considered. The amount of time permitted, the availability of references, notes, etc., and the method to be used by the student to record his answers to the questions must be determined. How written essay responses are to be judged in order to ensure accuracy and comparability of judgments must be decided. The examiner must also determine how the answers are to be graded or scored and the relative weight to be given each question or problem if the results are to be summarized into a single total score.

It may be necessary to combine the marks from a series of examinations in different subjects into one grand total. Statistical techniques have been developed by means of which the marks derived from a number of different examinations may be readily translated into sets of strictly comparable scores.

As marks have no absolute value, the form in which they are expressed must be to some extent arbitrary and will depend on the form which the examinations have taken—for example, whether they were competitive or diagnostic.

3. Reporting Results.—The translation of scores on the examination into pass or fail or into different levels of pass and fail marks must be determined. Usually a small committee of examiners determines this very crucial step in the examining process for certification purposes.

The examiner must also determine how to report the results for the different uses to be made of the examination, as for diagnostic purposes, for summarizing the results for the entire group of examinees or for summarizing and analyzing the results for the purpose of research and school comparison. Finally, the method of analyzing the examination results as a basis for the future improvement of the examination itself must be determined.

4. Examination Standards.—If examinations are competitive, it matters little if their level of difficulty fluctuates slightly from one occasion to another. Qualifying examinations, however, must be kept as close as possible to a defined standard. It would be manifestly unfair if, in England for example, the standard of the general certificate of education varied from year to year.

Certain precautions can be taken to prevent excessive fluctuations in difficulty and in marking standards from year to year. The most satisfactory method, although costly and time consuming, is that of calibration. This involves administering an examination to a sample group of candidates, calculating the average marks obtained and observing the distribution of the scores. A second examination, as nearly similar as possible to the first, is then administered to another group. Any difference between the scores yielded by the two examinations can then be noted and, if necessary, the second examination can be adjusted or appropriate allowances can be made when the marking is done to ensure that the same standards apply to both sets of candidates.

An alternative procedure is to keep constant the proportion of those who qualify and those who fail in an examination. This involves the assumption that the successive groups of candidates are equal in ability, which is amply justified if the number of entrants is always large.

See also PSYCHOLOGICAL TESTS AND MEASUREMENTS: *Characteristics Measured, Construction and Uses.*

V. EXAMINATIONS IN VARIOUS COUNTRIES

A. United States

The evolution of the examination system in the United States is comparable in general outline to that in the western hemisphere. In the early colonial colleges, for example, which were founded in the 17th century, students were examined orally in public. Before a large and learned assembly, a student was required to deliver an oration on a theme proposed by his tutor and his performance was judged by a panel consisting of teaching members of the college. This method of examination had obvious disadvantages. A student's performance in these circumstances did not necessarily afford a true indication of his academic capabilities. Comparisons

between students were difficult to make since each was required to respond to a different question or suggested theme, and assessments were subjective and unreliable. Consequently, by the middle of the 19th century these public exhibitions were rapidly giving way to written examinations.

By the beginning of the 20th century, written examinations were established at several levels: for the selection of entrants from high school to college, for the evaluation of students' achievements in college and for the assessment of professional qualifications of various kinds. During this century written examinations were developed in three ways. First, they were modified, in universities and colleges, to provide a more general and comprehensive assessment of students' attainments. Dissatisfaction with the limited objectives of the type of examination designed to evaluate a student's attainment at the end of a single elective course led to the development of examinations which had the more ambitious aim of estimating the degree to which he had succeeded in recognizing or establishing relationships among the various areas of knowledge with which he had become acquainted. Secondly, the 20th century saw the rapid growth of external examinations. Partly in order to separate the functions of teaching and examining and partly to resolve the chaos created by the multiplicity of requirements for college entrance, bodies such as the College Entrance Examination board (1900), pioneered by Nicholas Murray Butler, were set up to provide common entrance examinations whose results gradually became accepted and used by the majority of colleges. Thirdly, extensive experiments were conducted with a view to improving the construction of written examinations.

Standardized objective tests are used in the United States more than in other countries. Many colleges and universities use scholastic aptitude or intelligence tests in addition to the high school record as a basis for admitting students. Aptitude tests are also widely used in educational and vocational guidance. They are designed, not to yield a single index of the level of a person's general ability, but rather to provide a profile showing his specific strengths and weaknesses.

After 1930 many changes were introduced into the curriculums at the elementary, secondary and general education levels. In order to assess the effectiveness of these experimental courses and programs new types of tests and examinations were developed. Attempts were made, with some success, to devise tests to measure some of the more intangible results of educational programs, such as attitudes, interests, emotional maturity and improvements in critical thinking and aesthetic appreciation. The development and use of these tests not only served to evaluate the results of some of the experimental courses and programs but also stimulated teachers to analyze more closely the processes involved in education and to define with greater precision the objectives toward which their courses were directed. For example, the vague notion of developing appreciation of music was made increasingly precise and meaningful by defining just what a person does when he appreciates music, by defining the reactions he is expected to have when an excerpt of music is played for him in a test and by planning learning experiences that would help him to develop these reactions.

The results of tests of this kind helped teachers to find ways of improving their courses. They were also found useful for grading students and as diagnostic instruments.

Research into the construction and use of objective tests and examinations has been undertaken on a larger scale in the United States than in any other country, and the training of teachers generally includes courses in the development and use of these tests. As a consequence a large proportion of the teachers are more knowledgeable about this kind of test than are their counterparts elsewhere. The result has been a sufficiently large demand for standardized objective tests to justify the setting up of a number of specialist agencies such as the Educational Testing service in Princeton, N.J. These agencies supply to schools, colleges and other institutions a large variety of tests for the assessment of students' abilities, aptitudes and attainments. Some of these tests can be scored mechanically so that the results even of a large-scale testing program can be made speedily available.

During World War II many men and women in the armed forces took advantage of correspondence courses, study groups and training programs provided by the military authorities. In order to assess the competence of these students and to certify it to schools and employers, examinations were prepared by The University of Chicago for the examinations staff of the Armed Forces institute. Through these examinations more than 2,000,000 persons were given advanced credit when they returned to their high schools or colleges after military service. One of the consequences of this was that colleges began increasingly to use placement tests to determine the level at which each student should begin his college work. Such tests, it is believed, help to motivate students and to reduce academic failure.

By assessing for each student his levels of attainment it is possible to exempt him from work in which he is already demonstrably competent and to provide courses that are appropriate to his needs.

B. Great Britain

1. England.—In primary or preparatory schools most children become accustomed to examinations administered by their teachers. At the age of 11 the vast majority of children take an external examination set by the local education authority to determine the kind of secondary school courses to which they should be transferred. It is at this point that U.S. examination practice, as well as theory and philosophy, contrasts most sharply with that of Britain and most other countries. Under the U.S. common school or single-track plan any student who passes the 8th grade is eligible to pass directly from elementary school into a single secondary or high school. This tradition, as distinct from the British and European specialized or multitrack system under which students are assigned either to grammar schools or to vocational or other specialized schools, denies the reliability of an examination to determine which child shall be given or be deprived of the opportunity to obtain an education that would qualify him for college entrance (*see* Secondary Education). In Britain regular examinations continue during the secondary school course and, finally, grammar-school children and a growing number of children in secondary modern schools enter for the general certificate of education examinations. Other children enter for the examinations of the Royal Society of Arts, first held in 1858, or the College of Preceptors, whose first examination, open to girls as well as to boys, took place in 1853, and of other bodies; and a number of local education authorities offer school-leaving certificates to those who pass their examinations. Young people who proceed straight to universities and colleges of further education continue to take examinations. Those who start work may attend part-time courses in technical colleges or other establishments leading to qualifying examinations. Few people can win promotion or even obtain work without passing some kind of examination.

Although some educationists criticize examinations on the grounds that they distort the curriculums of the schools and encourage wasteful cramming at the expense of healthy, disinterested study, it is clear that the majority of English people consider them to be effective instruments for the evaluation of educational progress and for vocational selection. Their widespread use, however, is a relatively recent development, even in schools and universities. Until the end of the 18th century such examinations as existed followed the medieval pattern. The adoption by Oxford university of the Public Examination statute of 1800 marked the close of this period. This statute provided for both honours and pass degrees and was subsequently modified to require an increasing amount of written work. During the first half of the 19th century there was rapid reform and innovation, which included the inauguration of the Cambridge classical tripos in 1824 and culminated in 1858 with that of the Oxford and Cambridge local examinations for schools. The following 100 years saw the full development and refinement of the modern system.

Origins of Secondary and Professional Public Examinations.—The way in which the University of London developed exercised a considerable influence on the part public examinations came to play in English secondary education. The first charter (1836) instituted London university as a body charged with examining stu-

dents of two hitherto independent and rival colleges, University college (1828) and King's college (1829). In 1850 a revision of the charter allowed affiliation of other institutions inside and outside London, and the university's next charter (1858) opened its degree examinations, except those in medicine, to any student irrespective of where he had studied, thus establishing London as the purely examining university which it was to remain until 1900, when a new charter reconstituted the university to teach as well as to examine and produced the dichotomy of internal and external degrees.

Oxford and Cambridge had introduced written examinations in order to improve teaching methods and encourage their students to work; the first half of the 19th century also saw the ancient learned professions adopting written examinations to ensure the competence of new entrants. The Law society, which in its modern form dates from 1825, took a large part in having the mere formality of the earlier oral examinations replaced by written examination.

The Apothecaries act (1815) set going reform of medical and surgical training and practice. The insistence of the Society of Apothecaries that entrance to the profession should be by examination caused private teaching establishments to spring up throughout England and those in, for instance, Birmingham, Liverpool and Manchester developed into medical schools which in their turn contributed to the founding of the present-day universities in these cities. In 1844 the Royal College of Veterinary Surgeons was established and in 1852 the Pharmacy act gave rise to the Pharmaceutical society; both of these were examining bodies. This half century of reform culminated in the act of 1858 out of which came the General Medical council. The G.M.C. neither teaches nor examines; it keeps the register of those who satisfy its professional requirements in virtue of qualifications gained at approved teaching and examining institutions.

This period also saw the opening of the civil service in India to entry by competitive examination (1854), the first public competitive examination for entry to the Royal Military academy (Woolwich) and thereby to commissioned service in the army (1855) and the appointment of the civil-service commissioners, also in 1855. This latter step marked the victory of the claims of competence tested by examination over those of birth and patronage. The victory was not everywhere acclaimed.

Modern Secondary Public Examinations.—In the evolution of English secondary education external examinations were called upon to play that part in setting and maintaining standards which in France, for instance, was taken by the government. England had no board of education until 1899 and no "secondary education for all" until the Education act of 1944 adumbrated the conception.

This act abolished fee paying in all maintained schools and therefore concentrated public attention upon the 11-plus examination, as it is popularly styled. This examination, conducted by local education authorities, helps to determine the kind of secondary school course to which each child is allocated at or about the age of 11. Although its content varies from one locality to another the commonest form of the examination consists of verbal intelligence tests and standardized tests of attainment in English and arithmetic supplemented by assessments from the children's primary-school teachers. A number of large-scale investigations, involving follow-up studies, have been made into the effectiveness of this examination and have demonstrated it to be efficient, despite the amount of public criticism leveled against it.

The history of secondary-school examinations after the mid-19th century falls into three clearly defined stages discussed below: (1) from 1858 to 1917, when these examinations were conducted mainly by bodies independent of the central government; (2) after 1918, when external examinations for pupils in state-aided secondary schools were limited by the board of education to those conducted by the eight examining bodies approved by the board; (3) after 1946, when the ministry of education (which had replaced the board in 1944) assumed responsibility for external school examinations and, in 1951, introduced the general certificate of education.

1. The Education act of 1902 instituted a national system of secondary education financed by public funds. Before this date the Oxford and Cambridge local examinations and the external examinations conducted by the University of London had provided a goal for pupils in the existing secondary schools and for part-time students in technical and evening institutes. With the growth of the public secondary-school system other examining bodies were created. The Joint Matriculation board of the universities of Manchester, Liverpool and Leeds was set up in 1903. The University of Bristol also became a school examining body in that year. Previous to this a number of professional bodies had begun to conduct their own entrance examinations and by the beginning of World War I there were more than 100 external examinations for which secondary school pupils could be entered.

2. The president of the board of education reduced this chaos to order by ordaining that from 1918 no grant-aided secondary school could present its pupils for external examinations other than the first examination (school certificate) and second examination (higher school certificate) of eight approved examining boards; *i.e.,* Oxford, Cambridge, Oxford and Cambridge Joint board, London, the Joint Matriculation board, Durham, Bristol and the Central Welsh board.

The school certificate examination was for pupils of about the age of 16 who had completed the normal five-year course in a secondary school. The credit or higher pass standard in the individual subjects of this examination was designed to be equivalent to the previous matriculation standard, so that universities and professional bodies could accept the school certificate instead of conducting their own entrance tests. The higher school certificate was intended to be taken two years later at the end of the sixth-form course. The candidates were required to specialize in a group of subjects as follows: classics, modern studies, mathematics or mathematics and science. Local authorities used the results of this examination as a basis for awarding university entrance scholarships and from 1920 the board of education awarded state scholarships to universities upon it. The growth of the secondary-school population between World Wars I and II, and the consequent increase in the proportion of grammar-school children who did not aim to pursue an academic career, engendered doubts about the suitability of the school certificate and higher school certificate examinations, whose syllabuses were largely dominated by university requirements.

3. In 1946 responsibility for secondary-school examinations was assumed by the minister of education, and in 1951 the school certificate and higher school certificate examinations were abolished in favour of the more flexible general certificate of education (G.C.E.). The G.C.E. is a subject examination; *i.e.,* a candidate may offer one or more subjects and is awarded a certificate which indicates those in which he has passed. There are two main levels, ordinary and advanced, the ordinary level pass being equated with the credit standard in the former school certificate and the advanced level with a pass in the higher school certificate. It is also possible for pupils who wish to qualify for scholarships to universities to take papers of a higher level still, which are labeled "scholarship" papers.

The universities and professional bodies are able to use the G.C.E. for selection of entrants by naming the subjects in which they require passes. Also, because of the possibility of entering candidates for only a few subjects out of a large range, a growing number of secondary modern schools find it suitable as an examination for their older pupils. Perhaps the most striking development in English secondary education during the decade 1950 to 1960 was the steadily increasing number of secondary modern pupils who were successful in the G.C.E. During 1959, for example, about 15,500 pupils from these schools passed, compared with about 10,500 in the previous year.

2. Modifications in Wales, Scotland and Northern Ireland.—*Wales.*—Although the general pattern of examinations in Wales is similar to that in England, certain modifications were introduced as a result of determined local effort to defend the interests of the Welsh-speaking sections of the community and to preserve the Welsh language.

During the 19th century, Wales was provided with a system of elementary education; but "payment by results," introduced by Robert Lowe (later Lord Sherbrooke) in the Revised code of 1862,

seriously threatened the position of the Welsh language, because it laid down that grants to the elementary schools were to depend on the demonstration of their pupils' proficiency in arithmetic and English. As a result, Welsh was not only discouraged in the schools but was often expressly forbidden. Many of the head teachers, who, for the most part, spoke only English, saw the encouragement of Welsh as a threat to the recognition of their schools' efficiency; and so it became common for children to be fined or otherwise punished for speaking Welsh.

As a result of public outcry against this impediment to the preservation of the Welsh language, the terms of the Revised code were modified in 1888 to allow Welsh to become, with English and arithmetic, one of the specified subjects which could qualify schools for financial grants. The Balfour act of 1902 gave control of elementary education to local authorities, so that the position of the Welsh language in the schools was from this date safeguarded. One example of the way in which the interest of Welsh-speaking children were thenceforth taken into account is that several of the Welsh authorities provide alternative papers in the examinations held for the allocation of children to secondary schools. A child may opt to answer a paper in whichever of the two languages he prefers; and Welsh versions of some of the standardized objective intelligence tests have been prepared for use in these examinations.

Scotland.—The examination system in Scotland is similar to but not identical with that in England and Wales. The transfer of pupils from primary to secondary schools takes place at the age of 12 by an examination procedure comparable with the 11-plus. Some Scottish secondary schools enter their pupils for the English G.C.E., but the majority prepare them for the examinations set by the Scottish education department. This department issues a lower grade and a higher grade leaving certificate, with standards corresponding roughly to the ordinary and advanced levels of the G.C.E. Children who attend junior secondary schools (the equivalent of English secondary modern schools) rarely attempt the higher grade. This is sought by pupils of senior secondary schools and, like the advanced level of the G.C.E., is used by universities and colleges to select entrants. University and professional examinations in Scotland are similar to those in England except that in Scottish universities there is no bachelor's degree and the final examinations qualify directly for a mastership.

Northern Ireland.—Northern Ireland has its own ministry of education, which administers secondary-school examinations. A junior certificate examination, of a rather lower standard than the G.C.E. (ordinary level), serves as a school-leaving certificate for the majority of secondary-school children and for some of those who attend secondary intermediate schools, which correspond to English secondary modern schools. Pupils who pursue the full academic course in grammar schools are entered for the senior certificate, on the results of which entry to universities and technical colleges is determined and which is also accepted by many professional bodies as a preliminary qualification.

C. France

There could be no greater contrast with both U.S. and English practices than that offered by the centralized orderliness of the French system. At its root lies Napoleon I's Université de France, and it is based upon the 17 *académies* which cover the whole of France, each with a university at the centre of the regional organization. Examinations form an integral part of the system and they are under the direct administration and control of the *ministère de l'éducation nationale,* which is organized in four departments (*premier degré, deuxième degré, technique, supérieure*); the ministry is ultimately responsible for examination syllabuses and question papers and for marking the scripts. The examinations are therefore completely external, and insofar as members of the teaching staffs of universities and *lycées* take part in examining they do so as functionaries of the ministry. In 1905 the central government extended its control, including the administration of examinations, to cover the schools and colleges of the religious orders.

The reorganized scheme of the 1950s requires that French children take an external examination at the school-leaving age when they sit for the *certificat d'études primaires.* For many of them a *cours complémentaire* leads at the age of 15 to the external examination for the *brevet élémentaire* (*brevet fin d'études*); alternatively, after a course at an *école professionnelle, a brevet professionnel* gives entrance to a trade or to technical employment. If qualifications higher than the *brevet* are sought later, examinations on which such diplomas are awarded are controlled by the ministry's *département d'enseignement technique.*

If the pupil aims at an academic career or one in law, medicine, engineering, etc., at the age of 11 years he seeks entry to a *lycée* by means of the common state entrance examination, highly competitive and held simultaneously throughout the country. If selected, the pupil enters the *classe de sixième* and progresses through the school to the *classe de première* and on to the *première supérieure* by means of annual internal *examens de passage.* The *première partie* of the *baccalauréat* (*bachot*), an external examination common to all *lycées,* is taken at the age of at least 16 and, if this test of general education is passed, the *seconde partie* follows with a choice of nine specialized groupings of subjects. Until 1950 success in the *seconde partie* qualified for entrance to a university or polytechnic; since then an *année propédeutique* has been interposed for liberal-arts students which is spent in preparation for university study proper. At the end of this year there is the external examination for the *certificat d'études littéraires générales.*

At the university a student is at liberty to choose the order in which he will take the required examinations, and attendance at lectures is not compulsory. After a minimum period of three years from matriculation he may sit for the *licence.* One or two years later the *diplôme* may be obtained on submission of a thesis; this qualifies the individual for appointment by the *recteur* of the university to a post on the teaching staff of the regional *académie.* If he seeks promotion, the *licencié* will in due course seek *agrégation* by means of a severe and highly competitive examination in which the number of awards is strictly limited. A teaching qualification of a somewhat lower standing was instituted in 1952, the *certificat d'aptitude pédagogique à l'enseignement secondaire.* After ten years or so an *agrégé* may seek the *doctorat ès lettres* by defending two theses in public.

All these university examinations are under the control of the ministry. If instead of entering a university the student proceeds from a *lycée* to a *grande école,* the examinations for professional and technological qualifications are also under ministerial control. All external examinations including that for the *certificat d'études primaires* include written and oral tests and, by requiring that all oral examinations from those for the *bachot* upward must be open to the public, the French have increased the severity of the candidates' ordeal and preserved a living relic of medieval disputations. Since 1957 students, especially of technical and vocational schools, who do not hold the *bachot* may proceed to higher education by means of a special university entrance examination.

D. Union of Soviet Socialist Republics

In the U.S.S.R. education is an agency of government employed in the operation of the state (*see* School Administration: *Political Party Control*). Examinations in the higher educational establishments and specialized secondary schools or *technicums* are controlled by the ministry of higher and special secondary education. Nonspecialized secondary schools and primary schools are controlled by the ministries of public education in the individual republics, but the central authority issues statutes which govern examinations, curriculums and the choice of textbooks. Promotion from class to class in primary and secondary schools in the U.S.S.R. is determined by the pupils' records but an examination in Russian, arithmetic and the children's native language is held at the end of the fourth year of the course. The results of this examination, in some areas, determine the nature and length of the secondary-school course to be followed by the pupils.

The *technicums* prepare "middle-level specialists"; *e.g.,* nurses, medical aides, industrial technicians, etc. The students are given an examination at the end of the fourth year of the course. Those

who do well proceed directly to higher education, with science majors receiving preference. The rest are required to work for one to three years before passing on.

Part-time courses are organized for young workers whose education stopped at the primary level. Examinations are held at the end of these courses and those who qualify may proceed to institutes of higher education.

Entry to these institutes and to universities is by competitive examination. Annual internal examinations are also held, which the students must pass in order to gain promotion to the next stage. At the end of the course the examinations used for the award of degrees and diplomas are controlled by the state and not by the individual institutes or universities.

E. China

1. Nationalist China.—In 1956 and 1957, Nationalist China carried out a program to raise and equalize standards in institutes of higher education, and this included the arranging of joint entrance examinations for 26 colleges and military schools.

2. People's Republic of China.—There, the educational system is being reformed in accordance with the "common program" introduced in 1951. As in the U.S.S.R., centralized control of all aspects of education is exercised directly by the ministry of education over all universities and institutes and, indirectly, through regional and provincial authorities, over secondary and primary schools.

The conduct of all examinations is governed by the ministry. State examinations are sent to pupils in secondary schools and are used to select entrants to universities and higher technical institutes. Considerable use is also made of examinations to encourage adults to improve their educational standards. Examinations are held for the award of literacy certificates (a vocabulary of 1,200 words has been defined as the minimum requirement) and it is also possible for workers and peasants who have had no more than elementary schooling to qualify by examination for special short-term secondary schools and to graduate from these to People's universities. *See* also Education, History of.

Bibliography.—G. R. M. Ward (trans.), *Oxford University Statutes,* 2 vol. (1845–51); H.M.S.O., *Oxford University* (1852), *Schools Inquiry* (1864); H. Latham, *On the Action of Examinations Considered as a Means of Selection* (1877); Lord Bryce, *Secondary Education* (H.M.S.O., 1895); C. Cookson (ed.), *Essays on Secondary Education* (1898); A. D. Godley, *Oxford in the Eighteenth Century* (1908); H.M.S.O., *Examinations in Secondary Schools* (1911); A. I. Tillyard, *A History of University Reform from 1800 A.D. to the Present Time* (1913); P. B. Ballard, *Group Tests of Intelligence* (1922); J. M. Crofts and D. C. Jones, *Secondary School Examination Statistics* (1928); C. W. Valentine, *The Reliability of Examinations* (1932); A. M. C. Saunders and P. A. Wilson, *The Professions* (1933); Sir P. J. Hartog and E. C. Rhodes, *The Marks of Examiners* (1936); H. Rashdall, *The Universities of Europe in the Middle Ages,* 3 vol. (1936); E. R. Smith and R. W. Tyler, *Appraising and Recording Student Progress* (1942); J. L. Brereton, *The Case for Examinations* (1944); National Society for the Study of Education, *The Measurement of Understanding* (1946); J. B. Conant, *Education in a Divided World* (1948); N. A. Hans, *Comparative Education . . .* (1949); C. M. Fuess, *The College Board: Its First Fifty Years* (1950); E. F. Lindquist (ed.), *Educational Measurement* (1951); A. B. Moehlman and J. S. Roucek (eds.), *Comparative Education* (1951); K. E. Priestley, "China," *Year Book of Education* (1952); A. F. Watts, D. A. Pidgeon and A. Yates, *Secondary School Entrance Examinations* (1953); J. A. Petch, *Fifty Years of Examining: the Joint Matriculation Board, 1903–1953* (1953); N. C. Kearney, *Elementary School Objectives . . .* (1953); H. H. Remmers and N. L. Gage, *Educational Measurement and Evaluation,* rev. ed. (1955); B. S. Bloom (ed.), *Taxonomy of Educational Objectives . . .* (1956); P. E. Vernon (ed.), *Secondary School Selection* (1957); A. Yates and D. A. Pidgeon, *Admission to Grammar Schools* (1957); UNESCO, *World Survey of Education,* I (1955), II (1958); O. K. Buros (ed.), *The Fifth Mental Measurements Year Book* (1959); Ministry of Education, *Education in 1959* (H.M.S.O., 1960). (A. Ya.; X.)

EXARCH, in the Byzantine empire the title of certain chief officials, both secular and ecclesiastical, and thereafter of certain officials in the Eastern Church. The chief military and naval commanders in the early Byzantine empire were exarchs, as were the provincial governors of Africa and Italy (the latter being known as the exarch of Ravenna). A superior in charge of several monasteries, bishops of important sees and patriarchal legates also were called exarchs. In modern times an exarch is a deputy appointed by a patriarch to exercise jurisdiction over certain parishes abroad. Thus the patriarch of Constantinople appoints an exarch over the Greek Orthodox in North and South America, as does the patriarch of Moscow over the parishes of his jurisdiction in western Europe. The head of the Bulgarian Orthodox Church was an exarch before he became patriarch in 1953.

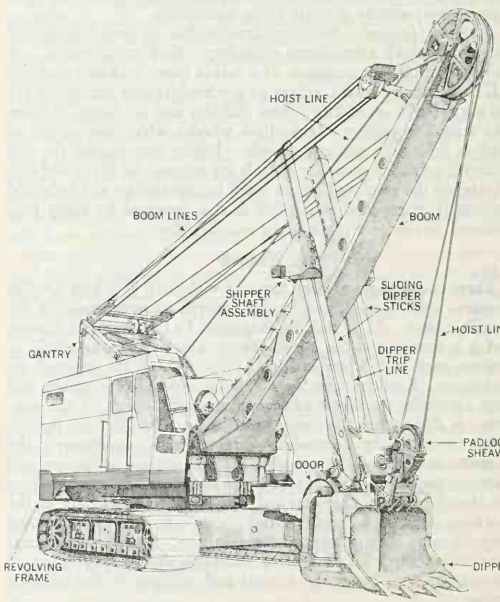
(H. M. W.)

EXCAVATION. Excavation includes the processes of loosening, digging, transporting, spreading and shaping soil and rock. Large-scale work is done by machinery wherever it is available.

Power Shovels.—The most important of the primary loosening and digging units are the power shovels with their many different pieces of interchangeable operating equipment. The basic equipment consists of a track or wheel-mounted superstructure, known as the deck, on which there is a power plant or power plants, boom, winches, cab, controls and, perhaps also, counterweights when the boom is very long. The deck of most shovels can be revolved so that the equipment may be used on any side.

Power Plants.—The power plants of shovels fall into two classes: (1) single engine, in which any type of internal combustion, electric or diesel-electric engine may be used, the power application being controlled by clutches; (2) multiple engined, in which a series of steam, electric or diesel-electric units are used with individual and direct-power drive to each operation.

Mountings.—A wide variety of mountings are available for most units; those commonly employed are: (1) crawler mounting, in which the heavy-duty chassis has two continuous, parallel crawler belts made of a series of tread shoes or links, supported by axles and driven by a power plant on the deck; the power plant may be the primary plant used by the shovel or may be a separate unit; (2) truck mounting, a heavy framed, rubber-tired chassis with two or more axles that has its own power plant and has the general characteristics of a heavy-duty motor truck; (3) wheel crane, a heavy framed, rubber-tired chassis with two or more axles that is driven by the primary power plant of the crane; (4) rail mounting, more commonly known as a locomotive crane, consisting of a heavy frame mount fitted with flanged railway



BY COURTESY OF HARNISCHFEGER CORP.

FIG. 1.—DIPPER SHOVEL

wheels so that it is able to run on rails under the power of the primary power plant; (5) other mountings, generally permanent or semipermanent, in which the deck is attached to piers, boat decks, barges and many other places not requiring a movable mounting.

Dipper Shovel.—This is the oldest and the most important machine used for hard excavation and for loading trucks. Primitive rail-mounted models were in use in 1835. A dipper stick is hinged to the midpoint of the heavy boom by a block through which it slides. The forward end of the stick carries the bucket or dipper, which is equipped with digging teeth and a bottom door held by a spring latch controlled by the operator. Capacity ranges from $\frac{3}{8}$ to more than 65 cu.yd. The boom is supported by a wire rope rigged from a deck winch over overhead sheaves (*see* fig. 1). The bucket is hoisted by another winch line extending from the deck over boom point sheaves. It is crowded (driven) forward and pulled back by a dual wire rope or roller chain set passed around the stick hinge and attached to the ends of the stick. The operator, by means of manual or power controls, can move the shovel forward and backward and steer it to either side; can revolve the deck and boom; raise, lower, crowd and retract the bucket; open and shut the bucket door and raise and lower the boom.

Power requirement varies from 40 to 12,000 h.p. A small shovel may have a gasoline engine; one of $\frac{3}{4}$ to 4 or 6 cu.yd. capacity is likely to have a diesel engine; larger machines have several electric motors, whose current may be supplied by a diesel generating set on the deck or on the ground, or by an external power line.

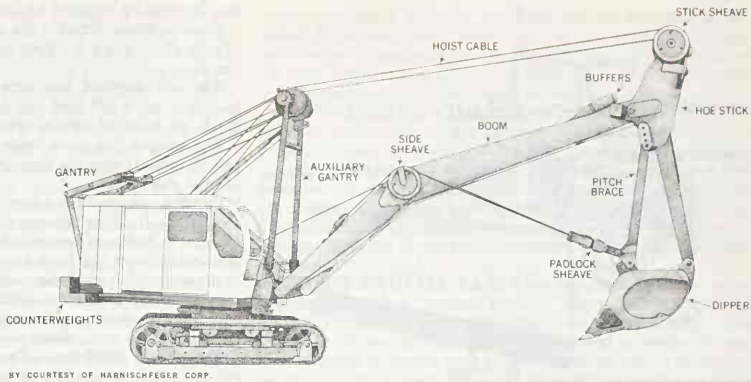
Small machines can make a complete cycle (dig, swing, dump, return swing) in 15 sec., while large ones may need 60 sec. Minor delays are estimated to cause loss of 10 to 15 min. in each hour. Most shovels dig between 70 and 170 cu.yd. per hour per cubic yard of bucket capacity.

Hoe.—The dipper shovel's boom, stick and bucket may be replaced with a hoe or backhoe attachment, designed for digging below the working level, that pulls toward the shovel instead of crowding away from it. The boom is raised and lowered with every digging motion. The stick is hinged to its end and does not slide. The bucket usually has no door and is dumped by raising and overturning. A vertical jack boom supports sheaves that carry the hoist line to the back of the stick. A double drag cable pulls the bucket inward for digging (*see* fig. 2).

There are also hoes of $\frac{1}{4}$ to $\frac{5}{8}$ cu.yd. capacity that are operated by hydraulic rams instead of cables, and are usually rear-mounted on tractors and trucks. The "wrist-action" bucket is hinged to the stick so that the operator can change its angle at any time for more efficient digging and dumping. The bucket may be reversible for use as a dipper shovel. A power-telescoping boom may permit pushing and pulling grading blades and other special attachments.

Lifting Crane.—The lifting crane is one of the interchangeable pieces of equipment that are generally used on the deck noted above.

The lifting crane boom and equipment is used to raise heavy loads, move them laterally by swinging or traveling and then lower them into new locations. The crane boom is a columnar structure of chord and lattice design pivoted at its connection with the deck and supported by cables at the top so that it may be raised or lowered. The hoist cable which runs over the top of the crane boom may be equipped with a hook, hook block, sling, tongs or a variety of similar load carrying devices, or by an electric lifting magnet for such materials as pig or scrap iron. A power-driven or gravity hammer may be used with a crane boom to drive piles. (H. L. N.; H. S. B.)



BY COURTESY OF HARNISCHFEGER CORP.

FIG. 2.—BACKHOE SHOVEL

Dragline.—The dragline has a long light boom of lattice construction that is held at a 25° to 35° angle to the ground. A bucket is hung from its point sheave by the hoist winch line and is drawn toward the shovel by the dragline. In digging, the bucket is lowered to the ground under the boom point or swung pendulum fashion out beyond it and is pulled in, cutting as it moves. When full, it is hoisted; then the earth is dumped by releasing the drag-rope.

Bucket sizes range from $\frac{3}{8}$ to 35 cu.yds. Booms are from 25 to about 275 ft. long. Increasing the length of the boom on a machine slows its cycle and requires a reduction in bucket size or installation of counterweight on the rear of the deck. With a standard boom and bucket production is only slightly less than that of a dipper shovel of similar capacity. A dragline cannot dig hard soil effectively and control of the bucket is not positive, but it digs below its own level in soil, mud and under water, moves material farther in sidecasting and can build high piles.

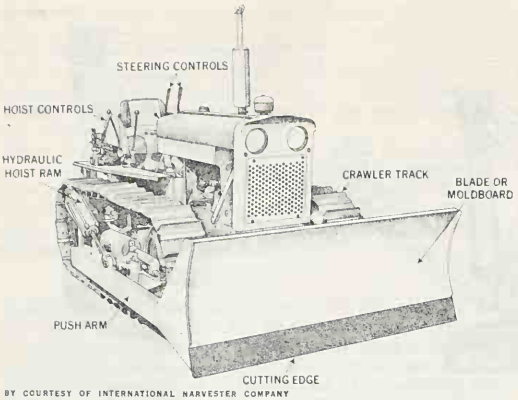
A "walking" dragline rests on a flat base and moves by means of shoes mounted on cranks on the ends of a cross shaft on the revolving deck.

Clamshell.—The clamshell uses a long light boom and a two-jaw bucket attached to two lines. One line is used to raise and lower the bucket, the other pulls the jaws together with a digging action. They open by gravity when the digging line is released. The bucket is prevented from spinning by a light rope, the tagline, attached to two bucket corners and kept under tension by a spring wound drum on the boom.

The clamshell can dig straight down to any depth its rope length and winch drum capacity will allow, and can also work at its own level and above. Hard digging requires a heavy bucket with a slow, powerful closure. Production may be only half that of a dipper shovel even under good conditions. The clamshell is used chiefly for deep narrow excavations and for stockpiling and rehandling loose material such as sand or gravel, but it can do almost any type of excavation.

Bulldozer.—The bulldozer, often called dozer, is a heavy front pusher blade mounted on a crawler tractor or on a four wheel-drive wheel tractor (*see* fig. 3). The blade hoist is usually hydraulic on small and medium machines and winch and wire rope on big ones. It is a powerful excavator and land-clearing machine, and can pile soil and debris at the edge of the work area or push it up to 200 ft. When the blade is held below surface level and the tractor is moved forward, it digs; if at the surface level it transports; and above the surface it spreads its load. An angle dozer is a bulldozer whose blade can be angled for sidecasting.

The bulldozer working alone can do many types of excavation efficiently and is indispensable as a helper machine on almost every



BY COURTESY OF INTERNATIONAL HARVESTER COMPANY

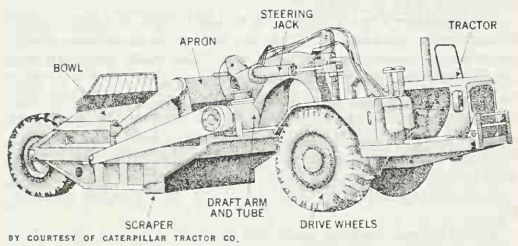
FIG. 3.—BULLDOZER

earth-moving job. Its uses include digging cellars, road cuts and shallow ditches; transporting, spreading and grading soil; removing trees, stumps and boulders; pushing loading scrapers; rescuing stuck equipment and cleaning and leveling around shovels. Power range is from 20 to more than 400 h.p., weight from 2 to 40 tons.

Shovel Dozer.—This machine, also known as a dozer shovel and front loader, was developed after 1944. It replaces the bulldozer blade with a tractor-width bucket that can be lifted hydraulically high enough to dump into trucks. The operator can pivot and hold the bucket in various digging, bulldozing and dumping positions. The bucket is crowded into the digging by driving the tractor forward, retracted by reversing and swung by steering. This machine can do most bulldozer work, and can load trucks both at the pit face and at a distance from it. Bucket capacities range from ½ to 4 cu.yd.; cycle time is about 30 sec. at minimum distance; and production is half that of a dipper shovel of the same bucket size.

Overhead Shovel.—The overhead shovel digs with a front-mounted bucket in the manner of the shovel dozer, but swings the load overhead and dumps it to the rear. Small models powered by compressed-air motors and mounted on narrow-gauge rails, with cycles as fast as 8 sec., are used extensively in tunnel excavation. They may dump directly into trucks or cars, or onto conveyor belts.

Scraper.—There are two important types of carrying scraper. A trailer model has two axles and is towed by a crawler tractor. The self-propelled type has a rear axle only and the front is fastened to a two- or four-wheel rubber-tired tractor by a vertical kingpin (*see* fig. 4). Either scraper body includes a bowl with a cutting edge that can be lowered into the ground for digging or raised for carrying, an apron that holds the load from spilling out the front, a tilting or sliding tailgate for dumping the load and rear tires to carry the body. Control is by means of winch lines or hydraulic rams operated from the tractor. Size range is from 1 to 40 cu.yd.; cycles vary from 3 to 30 min. Length of haul and speed are the most important variables. Hard soil



BY COURTESY OF CATERPILLAR TRACTOR CO.

FIG. 4.—SCRAPER

may have to be loosened by tractor-drawn rippers.

Crawler-drawn scrapers dig powerfully but have a top speed of only about 7 m.p.h. They are used chiefly for hauls of 200 to 500 ft.

The self-propelled machines have to be pushed by another tractor to get a full load, but can then travel at speeds up to 25 m.p.h. on hauls of over a mile.

The scraper is a most important tool in highway construction and in most large-scale short-haul excavation that is fairly free of mud and rock.

Grader.—A grader is a precision finishing machine that carries an 8- to 14-ft. blade between the front and rear wheels, with mechanical or hydraulic controls that permit pivoting to cast on either side, extending it out from either side of the machine and raising or lowering either end independently. Most models do little direct excavation except shallow ditching, but assist other earth-moving equipment by shaping and smoothing fills, cuts and roads. Weight is from 3 to 19 tons; diesel-power range is from 40 to 195 h.p.

Trenching Machine.—The trenching machine or ditcher is self-propelled on crawlers or rubber tires. It carries a boom or ladder on which an endless chain of buckets revolves, or has a wheel fitted with rim buckets. The machine moves forward and digs behind as the buckets dump by gravity on a conveyor belt that stacks the cuttings on one or the other side.

Trenchers may dig over 500 ft. of small ditch an hour under good conditions. They can be equipped to cut hard ground and even soft limestone, but have difficulty with boulders. They are not adaptable to other types of work.

Elevating Grader.—This machine consists of a plow or disk that cuts the soil and a conveyor belt that lifts the cuttings and discharges them into a truck or in a pile. Current models are mounted in heavy grader chassis.

Drilling.—Most rock must be drilled and blasted before digging. Hand and wagon drills use air from a separate compressor to hammer and rotate a steel rod fitted with a cutting bit. A wagon drill has a trailer or power track mounting, a swivel mast to hold a heavy drill and a power-feed mechanism. In hard rock drill bits are tipped with tungsten carbide. These light and medium compressed air drills are used for holes from 2 to 4½ in. in diameter and to depths of 30 ft. or more. Drilling rate varies from 20 to 100 ft. per hour.

Larger and deeper holes have been made chiefly by churn drills that operate by raising and dropping a heavy bit suspended on wire rope. Churns are now being replaced by rotary drills with bits up to 12 in. in diameter that are rotated by a drill rod and that discharge cuttings to the surface by means of an air stream. In very hard rock a thin hammer drill may be lowered into the hole by a slowly rotating rod. Soft rock may be drilled by augers. Hole size and spacing are calculated so that a pound of dynamite or other explosive can be placed to break from one to five yards of rock.

Highways.—Construction of a first-class highway with wide roadbed and low gradients may require moving of large amounts of soil and rock, sometimes over 500,000 cu.yd. in a mile. In the early estimates the gradients of a road are so calculated that all materials dug may be used in fills across nearby lower areas in one economic operation.

In cuts through hills the ground is dug in horizontal layers that may be 4 to 12 in. deep in scraper work, or 12 to 15 ft. for 2½ cu.yd. shovels. Rock is drilled and blasted before digging. If the route is along a hillside the upper line of the cut is started by sidecasting with a bulldozer to provide a working shelf. A bulldozer may complete a narrow roadbed in this manner.

In fills, soil is placed in layers 5 to 24 in. deep. Each is thoroughly compacted by rollers before the next layer is spread. Depth of layers of blasted rock is governed by the size of the pieces to be used. Both cuts and fills are started full width, and side slopes are finish-graded as the work progresses.

Permissible slopes of soil cuts may be as steep as 1 on 1 (100% grade or 45°) and of fills 1 on 1½, but they vary down to 1 on 6, depending on the stability of the material. Firm rock may be left vertical or even overhanging.

Compaction.—When soil is loosened by digging it increases in bulk 10% to 40%. It may take years for it to settle to its original volume. Such settlement will damage any surfacing or structure placed on loose fill. Damage is avoided by compacting soil as it is placed, partly by hauling equipment, but chiefly by rollers. These may be the tamping (sheepsfoot) type, that is, a weighted drum with projecting lugs that pack soil below the surface, or weight boxes of up to 60 tons carried on pneumatic tires, which compact by direct weight and by a kneading action. Effectiveness may be increased by a vibrating unit mounted on the axle. Good compaction reduces many soils to 90% of their original volume.

Strip Mines.—Mineral deposits near the ground surface may be mined in open cuts by first removing the overburden. If the deposit must be worked over a large area at one time, as in stone quarries and some extensive copper and iron ore formations, the overburden is dug and hauled by highway cut methods. If the pay layer is shallow and can be taken in narrow strips (coal, for example) the overburden is sidecast, the exposed deposit is dug, and then another strip is uncovered by sidecasting its waste into the mined area.

Such sidecasting may involve removing as much as 100 ft. of overburden in a single layer, so as to require the use of the largest shovels and draglines. Deep cuts usually include rock that must be drilled and blasted. State laws may require releveling the area when mining is completed.

BIBLIOGRAPHY.—Herbert L. Nichols, Jr., *Moving the Earth* (1955), *Machinery Around Your Country Home* (1957); R. L. Peurifoy, *Construction Planning, Equipment, and Methods* (1956). (H. L. N.)

EXCELLENCY, a title or predicate of honour. The earliest records of its use are associated with the Frankish and Lombard kings; *e.g.*, Anastasius Bibliothecarius (9th century A.D.) in his life of Pope Honorius refers to Charlemagne as "his excellency" (*ejus excellentia*); and during the middle ages it was freely applied to or assumed by sovereigns generally, though rather as a rhetorical flourish than as a part of their formal style. In the various charters in the *Red Book of the Exchequer*, for instance, the addresses to the king vary between "your excellency," "your dignity" (*vestra dignitas*), "your sublimity" (*vestra sublimitas*) and the like, according to the writer's taste. C. du Fresne du Cange's *Glossarium* gives examples also of the style's being applied to the pope or even to a bishop.

With the gradual stereotyping of titles of honour, that of "excellency" became obsolete for sovereigns of the highest rank about the beginning of the 15th century. Dukes and counts of the Holy Roman empire and the Italian reigning princes continued, however, to be "excellencies" for a while longer. In 1593 the bestowal of the title of *excellence* by Henry IV of France on his ambassador at Rome, Louis de Gonzague, duc de Nevers (who, as a younger son of the Gonzaga duke of Mantua, was a member of a sovereign family) set a precedent that was universally followed for ambassadors from the time of the treaty of Westphalia (1648). This led the Italian princes to adopt the style of "highness" (*altezza*) instead of "excellency."

In France from 1654 onward the title of *excellence* was given to all high officials. In the 18th century this example was followed in Germany and in Austria, where ministers, civil servants with the rank of *Geheimrat* (privy councillor), generals and admirals used to be addressed as *Exzellenz;* but this ceased officially after the revolutions of 1918. In Spain, ministers, bishops, high-ranking civil servants, captains-general of the armed forces, *grandees* and members of certain ancient orders of chivalry bore the title of *excelencia.* In Russia, Peter the Great introduced a "table of ranks" dividing the state officials in 14 classes: the dignitaries of the first three ranks were given the title of *prevoskhoditelstvo* (excellency). In Poland the style is given only to bishops (*ekscelenja*).

The only uniformity about the title in modern times is in the diplomatic service, where it is reserved for ambassadors (and their wives), though even here envoys, especially in the Latin-American republics, are often accorded the courtesy style of *vuestra excelencia* or *excelentisimo señor.* In 1948 the title was extended to commonwealth high commissioners.

Some republican heads of state, but not the president of the United States, are given "excellency," as also are cabinet ministers in some countries, but not in the U.K., the Commonwealth or the U.S. British colonial governors enjoy the title, as did viceroys of India and most Spanish colonial viceroys.

EXCEPTIONAL CHILDREN, those children who, because of abnormally high or low mental abilities or because of unusual talents or physical handicaps, should have specialized schooling in addition to or in place of that given to normal children. *See* PHYSICALLY HANDICAPPED, EDUCATION OF; RETARDED CHILDREN; SUPERNORMAL CHILDREN, EDUCATION OF; BLIND, TRAINING AND WELFARE OF; DEAF AND HARD OF HEARING, TRAINING AND WELFARE OF; GIFTED CHILDREN; REHABILITATION, MEDICAL AND VOCATIONAL.

EXCESS-PROFITS TAX, a tax levied on profits in excess of some standard of "normal" income. There are two theoretical principles governing the determination of excess profits. One, known as the war-profits principle, is designed to recapture wartime increases in income over normal peacetime profits of the taxpayer. The other, identified as the high-profits principle, is based on income in excess of some statutory rate of return on invested capital.

The modern excess-profits tax originated during World War I as a revenue measure and instrument for curbing excessive profits attributable to the war. Denmark and Great Britain were the first countries to impose wartime excess-profits taxes, in May 1915. By 1919, when India levied its tax, a score of countries had enacted levies based on either the war-profits or high-profits principle. These were repealed after the war but were restored in one form or another during World War II. Several countries revived their World War II levies during the Korean War.

Excess-profits taxes also have become part of the peacetime tax structure of Denmark and several South American countries. These levies, all based on the high-profits principle, are illustrated by the Argentinian tax, which ranges up to 30% on the amount by which 85% of profits exceeds 12% of capital and surplus. In 1950 Japan abandoned its peacetime high-profits tax, which progressed from 10% on profits of 30%–50% of capital to 20% on profits over 100% of capital.

The income subject to tax is the amount which exceeds an "excess-profits credit." A credit based on the war-profits principle is usually the average income during a base period immediately prior to the wartime increases in profits. Alternative credits based on a percentage of invested capital, a minimum amount of income or other standard may be provided as relief from hardship. A credit based on the high-profits principle is a statutory rate of return on invested capital which is usually graduated to favour small businesses. The invested capital may be limited to the amount by which the book value of assets exceeds liabilities, or it may include all or part of borrowed capital.

Tax rates may range up to 100% of the excess profits. Such high rates, however, may be reduced by postwar refunds or may be limited by a ceiling rate based on total net income. High-profits tax rates are frequently graduated so that they rise as the rate of return on investment rises. While excess-profits taxes may be levied on all forms of business, whether or not incorporated, they have increasingly been limited to corporations.

World War I.—The British Munitions levy of May 1915, a 100% tax on profits in the munitions industry in excess of 120% of prewar profits, was superseded in October by the Excess Profits duty. This was essentially a levy on profits of all businesses exceeding the average two out of three prewar years, or a minimum credit based on 6% to 7% of invested capital. Originally imposed at a rate of 50%, the tax rate was increased to 60% in 1916 and 80% in 1917; it was reduced to 40% in 1919 and increased to 60% thereafter until its repeal in 1921. The Excess Profits duty was enormously productive, yielding about one-fourth of total British tax revenues during the period of its enactment.

In 1915 Denmark, Sweden and Norway levied taxes at graduated rates on increases in profits of individuals and corporations due to the war. Average prewar income was taken as the standard, but invested capital credits were later introduced. Tax rates

were moderate. Other European countries followed with excess-profits taxes.

While British dominions generally followed the British model, neither Canada nor the United States adopted the war-profits principle during World War I. The Business Profits tax adopted by Canada in 1916 represented the first experience of the dominion government with an income tax on corporations. This was a levy of 25% on corporation profits in excess of 7% of capital and on unincorporated business profits over 10% of capital. In 1917 tax rates were made progressive up to a peak rate of 75% of profits in excess of 20% of capital, where they remained until the end of 1919. Rates were reduced until the end of 1920, when the tax was repealed. During the period 1915–20 the Business Profits tax accounted for about 10% of dominion tax revenues.

The first excess-profits tax of the United States was enacted in March 1917, but this was replaced by the greatly stiffened excess-profits tax of Oct. 1917. This tax was imposed at graduated rates on profits of all businesses in excess of prewar earnings but not less than 7% or more than 9% of invested capital. The choice of a high-profits tax was dictated largely by its potential role as a permanent levy, as well as by the precedent of Canada. The Revenue act of 1918 superimposed on this, for the year 1918 only, an alternative war-profits levy ranging up to 80%, with the taxpayer paying whichever was higher. The excess-profits tax was repealed in 1921, despite strong efforts to retain it permanently. It is estimated to have raised as much as one-fourth of total federal revenues during its enactment.

World War II.—With rearmament and the outbreak of World War II, excess-profits taxes were reintroduced by most countries to raise revenue and to prevent wartime enrichment. The excess-profits tax of Great Britain, enacted in Sept. 1939, superseded the Armaments Profit duty of that summer, which was limited to profits arising directly from rearmament. The new tax was in direct line of descent from the World War I duty, with an initial rate of 60% on profits of all businesses in excess of the prewar standard, in this case 1935 or 1936 or the average of the two highest of the three years 1935–37. An alternative invested capital standard was available at 8% of capital for corporations and 10% for individuals and corporations. Initially 60%, the tax rate was raised to 100% in May 1940. Because of the possible adverse effects of this rate on business initiative, a postwar rebate of 20% was later provided. Since the rebate itself was subject to a tax of about 50% when paid, the net excess-profits tax rate was about 90%. During the last year of the tax, 1946, the rate was reduced to 60%. Over the war period the excess-profits tax accounted for about one-eighth of total wartime tax revenues.

In adopting its World War II excess-profits tax, Canada departed fundamentally from the high-profits principle followed during World War I. The act of 1939 never went into effect and was replaced by the Excess Profits Tax of 1940. This act imposed a levy of 75% on incomes in excess of average earnings for the years 1936–39 or 12% of total profits, whichever was greater. Where base-period earnings were less than 5% of capital employed, a board of referees was authorized to increase the standard up to 10% of capital. The tax applied to nonincorporated businesses as well as to corporations.

Between July 1, 1942, and Dec. 31, 1945, rates were raised to 22% of net income or 100% of excess profits, whichever was higher. Taxpayers subject to the 100% rate were eligible for a postwar refund of 20%. The tax remained in effect at reduced rates after 1945 and expired with the end of 1947. Total excess-profits tax collections through the fiscal year 1949, after refunds, amounted to about 12% of total dominion tax revenues for the period. Corporations accounted for 95% of the revenues.

In the United States the choice between a high-profits and war-profits tax was the principal issue in the enactment of an excess-profits tax by the Second Revenue act of 1940. The treasury advocated a high-profits tax based on the World War I prototype, but congress adopted a credit based on 95% of average earnings for the years 1936–39. An alternative invested capital credit was provided taxpayers with a low prewar earnings base, initially 8% of capital and later graduated from 5% to 8%. A "growth formula" permitted growing companies to use high earnings at the end of the base period, and special relief administered by an Excess Profits Tax council was provided corporations with base period abnormalities. Only corporations were subject to the tax.

Tax rates ranged from 25% to 50% in 1940 and from 35% to 60% in 1941; in 1942 a flat rate of 90% was adopted with a post-war refund of 10%, and in 1944 the rate was raised to 95% with a 10% refund. The tax was repealed by the Revenue act of 1945, effective Jan. 1, 1946. The yield of the World War II excess-profits tax was estimated at about $16,200,000,000 after allowance for postwar refunds of $2,200,000,000 and the corporation income tax of $16,700,000,000 it displaced. This represented roughly 10% of total federal tax revenues during this period.

Australia, moving against the universal trend toward a war-profits tax and despite its own World War I levy, adopted a high-profits tax in 1940. Other countries generally followed their World War I patterns.

The Korean War.—The Korean rearmament period, 1950–53, witnessed the reintroduction of excess-profits taxes only by the United States (1950–53), Great Britain (1952–53), Belgium (1951) and Sweden (1951).

In the United States, in accordance with the president's promise of a tax program that would "have as its major aim the elimination of profiteering," the Excess Profits Tax act of 1950 was signed Jan. 3, 1951. It was made effective for three years beginning July 1, 1950, but was later extended to Dec. 31, 1953, because of the continued need for revenues. This measure was patterned on the World War II tax, with a standard based on 85% (later 83%) of average earnings during the three highest of the four years 1946–49. An alternative invested capital credit ranged from 12% of the first $5,000,000 to 8% on capital over $10,000,000. It also provided a minimum credit of $25,000 which excluded about 95% of all corporations. Special provisions were made for new and growing businesses, and so-called automatic relief was provided for companies with abnormalities during the base period. Additional credit also was given for increases in capital during the taxable years, at a rate of 12%.

The tax was assessed at 30% of excess profits but was limited to 18% of total net income. Including the ordinary income tax rate of 52%, the combined marginal rate was 82%, but the overall effective tax rate was limited to 70% of net income. The excess-profits tax yielded only about $7,500,000,000 and accounted for 10% of all corporation taxes. In the peak year, 1951, only 41,000 corporations were subject to the tax. About 4,600 of these, with incomes over $1,000,000, accounted for 80% of the total excess-profits tax liabilities.

Following the rise in Korean armament expenditures, the British Finance act of 1952 proposed re-enactment of an excess-profits levy to capture the "fortuitous profits arising from rearmaments," to be effective Jan. 1, 1952. This levy was enacted at a rate of 30% of profits in excess of the average in two out of the three years 1947–49, but not to exceed 15% of total taxable income. An alternative invested-capital standard was adopted equal to 8% of capital employed or 10% of paid-up capital. Unlike the previous taxes, the Excess Profits levy did not apply to unincorporated businesses. The low apparent rate of 30% is explainable by the fact that the levy was not deductible for income-tax purposes, whereas the wartime excess-profits tax was deductible. In this and other respects its structure was similar to that of the United States Korean excess-profits tax. With its repeal by the Finance act of 1953 the Excess Profits levy remained in effect for only two years, until Dec. 31, 1953.

Economic Effects.—Although excess-profits taxes serve a moral purpose by "taking the profits out of war" and achieving an equitable distribution of war costs, they are opposed by business as economically unsound and administratively impractical. The determination of what is "excess" bristles with conceptual difficulties and is bound to be arbitrary in practice. It is charged that high tax rates impair economic incentives, discourage production and undermine controls over business expenditures. They may encourage wasteful business expenditures for excessive salaries and bonuses, advertising, entertainment and other expenses

and thereby contribute to the inflationary forces of wartime.

On the other hand, if the profits taxed are really wartime "windfalls," their recapture by the government does not impinge upon essential business incentives. The extraordinary record of production which has characterized war emergencies tends to refute the contention that excess-profits taxes destroy incentive, and studies have shown that such taxes have had little effect on business investment decisions. The temporary character of the tax and wartime patriotic motives may make tolerable what could not be endured over an extended period.

Because of this temporary character, however, businesses may be encouraged to increase certain optional business expenditures where the benefits of their deduction at high rates can be realized in future profits taxable at lower rates. Such a tax theory of business expenditures has not been adequately tested by the facts, and in general there appears to be no appreciable correlation between high tax rates and expenditures except possibly for contributions. Neither has careful audit of U.S. tax returns shown that abuses with respect to deductions for entertainment and similar expenses are more associated with the excess-profits tax than with the income tax.

Broadly speaking, an excess-profits tax during a war period probably has a greater stabilizing effect on the economy than an alternative increase in corporation income tax. This is because limitation of profits makes it easier to increase individual income taxes, freeze wages and salaries and restrict consumption by rationing. An excess-profits tax also will absorb a larger share of inflationary profits than will a flat income tax. In addition, it facilitates price controls and greatly eases the task of renegotiating war contracts.

Legislative attempts to recognize differences in industry circumstances temper the harshness of an excess-profits tax but contribute greatly to its complexity. Special credits are usually provided for growing industries and new businesses, and special relief is given for depressed industries and those with base-period abnormalities. Other concessions or exemptions are also usually provided to remove tax deterrents to production, as in the case of extractive industries. Considerable administrative discretion is usually granted in determining relief, but it is occasionally limited to automatic formulas. The result is probably the most complex type of tax legislation on record, with a relatively high cost of administration and compliance.

High-profits taxes tend to cast different types of businesses in the same mold, regardless of risk, variability of profits and the importance of capital. Such a principle may be justified during national emergencies by wartime stabilization requirements. A high-profits levy can be more productive of revenue than a war-profits tax and may be found essential to wage controls and high personal income taxes. As a permanent peacetime measure, however, a high-profits tax may distort the allocation of resources into less productive channels and interfere with sound economic development. Rates of peacetime high-profits taxes are therefore usually moderate.

See also references under "Excess-Profits Tax" in the Index volume.

BIBLIOGRAPHY.—Carl S. Shoup, "The Taxation of Excess Profits," *Political Science Quarterly* (Dec. 1940; March, June 1941); J. R. Hicks, U. K. Hicks and L. Rostás, *The Taxation of War Wealth* (1941); K. J. Curran, *Excess Profits Taxation* (1943); M. H. Gopal, *The Theory of Excess Profits Taxation* (1947); "Symposium on the Excess Profits Tax," *National Tax Journal* (Sept. 1951); George E. Lent, "Excess Profits Taxation in the United States," *Journal of Political Economy* (Dec. 1951), "The Excess Profits Tax and Business Expenditures," *National Tax Journal* (Sept. 1958); W. Miller, *Excess Profits Tax, Law and Practice*, 3rd ed. (1944), and *Excess Profits Levy: Law and Practice* (1952); Tax Institute Symposium, *Excess Profits Taxation* (1953); R. S. Sayers, *Financial Policy 1939–1945* (1956); Robert M. Haig, "The Taxation of Excess Profits in Great Britain," *American Economic Review*, supplement (1920). (G. E. Lt.)

EXCHANGE, FOREIGN. The term foreign exchange denotes foreign money or any liquid claim on foreign money, such as a bill of exchange, that is offered for sale on the foreign exchange market where it may be bought with domestic money. Foreign exchange is bought and sold because payments from one country to another require the exchange of domestic currency for foreign currency. The price in terms of domestic money that is paid for another country's currency unit (*e.g.*, the dollar price for one pound sterling) is called the rate of exchange. *See* INTERNATIONAL MONETARY FUND; INTERNATIONAL PAYMENTS; *see* also references under "Exchange, Foreign" in the Index volume.

EXCHANGE CONTROL. This term refers to various forms of governmental restrictions upon private transactions in foreign money or claims on foreign money (commonly called foreign exchange). Some exchange controls have sought merely to relieve mild pressure on the balance of payments caused by capital export and have involved nothing more than the careful scrutiny of all applications for the purchase of foreign exchange and the restriction of allotments of exchange for capital export. The enforcement of such mild controls may be entrusted to established foreign exchange dealers and not require the setting up of any official control agency.

Most systems of exchange control, however, are more comprehensive and rigorous. Their chief function is to prevent or redress an adverse balance of payments by limiting purchases of foreign exchange to an amount not in excess of foreign exchange receipts. An exchange control authority is designated, usually the central bank or a specialized government agency, and residents are required to sell to this authority, at rates set by the authority, all foreign exchange that comes into their possession, although some systems permit recipients of exchange from certain sources to sell a portion of these receipts in a free market. The control authority thus becomes the only market in which foreign exchange may be sold and the only source from which residents may acquire foreign exchange for making payments abroad. Since it is a monopoly, the control authority can determine the purposes for which foreign exchange may be spent, fix the amount available for each purpose and set the rates at which exchange shall be bought and sold.

The exchange rate maintained by the control is usually higher than the rate that would prevail in a free market. Such an artificially high rate has the effect of curbing exports and of stimulating imports. By limiting its allotments of foreign exchange, however, the control authority can control imports and thus prevent a decline in its total gold reserves and foreign balances, as well as in its reserves of any particular foreign currency. *See* INTERNATIONAL PAYMENTS. (L. W. Te.)

EXCHANGE MARRIAGE, a form of marriage involving an arranged and reciprocal exchange of spouses between two groups. In general, a marriage establishes an affinal alliance between the immediate kin of the husband and those of the wife. The political and economic consequences of such alliance can be great, and it is not surprising that, in human society as a whole, marriages are more frequently negotiated by the senior kin of the contracting parties than by the spouses themselves. The advantages to the two sets of kin are not identical. As a rule, only one of the spouses changes residence on marriage; consequently, one group of kinsmen gains an adult member at the expense of the other. Moreover, when the children are born, they are commonly recruited to one side only, that through which descent is traced.

In societies which lay stress on the alliance aspect of marriage and which associate a doctrine of unilineal descent with a consistent rule of postmarital residence—*e.g.*, patrilineal descent with patrilocal residence—the symmetry of the alliance is often maintained by a systematic exchange: whenever group A contracts to give a bride to group B, group B will simultaneously contract to give a bride to group A. For example, in most tribes of Australian aborigines neighbouring communities (hordes) systematically exchange women, the ideal model of any marriage contract being that two men from different hordes should each marry the sister of the other. If a brother and a sister in group A marry a sister and a brother in group B and their children repeat this process, then the second set of marriages will be between "cross-cousins." Cousin marriage (*q.v.*) is thus often a form of exchange marriage. The logical difficulty that brothers and sisters are not ordinarily born in exactly equal numbers is disposed of by the use of classificatory kinship categories (*see* KINSHIP TERMINOLOGY). Institutions by which a man gives his sister in marriage in exchange for valuables

which he himself then uses to secure a wife are widespread and constitute a species of exchange marriage.

BIBLIOGRAPHY.—Sir J. G. Frazer, *Folklore in the Old Testament*, vol. ii, pp. 195–220 (1918); A. R. Radcliffe-Brown, *The Social Organisation of the Australian Tribes* (1931); C. Lévi-Strauss, *Les structures élémentaires de la parenté* (1949); E. R. Leach, "The Structural Implications of Matrilateral Cross-Cousin Marriage," *J. R. Anthrop. Inst.*, vol. 81 (1951); R. Needham, *Structure and Sentiment* (1962).

(E. R. L.)

EXCHEQUER, the English name for an accounting department, used of private as well as of public institutions, but more commonly of the government office receiving and having the care of the public revenues (Lat. *scaccarium*, "a chessboard" or, derivatively, a counting table and, from that, an accounting department).

As the name of a government department, it is not known whether exchequer is older than the 12th century. The treasury, with which it was in practice joined, dates from before the Conquest and those officers of the exchequer who were drawn from the treasury staff can be traced back to Domesday. The word came quite early to be applied to the two jointly. The "lower" exchequer, or "receipt," closely connected with the permanent treasury, was an office for the receipt and payment of money. The "upper" exchequer (the *scaccarium* proper) was a court sitting twice a year to regulate accounts. It was closely related to the Curia Regis, and is thus probably on the Norman pattern. The lower exchequer was conditioned in its development by the use of tallies (*see* TALLY); the upper, as its name implies, was based on the use of the abacus, or counting board, for reckoning, such a contrivance being almost universal before the introduction of Arabic numerals, which did not become common in England till the 16th century. The earliest information about the exchequer itself, apart from that afforded by the pipe rolls, rests on a treatise, *Dialogus de Scaccario*, written in the latter part of the reign of Henry II by Richard, bishop of London and treasurer of England. His father, Nigel, bishop of Ely, had been treasurer to Henry I and nephew of that king's great financial minister, Roger, bishop of Salisbury. Nigel is said to have reconstituted the exchequer after the troubles of Stephen's reign upon a model inherited from his uncle.

The Angevin Exchequer.—The sessions seem at first to have been confined to the actual times when the court sat at the "chessboard," around which were grouped the barons, officials and accountants. At the head of this table were the great officers of state except the treasurer, who sat on the right with the exchequer clerks below him: on the left was the calculator, and at the bottom appeared the officials whose accounts were to be audited. Of the officers of state, the justiciar was normally the president, in the king's absence; the chancellor was responsible, as keeper of the seal, for all writs and summonses; the two chamberlains, lay colleagues of the treasurer, shared with him the duty of receiving money and keeping the records and other contents of the treasury; the marshal, a subordinate of the constable, was responsible for the custody of prisoners and of the accountants' vouchers. The treasurer charged accountants with their debts and dictated the great roll of accounts (pipe roll), which was written by his scribe, the chancellor's scribe copying from this a duplicate (the chancellor's roll). The calculator worked out the sums on the exchequer table (abacus) by means of counters. Other subordinates were the *magister scriptorii* ("master of the writing office") and the chancellor's clerk (later chancellor of the exchequer) who settled the form of all writs and summonses, charged the sheriff with all fines and amercements and acted as a check on the treasurer in the composition of the pipe roll.

In the lower exchequer (the receipt), the staff was divided between the treasurer and chamberlains, the clerical members being under the former, the lay under the latter. The money brought in by the royal accountants was counted by the four tellers, weighed by the "pesour" (a knight) and assayed by the "melter," receipts being recorded by means of wooden tallies cut and split by the tally cutter, one half going to the payer while the other was retained at the exchequer to prevent fraud. Record of moneys issued was mainly secured by filing the warrants for payment (writs under the great seal) as they were honoured. Early in Henry II's reign written records of the tallies (rudimentary receipt rolls) began to be kept in addition to the originals; issue rolls were also used later for money paid out, but the old tally and warrant system continued throughout the history of the department. At the receipt the treasurer was represented by a clerk (later clerk of the pells), and each of the chamberlains by a knight (later the "chamberlains of the receipt"); these three had joint control of the treasury. In both upper and lower exchequers were ushers.

The business of the ancient exchequer was mainly financial, though some judicial business connected with the accounts was also done. The principal accountants were the sheriffs, who were bound, as the king's chief financial agents in each county, to give an account of their stewardship twice a year, in a preliminary "view" at Easter, when they "proferred" their first payments for six months, and in the effective account and audit on the pipe roll at Michaelmas. The fixed revenue consisted of the farms of the king's demesne lands within the counties and of certain boroughs which paid annual sums as the price of their liberties. Danegeld, though regarded as fixed revenue, was infrequently levied after Henry II's accession, but there were rents of assarts and purprestures, mining and other royalties. The casual revenue consisted of the feudal incidents (escheats, wardships and marriages), of the profits of justice (amercements, and goods of felons and outlaws) and of fines or payments made by the king's subjects to secure grants of land, wardships and marriages and of immunities, and to hasten (or sometimes to delay) justice. Besides these there were the revenues arising from aids and scutages of the king's military tenants, tallages of the crown lands, customs of ports and special "gifts," or general assessments made on particular occasions. For the collection of all these the sheriff was primarily responsible, though in some cases debtors and accountants dealt directly with the exchequer and were bound to make their appearance in person on the day when the sheriff accounted.

The farms of demesne lands were originally for the most part paid in kind but the evidence of Domesday Book shows that payments were then normally "by tale," occasionally by weight and more rarely in assayed coin.

The rapid deterioration of a small silver coinage gave rise to attempts at maintaining the value of cash payments, first by a scale reduction for wear, then by substitution of payment by weight for payment "by tale" and finally by the reduction of payments of most farms of the shire to their pure silver value by means of an assay. The casual revenue was still paid by tale, and for the determination of its amount it was necessary for the exchequer to receive copies of all grants made in the chancery on which rents were reserved or fines payable: these were known first as *contrabrevia* and later as *originalia*. The profits of justice were notified by the delivery of similar estreats from the justices' records from an early date; and, while for certain minor casualties the oath of the sheriff was at first perhaps the only security, later they were frequently determined by means of inquisitions. Exchequer business might be transacted anywhere in England; the original meetings were at Winchester, later they took place at Westminster as a rule but occasionally, even in the 14th century, elsewhere.

Developments in the 13th and 14th Centuries.—The Angevin "course" remained the ideal of the exchequer for six centuries: and, in spite of constant adaptation, its broad features were never completely effaced. Great officers of state ceased to attend personally. Even under Henry II it seems that all, except the treasurer and chancellor, had deputies; and under Henry III the chancellor was represented by his clerk, who became known as chancellor of the exchequer. On the other hand new officials emerged, notably the four permanent official barons and the king's and lord treasurer's remembrancers. Two memoranda rolls of these remembrancers survive for John's reign and show that the exchequer already sat regularly except during vacations, that memoranda were made of business reserved for attention at subsequent sittings and that this business had begun to be arranged systematically. During the 13th century the functions of the

remembrancers were differentiated, their roll developed to form a record of a large part of the exchequer's business, and by the ordinance of 1323–26 their precise duties were defined. Henceforward the king's remembrancer was more particularly concerned with the casual, the treasurer's with the fixed revenue. Hence in the preliminary stages of each account, including the registration of the king's writs, the king's remembrancer was responsible; he retained the corresponding vouchers, a collection which is still one of exceptional size, diversity and interest. The treasurer's remembrancer exacted the remanets of such accounts as had been enrolled, as well as reserved rents and fixed revenue, and so became closely connected with the clerk of the pipe. Before the end of the century these three offices had separated into different departments.

Judicial Procedure at the Exchequer.—One development at the exchequer during the 13th century was only indirectly concerned with accounting. By about 1250 convenience had led to the establishment of something like separate machinery, and definitely separate records for legal cases which arose out of accounts and which, in the earliest times, we find among the ordinary memoranda.

The affairs of the Jews, until their expulsion in 1290, had, since Richard I's reign, been subject to a mixed arrangement. Receipts from Jewish sources passed through the lower exchequer, while a special *Scaccarium Judaeorum*, whose records are known to us as the Jewish plea rolls, dealt with most of the financial business, other than that of actual receipt. At the same time it acted as a court for the trial of all cases, both crown and civil, in which Jews or Jewish business were concerned. The point of contact with the exchequer was that, all Jews and their property being at the absolute disposal of the crown, no case arising from their business was without a potential interest for the royal financial administration. Similar needs gave rise to an organized department of pleas which, while it had nothing to do with normal financial business and with normal pleas of the crown, took cognizance of the increasing bulk of disputes arising out of the conduct of the royal accounting together with all cases affecting exchequer officials. In course of time the exchequer of pleas, though it continued to be staffed by exchequer officials, became a court of civil pleas like any other and was ended only by the Judicature act of 1873.

Other Modifications in Exchequer Machinery.—During the 13th century the increasing length and variety of accounts led to various other efforts at reforms besides those already mentioned. Even before Henry II's death, two problems had become apparent: the annual repetition on the great roll of "desperate" or bad debts, and the increasing number of amercements in the royal courts. The 13th century saw a strenuous effort to meet the new conditions: its stages are marked by the reforms of 1236–42 (no statute), and the statutes of 1270 and of Rhuddlan in 1284; its termination by the codification of exchequer practice (already well established in many respects) in the Stapeldon ordinances, 1323–26. By 1242 the farms of the counties and boroughs had been revalued and the demesne lands removed from them and let to farm for a term of years; the estreats were no longer entered on the great roll in detail, the sheriff being charged on each estreat roll and only actual payments recorded on the great roll itself; while a new form of tally allowed the sheriff to pay large totals into the exchequer in his own name instead of having to get a tally for every small debt. This simplified work at the exchequer enormously: the machinery had stretched. The statute of 1270 codified the new procedure with regard to estreats; that of 1284 dealt with farms of counties, the new form of tally and a special "exannual roll" in which desperate debts were recorded.

But these were not the only changes. One development which began early was a practice of anticipating revenue, by which sheriffs were ordered to pay out locally the greater part of their receipts; the assignment of tallies for moneys not yet paid in was a feature of exchequer procedure throughout its subsequent history. Again, under Henry III and his son a large accession of financial business arose from the "foreign accounts," that is to say, the accounts of services which, naturally, were not included among those of the counties and which under Henry II had not formed part of exchequer business. In the course of the 13th century the exchequer gradually acquired partial control of these "foreign" accounts. Finally we have to record the appearance before 1242 of escheators, and from then onward the accounts of these and of the officials of the customs, subsidies and wardrobe also occupy an increasingly important position in the national revenue. During the reign of Edward I. the wardrobe account became unmanageable, since the wardrobe both financed the household, army, navy and diplomatic service, and received money direct from accountants as well as through the exchequer. (*See* WARDROBES, THE.)

Enough has been said to show that the 13th century was a period of rapid growth in many directions, and the reform of 1323–26, referred to above, codified and rounded off those already in existence. It greatly increased the number of foreign accounts by making the great wardrobe (the storing department), the butler, purveyors, keepers of horses and clerk of the hanaper of the chancery (who took the fees for the great seal) and the various ambassadors directly accountable to the exchequer. The whole business of foreign accounts was transferred to a separate building where one baron and certain auditors spent their whole time in settling the balances due on the accounts already mentioned, as well as those of royal lands not let to farm, of Wales, Gascony and Ireland, of aids (clerical and lay), of temporalities of vacant bishoprics, abbeys, priories and dignities, of silver and tin mines, of alnage, etc. The balances on these were accounted for in the exchequer itself, and arrears still entered on the pipe roll, but the preliminary accounts were filed by the king's remembrancer, and the actual accounts enrolled separately by the treasurer's remembrancer on the "roll of foreign accounts," a supplement to the pipe, or great, roll. At the same time the sheriffs' accounts were expedited by further simplification of the great roll and by the appointment of a special apposer to take the account of the "green wax," or estreats.

Reform in the rest of the 14th century apparently worked along the lines laid down in 1323–26, with the addition of some attempts about 1360 to draw up something like a balance sheet of revenue and expenditure—a notion foreign, in general, to medieval accounting methods, though an estimate of revenues had occasionally been attempted; *e.g.*, in 1284.

End of the Medieval Period.—The next important change came at the end of the 15th century, following a period during which assignment and other expedients for patching up the royal finances ran wild and the issue roll ceased to exist. It was inaugurated by Henry VII, who introduced summary methods of audit in certain connections, while carefully and personally supervising receipt and expenditure. In the next reign the dissolution of the monasteries led to the setting up of independent financial bodies—the courts of augmentations and of first fruits—with wide powers and with new methods: and although these, in 1554, became departments of the exchequer, a new act set up in 1560, in addition to the exchequer auditors, two new and independent "auditors of the prests" (predecessors of the commissioners for auditing public accounts, who appear in 1785), to take certain specified accounts, such as those of clerks of the works, the treasurer of ships and the master of the ordnance, the clerks of the hanaper and of the great wardrobe and (later) the customs. These new authorities followed improved methods of auditing, though it was only by degrees that the various stages of declaring an account were worked out and the final stage of development was not reached until the treasurership of Juxon (1636–41). By the end of the 17th century a system had been evolved of declaring at the treasury accounts presented in triplicate, one (on paper) being returned to the office of the auditors, while of the other two (on parchment) one went to the accountant and one was passed for enrollment through the offices of the two remembrancers to that of the clerk of the pipe, in order to secure the levying of any remanets or "supers" by process of the exchequer.

Later Developments.—It will be observed that by this device the connection with the old exchequer machinery was maintained.

though in effect the control over public accounting, asserted with so much trouble in the 13th and 14th centuries, passed from it to the audit office. Similarly when, in the 17th century, the other great step toward modern methods was accomplished by the putting of the office of treasurer into commission, the old machinery of the lower exchequer, with its tallies, receipt rolls and a multitude of added apparatus, continued to function under the direction of the new body. This survival of the old machinery continued for 150 years and there were even additions of new administrations to which the name of exchequer attached; there was, for instance, an exchequer court of equity, established in the reign of Elizabeth I, whose registrar was the king's remembrancer; and in the same reign there was added to the exchequer's departments one for the audit of land revenue. It was not until the 19th century that a series of acts swept away one by one all these departments leaving only the name of the exchequer and those of one or two of its officials as relics of the past.

Of these acts the most important from our point of view was that of 3 and 4 William IV cap. 99; this by removing the sheriffs' accounts from its competence ended the series of pipe rolls and in effect destroyed the last vestige of that *scaccarium* with which the tale began. (H. Je.; M. H. M.; F. M. P.)

FROM THE ACT OF 1834

The Exchequer act, 1834, swept away the cumbrous medieval procedure and the expensive sinecures that had survived at the exchequer till then. A simplified system of accounting was adopted and the great triple-locked chests in which the public money had been put away became unnecessary when accounts were opened at the Bank of England and Bank of Ireland for the consolidated fund. For the work which remained to be done a single comptroller of the exchequer with a small clerical staff sufficed. The comptroller's duty was to see that sums were issued from the consolidated fund only to the amounts and for the purposes authorized by parliament.

For a long time most of the issues from the exchequer had been made by way of advances to the paymasters of the departments concerned (the paymaster general of the forces, the treasurer of the navy and others). In 1836 the paymaster general of the forces, the treasurer of the navy and the treasurer of the ordnance were amalgamated into one post. In 1848 this post was combined with the paymaster of civil services into a single paymaster general for all voted services. Since that date all departments have drawn for their expenditure upon this single paymaster general who thus acts, as it were, as their banker.

Exchequer and Audit Departments Act, 1866.—The business of authorizing the issues from the exchequer in conformity with parliamentary authority, although constitutionally important, did not make enough work to justify a separate department. In 1866, under the Exchequer and Audit Departments act, the exchequer work was amalgamated with that of the audit department, thus giving effect to the recommendations of the Select Committee on Public Moneys of 1856–57. This was a suitable arrangement since constitutionally both the comptroller of the exchequer and the head of the audit office required the same kind of independence.

At the head of the new exchequer and audit department was placed a comptroller general of the exchequer and auditor general of the public accounts, usually known as the comptroller and auditor general. He is appointed by the government but he is not a servant of the government. He is a statutory officer with responsibility primarily to parliament, but also, in respect of some of his functions, to the government. In order to secure his independence his salary is borne on the consolidated fund mentioned below.

The Consolidated Fund.—All revenue, proceeds of loans and other public receipts (with certain restricted exceptions) have to be paid to the consolidated fund (originally created by an act of 1787). Consequently all the expenses of government have to be met from that fund, which is in the hands of the exchequer; and the balances to its credit at the Bank of England and Bank of Ireland (formerly at the head office in Dublin, now at the Belfast branch) are commonly called the exchequer balances. Issues cannot be made from the exchequer accounts at the Bank of England and Bank of Ireland without the consent of the comptroller and auditor general.

One of the essential duties of the comptroller and auditor general is that he does not give consent to the issues of any moneys from the exchequer accounts until he is satisfied that parliamentary authority has been given for the purposes for which the expenditure is required. The comptroller and auditor general thus plays an essential part in the machinery whereby parliament, by refusing supplies, can enforce its will on the government.

In order to safeguard certain services from the hazards of a political vote, permanent statutory authority is given for the exchequer issues required for such services. These, known as consolidated fund services, include the national debt charges (interest, maturities and sinking funds), the sovereign's civil list, the salaries and pensions of judges and of the comptroller and auditor general and some other items.

"Ways and Means" and "Supply."—Expenditure other than consolidated fund services is provided for by annual parliamentary votes. The parliamentary financial procedure for voting money comprises two separate processes. One is the appropriation of funds to particular services and purposes. This is handled by the committee of supply. The other is the issue of money from the consolidated fund to make good the grants of supply approved by the committee of supply and agreed to by the house of commons. This is dealt with by the committee of ways and means. An act called the Consolidated Fund act is always passed before April 1, when the new financial year begins, authorizing the issue of a sum equal to the totals of the votes passed in committee of supply. One of these votes (known as the vote on account) provides for expenditure by the civil and revenue departments and the ministry of defense from the beginning of the new financial year until complete provision has been authorized.

Appropriation Act.—The resolutions passed by the committees of supply and ways and means, when agreed to by the house, are valid so long as parliament is in session but require statutory confirmation before the end of the session. Accordingly an act, the Appropriation act, must be passed before the end of the session. The purpose of the act is to ensure that no money shall be spent on any service in excess of the amount granted in it for that service, and that no money shall be spent on any service which has not already been authorized by resolution of the house.

The treasury (fortified with a royal order under the sovereign's sign manual) applies to the comptroller and auditor general for authority to issue money as and when required, and he, in complying with the application for credit, is concerned only to know that the statutory authority exists for the total amount of the credit applied for. In directing issues from the exchequer against the credit granted, the treasury specifies the particular votes in respect of which the sums are drawn from the exchequer, and the duty of ensuring that the government conforms strictly to parliamentary authority, as expressed in the acts charging the consolidated fund and in the votes in supply, is enforced by the comptroller and auditor general, not through his control of the exchequer, but through his subsequent audit.

Exchequer Control.—The responsibility of the government to parliament is so well recognized that the very purpose of exchequer control, as the constitutional sanction of parliamentary government, is apt to be forgotten or overlooked. Occasionally a government (like the Liberal government engaged in 1910 in a conflict with the house of lords) may deliberately ask for ways and means in small installments in order that the house of commons may keep a hold over a possible alternative government, but this is rare.

Apart from its constitutional purposes, exchequer control provides, incidentally, a convenient index of the state of the public finances. As all receipts from revenue and loans are paid into the exchequer, and the money for public expenditure is issued out, the exchequer receipts and issues at any stage of the financial year show substantially the financial position up-to-date. Since the practice is to indicate the sources of the receipts under the principal

heads of revenue (customs, excise, death duties, stamps, income tax, surtax, etc.) and the application of the issues to the several consolidated fund services and to the supply services, it is possible to compile a detailed account showing the progress of each main head of revenue and expenditure. But this account would differ from a true account of revenue and expenditure on account of variations in the balances held on the one hand by the revenue departments and awaiting payment into the exchequer, and, on the other hand, by the paymaster general and the subaccountants of all the public departments and not yet spent. In normal times, however, this difference is not large.

The Annual Balance.—For the purpose of calculating the surplus or deficit for the year, revenue and expenditure are calculated on the basis of exchequer receipts and exchequer issues. To show the amount of the revenue surplus or deficit, the receipts and issues on capital account, commonly known as transactions below the line (such as the raising and repayment of loans, expenditure authorized by special acts to be met by borrowing, and repayments applicable to the redemption of debt), are shown separately in published statements of the transactions of the exchequer from those items which properly belong to the revenue and expenditure of the year.

Sinking funds arise in two ways. First come the contractual sinking funds set up under the terms of the prospectuses issued in respect of certain government securities. Secondly, under the Sinking Fund act of 1875, the surplus of revenue over expenditure in any financial year must be used for the redemption of debt and is known as the old sinking fund. This surplus is applied to debt redemption as it accrues during the year, the balance on the exchequer account being kept to a working level only.

On any voted service the government is not allowed to spend more than parliament has provided, and may spend less. Any foreseen excess is provided by supplementary votes; any unforeseen excess is disallowed unless, as invariably happens, an excess vote is provided in the following year. These supplementary and excess votes are accompanied by grants of corresponding ways and means. On the majority of votes there are savings within the amount of the parliamentary grant and the consequential balance on ways and means not spent is surrendered to the exchequer.

(E. E. Bs.; F. G. L.)

BIBLIOGRAPHY.—*Historical Aspects:* T. Madox, *History and Antiquities of the Exchequer of the Kings of England*, 2nd ed. (1769); *Dialogus de Scaccario*, ed. and trans. by C. Johnson (1950); Publications of the Pipe Roll Society, especially C. Johnson, *Introduction to Pipe Roll 2, Richard I* (1925); J. Hunter (ed.), *Pipe Rolls and Chancellors Roll*, for Record Commission (1844); H. Hall (ed.), *The Receipt Roll of the Exchequer for Michaelmas Term, 31 Henry II, A.D. 1185* (1899), *Court Life Under the Plantagenets* (1890), *The Antiquities and Curiosities of the Exchequer* (1898), *The Red Book of the Exchequer* (1896); R. L. Poole, *The Exchequer in the 12th Century* (1912); J. H. Round, *The Commune of London and Other Studies* (1899); H. Jenkinson, "Financial Records of the Reign of King John," *Magna Carta Commemoration Essays*, Roy. Hist. Soc. (1917); M. H. Mills, "Experiments in Exchequer Procedure" and "Reforms at the Exchequer," *Roy. Hist. Soc. Trans.* (1925, 1927), also her essay prefixed to *The Pipe Roll for 1295 Surrey Membrane*, Surrey Record Soc. (1924); Public Record Office, *List of Foreign Accounts* (1900), *List and Index of Declared Accounts* (1893); M. S. Giuseppe, *Guide to the Public Records*, 2 vol. (1923–24); D. M. Broome, "Auditors of the Foreign Accounts," *Eng. Hist. Rev.*, vol. xxxviii (1923); Mrs. Eric George, "Note on the Origin of the Declared Accounts," *Eng. Hist. Rev.*, xxxi (1916); T. Fanshawe, *Practice of the Exchequer Court* (1658); C. Vernon, *The Exchequer Opened* (1661); A. P. Newton, "The King's Chamber Under the Early Tudors," *Eng. Hist. Rev.*, xxxii (1917).

On the Jewish exchequer, *see* H. Jenkinson's papers, noted, with a difference of opinion on some points, by A. C. Cramer, in H. Jenkinson and B. E. R. Fermoy, *Select Cases in the Exchequer of Pleas*, Selden Soc. (1932); *Speculum*, xvi:226–9 (April 1941).

During the 20th century historical work threw light upon and led to discussions about the nature and use of exchequer materials and on the exchequer as a living and active element in the royal administration. *See* A. B. Steel in the *Amer. Hist. Rev.*, xxxiv (1929), and the article "Exchequer" in the index to T. F. Tout's *Chapters in Mediaeval Administrative History*, vi:241–251 (1920–33). In a long introduction to *The Memoranda Roll for Mich. 1 John, 1199–1200*, in Pipe Roll Soc. n.s. xxi (1943), H. G. Richardson raised many problems. J. F. Willard, *Parliamentary Taxes on Personal Property, 1290–1334*, Mediaeval Academy of America (1934), is an outstanding contribution to the study of exchequer practice. In addition to M. H. Mills's articles noted above, *see* also her paper "Exchequer Agenda and Estimate of Revenue, Easter Term, 1284," in *Eng. Hist. Rev.*, xi (1925); J. F. Willard's articles noted in the index to F. M. Powicke, *The Thirteenth Century* (1953) and K. B. McFarlane, "Loans to the Lancastrian Kings" in *Cambridge Hist. Jnl.*, ix (1947); also A. B. Steel's papers on English finance in the 14th century in *History*, xii (1928) and on the practice of assignment in *Eng. Hist. Rev.*, xliii (1928), and his book, *The Receipt of the Exchequer, 1377–1485* (1954); but see *Econ. Hist. Review*, 2nd series, vii:251–252 (1954–55); E. B. Fryde, in the *Bulletin of the Inst. of Hist. Research*, xxii, xxiii (1949–50). (F. M. P.)

Modern Developments: E. Hilton Young, *System of National Finance*, 3rd ed. (1936); R. G. Hawtrey, *The Exchequer and the Control of Expenditure* (1921); Sir Herbert Brittain, *The British Budgetary System* (1959); *Annual Reports of the Commissioners of Inland Revenue.* (F. G. L.)

EXCHEQUER BILLS. The exchequer bill was the credit instrument in which the principal part of the floating debt of the British government was embodied until its place was taken by the modern treasury bill in 1877. The exchequer bill was an obligation of the exchequer issued by direction of the treasury, negotiable and transferable by endorsement, and with a currency varying from three months to several years. Interest accrued on it daily at rates announced by the treasury from time to time. The original exchequer bills, introduced in 1696 during the chancellorship of Charles Montague, were payable at any time on demand, like bank notes. From 1697 holders were able to use them in payment of taxes or other sums due to the exchequer. In order to prevent bills from being presented for encashment to an inconvenient extent, the treasury relied on its power over the rate of interest. Thus if the bills were quoted in the market at a discount, making it profitable to holders to present them or sell them to taxpayers to be presented, the treasury would raise the rate of interest high enough to induce their retention. During the 18th and 19th centuries bills were issued under various authorities: as part of the permanent debt, renewable when due; to make up the charge on the consolidated fund at the end of a quarter; in anticipation of certain taxes; as "ways and means bills" for temporary borrowing; and to raise money to be lent to local authorities for public works. *See* DEBT, PUBLIC.

EXCISE TAX. An excise tax is levied on the purchase, production or sale of specific groups of commodities to raise revenue, to discourage the consumption of "luxuries" or to influence the allocation of resources. The term also includes exactions in the form of licences to engage in certain occupations and taxes paid for recording documents. Excise taxes have varied in nature and significance throughout different periods of history. In classical antiquity or the early middle ages, when a relatively small fraction of national incomes was embodied in marketable commodities, excises were a minor source of government revenue, though a major source of public annoyance.

The breakdown of feudalism and of the feudal fiscal apparatus of land taxes and per capita payments, the growth of industrialization and the acceptance of mercantilist policies increased the emphasis upon taxation of commodities to fill the treasuries of the new national states. Under mercantilist auspices excise taxes flourished in 16th- and 17th-century Europe, though distinguishing excises from the customs duties offers a difficult problem in classification. Following the practice of the times, the cities of France levied a tax upon articles of commerce entering their gates.

In the mercantilist period the sovereign often farmed out the excises to private monopolists who paid a fixed sum for the right to levy taxes on a designated list of commodities, a practice which Adam Smith in his *Wealth of Nations* denounced with exceptional vehemence. Contributing to this obnoxious expedient were the deficiencies of public administration, the desire to reward favourites and the need to secure revenues in advance of collections. Closely related to this device was the establishment of public monopolies by a number of continental governments for the manufacture and sale of commodities such as tobacco, where the tax appeared in the guise of monopoly profits. A famous example was the French tobacco monopoly inaugurated by the French statesman Jean Colbert in 1674.

From the Industrial Revolution to World War I, taxation of commodities generally appeared as a levy on the wholesale or retail value of a brief list of products, with alcohol and tobacco being

the favourite candidates for inclusion and sugar, soap, salt, leather, matches, perfume, jewelry and cloth appearing on the list at various times and places. The list grew and the rates rose during wars or fiscal emergencies. Beginning with World War I, a number of countries, states and cities employed a general tax on production or consumption as a major source of revenue. By the late 1950s and early 1960s these "indirect" excises—including sales and turnover taxes, customs duties and the revenue from public liquor and tobacco monopolies—provided half the public revenues for France and Italy, two-fifths for West Germany, Norway, Austria, Belgium, Finland and western Europe as a whole, and one-third for Sweden and Switzerland. Purchase taxes and other excises yielded one-sixth of the national revenue of Great Britain, and sales taxes produced one-fifth of the receipts of U.S. states.

Great Britain.—The British system of excises on goods produced within the country, most authorities agree, dates from 1643 when both the tax and the name were borrowed from the Dutch. Enlarged by the financial stresses of Cromwell's Commonwealth and Marlborough's wars later in the century, but diminished in the 18th century, excises were extensively used during the Napoleonic Wars. Following the suspension of the income tax in 1816, customs duties and excise taxes provided the bulk of the national revenue until the income tax returned and the tariffs were repealed during the 1840s. During the last half of the century, alcohol and tobacco were the mainstays of the excise.

In World War II the traditional excises on liquor, tobacco and other products were supplemented by the purchase tax at rates which varied from 16⅔% to 100% on the wholesale value of a wide variety of items. Intended to ration raw materials as well as to provide revenue, the purchase tax, following several postwar revisions, consisted in the early 1960s of 7 different rates from 5% to 90% on about 45 different classes of goods. (*See* also GREAT BRITAIN: *National Finance.*)

United States.—Debarred by the constitution from the employment of direct taxes, the federal government raised the bulk of its revenue, between the American Revolution and World War I, from tariffs and excises. A tax on distilled spirits in 1791 was followed by other excises in 1794. Spirited opposition to the liquor tax having been indicated by the Whisky Insurrection (*q.v.*) of 1794 in Pennsylvania, the "infernal system" of excises was swept away by Thomas Jefferson in 1802. To finance the War of 1812, taxes on liquor, sugar, carriages and several other items were levied from 1813 to 1818. No further excises were levied by the federal government until 1862 when taxes at low rates on a broad group of commodities were employed during the American Civil War. They were repealed by 1870, with the exception of the liquor and tobacco excises. The Spanish-American War and World War I repeated these advances and retreats. From the Civil War to 1930, excises accounted for roughly a third of all federal revenues.

During the depression of the 1930s federal excises at high rates were levied but most of them had been repealed by 1938. A number of state governments adopted a general tax on sales, at rates of 1% or 2%, to replace the revenues from property which they deeded to cities, counties and districts.

World War II saw the introduction of manufacturers' and retailers' excises at rates of 5% to 25% on a long list of goods and services. Many of these continued after the war and by the early 1960s were yielding about 15% of budgeted revenue. In 1965 congress repealed excise taxes on a great many items such as furs, jewelry, cosmetics, luggage, and home appliances, and reduced those on passenger automobiles and telephone service over a period of years. The taxes on distilled spirits, wine, beer, cigarettes, and airplane tickets were retained.

State Taxes.—Property taxes provided the bulk of state revenues till the early part of the 20th century, and thus the commonwealths did not enter the excise field substantially till the early 1920s. But as the states deeded the property tax to municipalities and school districts, excises on tobacco, gasoline and liquor increased. Some states also introduced liquor monopolies. Another widespread development was the general sales tax, covering virtually all retail transactions at rates averaging around 2% and providing up to a third or a half of the revenue of some U.S. states. By the early 1960s about three-fourths of the states levied sales taxes.

Economic Effects.—How is the burden of an excise shared between the persons who buy and the persons who make and sell the taxed commodity? The smaller the number of other things that he can buy in place of the product that is taxed, the greater is the share of the burden that falls upon the consumer. For items with few close substitutes, such as liquor, tobacco or perfume, the tax is shifted to the consumer in the form of higher prices. On the other hand, the smaller the number of other things that the producer can make in place of the product that is taxed, the greater is the share of the burden that falls upon the resources engaged in production. For resources with few alternative uses, such as coal mines or ore pits, the tax is partly shifted to the producer in the form of lower rents for the mineral rights. In general, the availability of substitutes in both consumption and production governs the share of the burden that the consumer and the producer bear. *See* also LUXURY TAXES; TAXATION; TAXATION, LOCAL.

BIBLIOGRAPHY.—Adam Smith, *The Wealth of Nations*, book v, ch. 2, part ii, art. 4 (1776); U. Hicks, *Public Finance*, ch. 8, 10 (1948); J. F. Due, *Sales Taxation* (1957), *Government Finance*, 2nd ed., ch. 18, 19 (1959); J. Frederic Dewhurst and Associates, *Europe's Needs and Resources*, ch. 13 (1961). (P. TN.)

EXCLUDED MIDDLE, LAW OF: *see* THOUGHT, LAWS OF.

EXCOMMUNICATION, defined by canon law as "a censure by which someone is excluded from the communion of the faithful" but not from membership in the church as such. Some form of exclusion belongs to the administration of all Christian churches and denominations, indeed of all religious communities.

Although the Christian practice of excommunication has roots in the Old Testament, the early church traced its origin to the explicit instructions of Jesus that anyone who refused to accept fraternal admonition should be denounced to the church, "and if he refuses to listen even to the church, let him be to you as a Gentile and a tax collector." It is evident that the earliest Christians regarded these instructions as binding upon their life in the church, for in his correspondence with the church at Corinth Paul charged the Corinthians to take action against one of their members guilty of incest: "When you are assembled, and my spirit is present, with the power of our Lord Jesus, you are to deliver this man to Satan for the destruction of the flesh, that his spirit may be saved in the day of the Lord Jesus." The later Epistles of the New Testament show that a judicial act of exclusion from the rights and privileges of church membership was the final step in dealing with those who "made shipwreck of their faith" either by flagrant and impenitent immorality or by a denial of fundamental Christian teachings such as the resurrection of Christ. The aim of this disciplinary action was principally to cure the offender, only secondarily to punish him. Yet the statement of the Epistle to the Hebrews is an indication of the strictness of church discipline in the church of the 1st century: "It is impossible to restore again to repentance those who have once been enlightened . . . if they then commit apostasy."

In the practice of modern Christian churches, excommunication is an application of this biblical material. Roman Catholicism distinguishes between two kinds of excommunication, that which renders a person *toleratus*, tolerated, and that which renders him *vitandus*, one who is to be avoided. Except for certain crimes that incur it automatically, the latter and more severe form requires that the culprit be announced by name in public as *vitandus*, in most cases by the Holy See itself; this is reserved for the gravest offenses. Both kinds of excommunication bar the excommunicated person from the sacraments of the church as well as from Christian burial. There is a specified list of actions that incur excommunication; it includes not only heresy, schism, blasphemous treatment of the eucharistic host, personal violence against the pope, and the like, but such offenses as joining forbidden secret societies or contracting marriage before a Protestant minister or a Jewish rabbi. If an excommunicated person confesses his sin and undergoes penance for it, he is absolved; in some cases this absolution may come from any priest, but in many others it is reserved

to the bishop or even to the Holy See alone, save *in periculo mortis*. Yet it must be remembered that "excommunication is a medicinal punishment; its primary and immediate purpose is to bring the delinquent back to a sense of duty" (F. E. Hyland). Excommunication should be distinguished from two related forms of censure: suspension and interdict. Suspension applies only to clergy and denies them some or all of their rights; interdict does not exclude from the communion of the faithful, but forbids certain sacraments and sacred offices, sometimes to an entire area.

Some churches do not use the term excommunication, preferring to speak of church discipline. Churches holding the Reformed order vest the authority for exercising discipline and, if need be, carrying out excommunication, in the session, which consists of the minister and the elders. The 30th article of the Westminster Confession of 1647 specified "admonition, suspension from the sacrament of the Lord's Supper for a season, and excommunication from the church" as the proper steps of discipline. The Lutheran tradition has followed Luther's catechism in speaking of "the power of the keys" and in defining excommunication as the denial of the Lord's Supper to public and obstinate sinners; the clergy and the congregation together have the right to exercise such discipline. In the Anglican Church the bishops have the right to excommunicate, as the Book of Common Prayer also suggests, but this right is almost never exercised. Where a congregational polity and the principle of "believers' baptism" are observed, discipline is often very rigorous. Thus American denominations of the Free Church tradition speak of "churching" a sinner; *i.e.*, of excommunicating him.

See Francis Edward Hyland, *Excommunication: Its Nature, Historical Development and Effects* (1928); Walter Doskocil, *Der Bann in der Urkirche* (1958). (J. J. Pn.)

EXCRETION (KIDNEY): *see* Urinary System: *Physiology.*

EXE, a river in England (Brit. *Isca,* "water"), rises at 1,500 ft. on Exmoor, northwest Somerset, 5 mi. from the Bristol channel and flows 60 mi. across the southwestern peninsula of England to the English channel at Exmouth. With its tributaries, principally the Barle, Culm, Creedy and Clyst, it drains 580 sq.mi. Its upper valley in Somerset, deeply incised in high plateau country, displays long east-west sections parallel with structures in the underlying Devonian rocks. Its middle portion, from the county boundary south to Exeter, is perpendicular to the east-west structural grain of central Devon and therein the valley is alternately narrow and steep-sided on Carboniferous rocks and broad, terraced, flat floored and liable to floods on Permian rocks. On the Permian outcrop below Exeter the river flows southeast in a broad, open valley with extensive terraces and a mile-wide flood plain and runs into its shallow estuary 10 mi. above a bar at Exmouth. A ship canal (1564–67) from the head of the estuary near Topsham makes Exeter, at the lowest bridging point, the head of navigation for small oil tankers and cargo boats. The Exe provides water for Exeter and power for paper and flour mills. It is an important salmon and trout river and yachting is popular on the estuary. (R. S. W.)

EXECUTIVE ORDER. The executive order is the principal mode of administrative action on the part of the president of the United States. It came into use before 1850, but the numbering system which was inaugurated about 1895 goes back only to the administration of President Lincoln. A famous executive order which is still in force except as amended from time to time is executive order no. 9, issued Jan. 17, 1873, by Pres. U. S. Grant to curb the abuse by which an individual often held at the same time a state and a national office. It received considerable publicity when it was modified by no. 4439 issued May 8, 1926, by Pres. Calvin Coolidge for the purpose of aiding prohibition enforcement by allowing the same persons to be simultaneously state and federal enforcement officers.

The president derives his authority to take administrative action by executive order either from the statutes of congress or from his constitutional powers, including the general power of direction or over-all management which flows from his implied constitutional position as the administrative superior of the heads of the executive departments of the national government and which finds its ultimate sanction in his power to remove those department heads.

The executive order stands in contrast with the proclamation (*q.v.*), which as a historic means of giving public notice of matters of widespread concern antedates the presidency, and which may or may not produce legal effects. Executive orders as well as proclamations of the president are published in the *Federal Register* and are made the subjects of White House press releases.

While many presidential proclamations have been related to such observances as Thanksgiving day, Constitution week and Red Cross month, those administrative actions which the president himself takes by executive order instead of leaving to his subordinates are apt to be significant.

In form the president's executive order is simpler than his proclamation. Its body, which usually has no "whereas" clauses, is followed in turn by his signature, the words "The White House," and the date expressed in the ordinary way. (J. Ht.)

EXECUTORS AND ADMINISTRATORS: *see* Estates, Administration of.

EXEDRA, a term originally describing any outdoor seating arrangement accommodating a number of people. In this sense, the word is properly applied to the exedra of Attalus II at Pergamum and the tomb of the priestess Mamia at Pompeii, both uncovered semicircular benches of monumental proportions. According to Vitruvius, enclosed exedrae formed an integral part of the Roman patrician house and were used as gathering places for conversation. Such rooms, open on one side like porticoes and usually facing the peristyle court, have been recognized in many houses at Pompeii. Exedrae could also be of large size, and, like apses, be covered with a hemispherical vault. Noteworthy examples of this architectural type exist in the baths of Caracalla, the temple of Heliopolitan Jupiter at Baalbek and the great exedra of the hippodrome on the Palatine hill in Rome. This semicircular exedra form was also applied by the Romans to large-scale civic design such as the imperial forums of Augustus and Trajan, and this usage has again found favour in modern times; *e.g.*, the 19th-century Piazza dell'Esedra in Rome. (H. Hn.)

EXEGESIS AND HERMENEUTICS, BIBLICAL, the branch of Christian theology that deals with the interpretation of the Bible. The distinction between hermeneutics and exegesis is between theory and practice, hermeneutics being the definition of the principles according to which the Bible is to be interpreted and exegesis being the application of those principles; historically and practically, of course, hermeneutics and exegesis are inseparable. Thus biblical exegesis, based upon hermeneutical principles, occupies much the same place in Christian theology that the Midrash, the Talmud and later commentaries occupy in the thought and literature of Judaism.

It has been suggested that the entire history of the Christian church is actually the history of biblical exegesis and hermeneutics. Not only in its formal expositions of the Scriptures but in its concrete life and institutions Christianity has been engaged in trying to make sense of the Bible and to apply it to "the changes and chances of this mortal life." What was implicit in the text of the Bible became explicit in the life and history of the church. For instance, the history of the papacy has been interpreted as the gradual explication, over a period of many centuries, of the words of Jesus to Peter: "On this rock I will build my church." Nevertheless, it is useful to concentrate more specifically upon the history of biblical hermeneutics and exegesis as a special aspect of the teaching function of the church and an important illustration of the dialogue between Christian theology and the history of culture. The exegesis of the Bible today, moreover, is largely shaped by the history of exegesis, and without a knowledge of the history of biblical hermeneutics neither *The Divine Comedy* nor *Paradise Lost* nor *Faust* will receive just treatment from an interpreter. This article (to be read in close conjunction with the article Bible) will sketch the most important periods in the history of the Christian interpretation of the Bible.

The New Testament.—The history of the Christian interpretation of the Bible begins within the Bible itself, for the New

Testament is a confession about Jesus Christ that is based upon an exegesis of the Hebrew Scriptures as the "Old Testament." Scarcely a book of the New Testament is without a quotation from the Old Testament; some books of the New Testament, such as the Gospel of Matthew or the Epistle to the Hebrews, are practically series of passages from the Old Testament with a commentary. Most of the central terms of the New Testament writers are unintelligible apart from their origin and context in the Old Testament—a judgment that applies even and especially to the sayings of Jesus themselves.

In the Sermon on the Mount, as set down in the Gospel of Matthew, Jesus is represented as saying: "Think not that I have come to abolish the law and the prophets; I have come not to abolish them but to fulfil them." Yet a few verses later he says: "You have heard that it was said to the men of old . . . . But I say to you." Taken together, these two statements summarize two of the main features in Jesus' exegesis of the Old Testament. In his treatment of the Mosaic Law he manifested great independence, looking for the intent and spirit of the Old Testament legislation and elevating this above mere observance. At the same time he saw himself—or so, at least, the four Gospels unanimously describe him—as the personal fulfillment of the messianic prophecies in the Old Testament. "Today this scripture has been fulfilled in your hearing," he said in the synagogue at Nazareth after reading from Isaiah. When he was tempted by the Devil in the wilderness, he defended himself with citations from the Book of Deuteronomy, and his final cry of dereliction on the cross was a quotation from one of the Psalms. Many individual interpretations of the Old Testament in the Gospels bear a strong resemblance to the *haggadah* and the allegorical spiritual methods of Hellenistic Judaism, but more prominent are the ways in which Jesus set himself apart from contemporary exegesis of the Old Testament.

The dual themes of fulfillment and freedom are characteristic also of the hermeneutics of the apostle Paul as he interpreted the Old Testament. His Epistle to the Galatians stated the theme of freedom from the Law more vigorously than did any other New Testament writing, but the same epistle presented the most familiar allegory upon the Old Testament to be found anywhere in the New, the contrast between Sarah and Hagar. The point of that allegory was the fundamental presupposition of Paul's hermeneutics and exegesis: that in Jesus Christ and in his redemptive work the history and message of the Old Testament had found their consummation. "Christ," he wrote elsewhere, "is the end of the law"—"end" both as goal and as conclusion. Paul, who had studied under the great rabbinical exegete Gamaliel I, employed many of the exegetical devices developed by the rabbis, but he used them to support a hermeneutic that the rabbis rejected, namely, the fulfillment of the prophecies in Jesus as the Christ. The cloud and the rock in the Exodus were actually Christ; the story of Adam in Genesis supplied a comparison and contrast with Christ as the second Adam; Abraham was the father of all believers—*i.e.,* of Christians as the true Israel. This interpretation of the Old Testament as a Christian book has dominated the history of Christian hermeneutics and exegesis, even though later interpreters have gone beyond Paul in defining the authority of the Old Testament as well as in their use of allegory, typology and other techniques for christianizing the Law and the prophets.

Even some books of the New Testament went beyond Romans and Galatians in both these respects. The two Epistles to Timothy and that to Titus, whose composition by Paul has been questioned by modern critics, contain the most complete statement of the inspiration of the Old Testament to be found anywhere in the Bible: "All scripture is inspired by God and profitable for teaching, for reproof, for correction, and for training in righteousness." But no book of the New Testament presented as refined an exegesis of the Old Testament as did the Epistle to the Hebrews. Whether or not Paul was its author, this epistle reflected upon the Christian meaning of various Old Testament passages with a hermeneutic unmatched anywhere else in the New Testament for thoroughness and consistency. Christ was the eternal priest, fulfilling the liturgical history of Israel; it was to Christ that the mysterious story of Melchizedek in Genesis pointed; the saints of the Old Testament had achieved their elevated status "by faith"—*i.e.,* by a trustful anticipation of what Christ was to bring in the days of his flesh.

Other books of the New Testament, though not so elaborate in their exegesis, shared its hermeneutical assumption and joined with it in the adaptation of Old Testament materials to Christian use. Thus the story of the flood and of Noah's ark became a type of baptism for the First Epistle of Peter; in the Epistle of James the story of Job provided the ground for an admonition to Christian steadfastness. By the time the last book of the New Testament had been written, Christians had so identified themselves with the Old Testament that they could regard Judaism as the interloper upon the ancient Scriptures and the Christian church as the rightful possessor and authoritative interpreter of the Jewish Bible.

More fundamentally than any subsequent stage in the history of Christian hermeneutics, the New Testament laid down the lines for later developments. Its exegesis of the Old Testament in the light of the life, death and resurrection of Jesus Christ has been normative for most Christians of all centuries, whose creeds, dogmas and liturgies echoed the formula of the apostle Paul: "Christ died for our sins in accordance with the [Old Testament] Scriptures, and was raised on the third day in accordance with the [Old Testament] Scriptures."

The Early Church.—During the 2nd and 3rd centuries Christianity successfully defended its hermeneutical approach to the Old Testament against attack, and it developed principles for the exegesis of the New Testament as well. The attack upon the place of the Old Testament in the church came chiefly from Marcion (*q.v.*), who separated the God of the Old Testament from the Father of Jesus Christ and compiled a canon of Christian writings to set into opposition with the Jewish Scriptures; this canon consisted of an emended Gospel of Luke and of those writings of Paul that Marcion acknowledged as authentic. The outcome of the controversy with Marcion over hermeneutics was the vindication of a "spiritual" over a literal exegesis, but the establishment of a New Testament alongside the Old provided Christian exegetes with a Bible consisting of two testaments in one sacred book. Henceforth Christian exegesis was to be the interpretation of the entire Bible, setting the Old Testament into the framework supplied by the New and putting the New Testament into the continuity supplied by the Old. As a result, a "spiritual" hermeneutic for the Old Testament required the allegorization of its narratives, while the narratives of the New Testament could be read chiefly for their literal or historical import, since this was also their "spiritual" significance.

Irenaeus (*q.v.*) was perhaps the most important of the theologians who established a "spiritual" hermeneutic of this kind. He interpreted the historical accounts of the Old Testament as authentic history but as a history that contained types of Christ and of the Christian community. Both the interpretation of the Old Testament and the exegesis of the New were subject to the hermeneutical authority of the "rule of faith," the normative tradition of the church as preserved in its succession from the apostles. This he contrasted with the claims of the Gnostics that their communities had preserved a secret tradition going back to Christ and the apostles (*see* GNOSTICISM). Biblical history was the history of salvation, in which Christ as the second Adam came to recapitulate and to redeem the history of humanity in the first Adam.

The hermeneutics of Tertullian (*q.v.*) was in many ways a further development of that of Irenaeus. For Tertullian, too, there was an authoritative interpretation of Scripture set down in the rule of faith; because heretics rejected this hermeneutical a priori, they were incapable of interpreting the Bible properly. In his later years Tertullian became an adherent of Montanism (*q.v.*) and read the Scriptures from the hermeneutical a priori of that sect's apocalypticism and its insistence upon moral rigour. Both Irenaeus and Tertullian appear to have drawn extensively upon an exegetical tradition—reflected, as has been shown, already in the New Testament—that was at work also in the liturgies, the sermons and the devotional life of the Christian community.

Schools of Alexandria and Antioch.—But the most important Christian exegete of the 2nd and 3rd centuries was neither Irenaeus nor Tertullian, but Origen (*q.v.*). Clement (*q.v.*) of

Alexandria, who was Origen's predecessor in Alexandria and probably his teacher, adapted to Christian use the hermeneutics developed by Philo, the Jewish theologian of Alexandria and contemporary of Jesus. Philo debated against an extremist school of interpreters, who negated the literal meaning of Scripture in the belief that only the "spiritual" meaning was valid. Continuing the Philonic tradition, the hermeneutics of Clement sought the allegorical sense, which conveyed deep philosophical truths in the plain-sounding narratives and precepts of the Bible. But it was left to Origen to systematize these hermeneutical principles and to apply them in the grand manner to the exegesis of the Scriptures. The basis of this method was a minute study of the biblical text; for this purpose Origen compiled the *Hexapla,* which lined up in parallel columns the Hebrew text of the Old Testament (written in both Hebrew and Greek characters) and four (in some portions as many as seven) Greek translations of the Old Testament. Such minute study persuaded Origen that the language of the Bible had never been intended to convey the simple and crude meaning that both pagan critics and naïve believers found in it. "The letter killeth, but the spirit giveth life"—these words from II Cor. iii, 6 provided him with justification for elevating spiritual, that is, allegorical exegesis far above the literal.

> The individual [Origen wrote] ought to portray the ideas of Holy Scripture in a threefold manner upon his own soul; in order that the simple man may be edified by the "flesh," as it were, of Scripture—for so we name the obvious sense—while he who has ascended a certain way [may be edified] by the "soul," as it were. And the perfect man [may receive edification] from the law of the spirit. For as man consists of body and soul and spirit, so does Scripture.

Applying these hermeneutical principles to the concrete task of interpreting the Bible, Origen composed exegetical works of three kinds: explanations of individual passages that had proved troubling or were especially important (scholia); commentaries on nearly all the books of the Bible *in toto;* and "homilies," which seem to have been lectures rather than sermons. The spiritual depth and intellectual power of Origen's exegesis and hermeneutics dominated Alexandrian theology for generations. Although the Alexandrian orthodoxy represented by Athanasius (*q.v.*) repudiated many of Origen's doctrinal views, specifically his subordination of the Son to the Father in the doctrine of the Trinity, it continued to rely upon the philosophical and hermeneutical tradition of Origenism, which came to mean an extensive use of allegorical exegesis and a preoccupation with the "spiritual" rather than the literal sense of Scripture even in the life of Jesus. Thus the exegesis of Cyril (*q.v.*) of Alexandria sought to interpret the Gospel records as accounts about the earthly career of the incarnate Logos rather than of the man Jesus.

Both this Alexandrian view of Jesus Christ and the hermeneutics that supported it came under attack from theologians who were identified with the see of Antioch. Of these the most important was Theodore (*q.v.*) of Mopsuestia. He composed a treatise *On Allegory and History,* no longer extant, in which he criticized Origen for ignoring the literal sense of Scripture. Those who interpreted Scripture allegorically, Theodore said elsewhere, "turn everything backwards, since they make no distinction in divine Scripture between what the text says and dreams in the night." Like all his predecessors since the New Testament, Theodore found "types" of Christ in the Old Testament, but he urged that typology was valid only if the exegete took the historicity both of the type and of the antitype seriously: both the life of Adam and the life of Christ as the second Adam were to be read literally. Even the messianic prophecies of the Old Testament were to be viewed first in their historical context as sayings addressed to contemporary conditions; only then was it permissible to apply them to Christ and to the New Testament.

More famous than Theodore, though less original as an exegete and scholar, was John Chrysostom (*q.v.*), who managed to escape the opprobrium that came upon Theodore and upon much of Antiochene exegesis through its association with the cause of Nestorianism. As a preacher and expositor of Scripture, Chrysostom concentrated upon the literal sense of the text and upon its moral application. His standing in both the eastern and the western parts of the church served as a counterweight to the prestige of Alexandrian hermeneutics and exegesis in the church's tradition.

During the 6th century and afterward, that tradition was gradually achieving consolidation among the Greeks. Then most of the exegetical writing consisted of catenae or chains, which strung together passages from various Fathers. These catenae took the place both of independent exegetical work and of the study of the writings of the Fathers in their entirety. So utterly did the catenae dominate exegesis that we are dependent upon them for much of what we know about earlier exegesis. In A.D. 692 the "Quinisext" council decreed:

> If any controversy in regard to Scripture shall have been raised, let [the clergy and bishops] not interpret it otherwise than as the lights and doctors of the church in their writings have expounded it, and in these let them glory rather than in composing things out of their own heads.

Both Alexandrian and Antiochene hermeneutics were thus incorporated into the exegetical tradition of the fathers, from which both Greek and Latin theologians drew much of their interpretation of the Bible.

Latin Fathers.—Although the exegetical learning of the Greek Fathers shaped Christian hermeneutics also in the Latin-speaking west, occidental theology underwent its own distinctive development. Probably the most influential biblical scholar in the history of the western church was Jerome (*q.v.*), *doctor maximus in exponendis Sacris Scripturis,* "the greatest doctor in interpreting Sacred Scripture." Primarily, of course, this reputation and influence were due to Jerome's work as an editor and translator of the Latin version of the Bible, but he also became an important champion of literal exegesis in opposition to allegorical excesses. Initially he admired and imitated the allegorical method of Origen, but gradually he came under the influence of Antiochene exegesis; and in his hermeneutical theory, if not always in his exegetical practice, Jerome preferred the literal to the allegorical sense. As Pope Benedict XV said of him in 1920,

> Jerome, in imitation of the Latin and Greek exegetes who preceded him, made use first of allegorical interpretations to perhaps an excessive degree. But his love of the Sacred Books and his unceasing toil in constantly reading them and in weighing their meaning led him ever on to a right appreciation of their literal sense and to the formulation of sound principles regarding it.

Jerome's term for this literal sense was *Hebraica veritas,* the truth in Hebrew, upon which any allegorical or spiritual exegesis had to be founded. As one of the few early Christians to learn Hebrew, Jerome mediated to later generations of exegetes not only Alexandrian and Antiochene exegesis but the work of the rabbis as well.

Augustine (*q.v.*) may have been less a scholar than Jerome, but he was vastly more profound a mind. Both by his many commentaries on individual books and by the handbook on hermeneutics that he wrote about 397, *De doctrina christiana,* Augustine shaped the history of western exegesis. During his early Manichaean period he had spurned the Old Testament, but the allegorical exegesis of Ambrose showed him how the Old Testament could acquire a spiritual meaning. As his Christian understanding matured, he relied less exclusively upon allegory, but he continued to employ it where it seemed to bring out the deepest sense of the text. Augustine's sermonic expositions of the Psalms and his exegesis of St. John were perhaps the masterpieces of his work as an interpreter. In both he displayed the significance of his hermeneutics for the explanation of the Bible, interpreting Scripture by means of Scripture and finding parallels from many other parts of the Bible to illumine the verse at hand. If these parallels often seem bizarre to the modern reader, they are more often a brilliant penetration into the language and the thought of the biblical writers. The same hermeneutical principle was at work in the exegetical arguments of Augustine's controversies with the Donatists and the Pelagians. It is the opinion of many scholars, however, that Augustine's *On the Trinity* was the high point not only of his philosophical work but of his hermeneutics and exegesis as well. The central dogma of Christian theology served here as a hermeneutical key to the understanding of individual passages and themes from both the Old and the New Testament. The unrivaled eminence of Augustine as a doctor of the church has guaranteed to

his exegesis a constitutive role in the history of biblical interpretation, both Roman Catholic and Protestant.

In the western as in the eastern sections of the church, the close of the patristic period witnessed the creation of an authoritative tradition of hermeneutics and exegesis. Vincent of Lérins, a contemporary and a tacit critic of Augustine, coined the classic formulation of the norm for this tradition:

> The line of the interpretation of the prophets and apostles must be directed according to the norm of the ecclesiastical and Catholic sense. . . . What has been believed everywhere, always, by everyone [*quod ubique, quod semper, quod ab omnibus creditum est*].

Middle Ages.—The authority of the tradition was the dominant element in the exegesis and hermeneutics of the medieval church. Among early medieval writers, two of the most important for the history of interpretation were Gregory the Great and the Venerable Bede (*qq.v.*). Both were essentially transmitters of the exegetical tradition rather than bold and original exegetes, but the *Moralia* of Gregory (a sermonic exposition of Job) and the many commentaries of Bede displayed a religious insight and a rhetorical power that cannot be dismissed as sheer traditionalism. The catena was still the chief form for the compilation of exegetical authorities; to it the early middle ages added the "gloss." The settings for the biblical gloss were the monastic and cathedral schools, later the medieval universities. There the exegetical tradition—or, more precisely, traditions—received the codification that prevailed in western hermeneutics and exegesis for nearly a millennium, the *Glossa ordinaria* compiled by Anselm of Laon.

From the tradition, medieval compilers like Anselm drew not only the exegesis of words and passages in the Scriptures but also the hermeneutics that had produced his exegesis. Some time in the middle ages there was composed a mnemonic verse to distinguish four senses of Sacred Scripture:

> *Littera gesta docet, quid credas allegoria,*
> *Moralis quid agas, quo tendas anagogia.*
> (The letter shows us what God and our fathers did;
> The allegory shows us where our faith is hid;
> The moral meaning gives us rules of daily life;
> The anagogy shows us where we end our strife.)

Using this hermeneutical theory, medieval exegesis could take into account both the allegorism of the Alexandrian school and the emphasis upon literal interpretation that had come through theologians like Chrysostom. Of course, medieval hermeneutics still had to stratify the relations among the four senses. The two theological luminaries of the latter part of the 13th century, Thomas Aquinas and Bonaventura (*qq.v.*), interpreted the relation in a way that assigned priority to the literal sense but acknowledged the legitimacy of the other three senses.

At the end of the very first question in his *Summa Theologica* Aquinas took up the fundamental hermeneutical issue: "Whether in Holy Scripture a word may have several senses?" His answer established the literal sense as the primary and decisive one, from which any "spiritual" exegesis had to proceed. For the literal sense was the sense according to which words signified things. But there was also a sense according to which these things signified other things in turn; this was the spiritual sense, divided into the allegorical, the moral and the anagogic:

> So far as the things of the Old Law signify the things of the New Law, there is the allegorical sense; so far as the things done in Christ . . . are signs of what we ought to do, there is the moral sense. But so far as they signify what relates to eternal glory, there is the anagogical sense.

Nevertheless, the literal sense stood as "that which the author intends." When he applied this hermeneutic to the practice of exegesis, as he did in his commentaries and in his *Catena aurea* on the Gospels, Aquinas manifested his personal preference for the literal sense but also his readiness to accept the allegorical interpretations he was taking over from the Fathers, particularly the Greek Fathers, whose writings were being rediscovered in the west as a result of new contacts between the Greek and Latin churches. Among the Franciscan theologians of the later middle ages, the most important exegete was Nicholas (*q.v.*) of Lyra, whose *Postillae perpetuae in universam S. Scripturam* shared with Aquinas an insistence upon the primacy of the literal sense of Scripture. Nicholas' commentary and the *Glossa ordinaria*, reflecting a deep familiarity with the work of the Jewish commentator Rashi, were the indispensable handbooks for most exegetes at the end of the middle ages.

Reformation.—In some ways the most revolutionary and least traditional hermeneutics of the 15th and 16th century was not that of the Protestant reformers but that of the biblical humanists. What Erasmus (*q.v.*) called "the philosophy of Christ" was an interpretation of the Gospels that claimed to recover the religious and ethical message of Jesus, while by his achievements in the textual criticism of the New Testament Erasmus helped to establish the foundations for the later historical-critical exegesis of the Bible.

Nevertheless, Erasmus remained devoted to the tradition of the church, including its exegetical tradition, and in his hermeneutics he championed the cause of allegorical interpretation. Despite surface appearances, therefore, the Reformation, not the Renaissance, was the new stage in the history of Christian hermeneutics and exegesis. Using the textual and historical tools developed by Renaissance humanism, the Protestant reformers derived from the Bible a hermeneutic far more radical than that being propounded by the leaders of the new learning.

Distinguishing the hermeneutic of the reformers from the tradition was the principle that Scripture was its own interpreter. Luther did not hesitate to break with the tradition when the clear text of the Scriptures appeared to demand another interpretation. His early lectures on the Psalms and on the Epistle to the Romans were the setting of his struggle for religious certainty and for the meaning of the Gospel. When he concluded that justification by grace through faith was the theme of the Scriptures, he applied this as a hermeneutical principle to the exegesis of both the Old and the New Testament. Perhaps his most influential work in exegesis was his *Commentary on Galatians* of 1535. But Luther stressed also that it was necessary for the interpreter of Scripture to experience the meaning of the text personally; the experience came from the word of God, and the true meaning of the word of God could be discovered only in the experience. "That which urges Christ" thus became the criterion for a canon within the canon, as well as the key to a proper hermeneutics. Luther broke sharply with the allegorical tradition, at least in his hermeneutical principles, and demanded that the literal sense of the text, and it alone, be the arbiter of Christian doctrine. In his own exegesis, however, he continued to practise allegorical interpretation, even in his last commentary, the *Lectures on Genesis*. The work of Luther made the exegesis of Scripture the central concern of Protestant theology.

John Calvin took exegesis as his central concern, too. Although he is best known for his *Institutes of the Christian Religion*, he prepared commentaries upon most of the Old and the New Testament. There are many similarities between Calvin's hermeneutics and Luther's, in their insistence upon the sole authority of Scripture and their attention to literal interpretation. But while Luther's method of interpretation was intuitive, striking at the meaning of the text in flashes of illumination, Calvin's was more sober but also less exciting. Both Luther and Calvin continued to find the doctrine of the Trinity in the Bible, even in the Old Testament, and otherwise to repeat traditional teachings, in spite of their principle of *sola Scriptura*.

The left wing of the Reformation, however, came to more radical conclusions in the exegesis of the Bible. Some of the radical reformers employed the antithesis between the spirit and the letter, not to justify allegorical exegesis but to claim private revelations by which they interpreted Scripture. Others read the ethical imperatives of the Sermon on the Mount and even the Mosaic legislation of the Old Testament literally—an exegesis that led either to pacifism or to a militant effort to establish the Kingdom of God on earth. The Socinians and Anti-Trinitarians, accusing the classical reformers of inconsistency between their hermeneutics and their exegesis, rejected the traditional doctrines of the Trinity and the deity of Christ on the grounds that they were not taught *expressis verbis* in the Bible.

By the time the Reformation of the 16th century was over, Protestant hermeneutics and exegesis had manifested great variety

both in its methods and in its conclusions; in this sense it had broken with much of the tradition. But it had also introduced into theology a new seriousness about the meaning and message of the Scriptures; in this sense it had continued and deepened the tradition. There is, paradoxically, greater continuity between the hermeneutic of Aquinas and that of Luther or Calvin than the theological descendants on either side would acknowledge.

Protestantism Since the Reformation.—The theological results of the Reformation, achieved through exegetical exploration, were codified in confessional standards that inhibited further exploration. Protestant Orthodoxy, both in its Lutheran and in its Calvinist form, was responsible for the production of massive and learned works in biblical exegesis. The Calvinist scholars Johannes Buxtorf the Elder and Younger, as well as the Anglican scholar John Lightfoot, made available a vast amount of philological, historical and especially rabbinical erudition; Lutheran theologians like Martin Chemnitz, Johann Gerhard and Abraham Calovius composed extensive commentaries upon the Bible in whole or in part. But all this learning in biblical lore and language was used to support their previous theological positions rather than to achieve any new interpretations of the Scriptures. Hermeneutically the most creative of 17th-century Protestant theologies was perhaps that of Johannes Cocceius, who saw in the notion of the covenant between God and man the key to biblical hermeneutics and who attempted to make the history of the covenant, rather than philosophical or creedal categories, the organizing principle of his theology. A call for the priority of exegesis over dogmatics was one of the central elements in the platform of Protestant Pietism. Although the exegesis of Pietism paid more attention to the devotional and practical use of the Bible than to the scholarly explication of its meaning, its emphasis upon the Bible undoubtedly helped to stimulate the exegetical work of the 18th and 19th centuries. The *Gnomon* of Johann Albrecht Bengel was the climax of Pietist exegesis, influencing not only the subsequent interpretation of the Bible on the continent (it had gone through at least six editions by the end of the 19th century) but British exegesis as well; it was, in fact, basic to the *Notes on the New Testament* of John Wesley and to the Greek New Testament of Henry Alford.

Not in Pietism, however, but in a movement opposed both to Pietism and to Orthodoxy, in the Enlightenment, the study and interpretation of the Bible broke through to new ground. One of the earliest thinkers of the Enlightenment to reopen the question of the interpretation of the Bible was the Jewish philosopher Spinoza, who cast doubt upon the Mosaic authorship of the Pentateuch and interpreted the Bible according to the rational and historical principles appropriate to the exegesis of other books. Such ideas arose among Christians of the 18th and 19th centuries, too, and thus the historical criticism of the Bible became an indispensable part of Protestant exegesis and hermeneutics, as it has been ever since; Johann Ernesti, Johann Michaelis and Johann Salomo Semler were chiefly responsible for this change. Like them, most of the leaders of the critical-historical method of exegesis were German professors of theology, but the same approach to the study of the Bible appeared elsewhere as well. In Great Britain, for example, the historical exegesis of the scholars received support from men of letters like Samuel Taylor Coleridge and Matthew Arnold. Yet Protestant Germany remained the home of biblical criticism, and the 19th century was its heyday.

Among the many historical and literary issues that engaged the exegesis and hermeneutics of German Protestantism in the 19th century, three were of decisive importance. The historical criticism of the Pentateuch, already referred to, came to a climax in the work of Julius Wellhausen (*q.v.*), who based his exegesis of the Old Testament upon the historical evolution of the religion of Israel. As pentateuchal criticism endeavoured to isolate the several sources of the Books of Moses, so the study of the Synoptic Gospels attempted to chart the literary history of the New Testament and thus to find "the essence of Christianity" by locating the authentic facts about the life and teachings of Jesus; the sensation caused by the works of David Friedrich Strauss on the life of Jesus drew the attention of many people in the churches to what was going on in biblical exegesis. The third issue debated in the exegesis and hermeneutics of the 19th century was the proper method for the study of the thought and writings of Paul; Ferdinand Christian Baur, one of the most seminal theological scholars of the 19th century, saw in the conflict between Peter and Paul a key to the exegesis of the entire New Testament. Although the scholars were by no means agreed about most of the details of exegesis, by the end of the 19th century the historical-critical method had been widely accepted as the only hermeneutics that was both faithful to the historical character of the Bible and capable of making its message understandable to modern men.

Roman Catholicism Since the Reformation.—Not only for Protestantism but also for Roman Catholicism, the Reformation of the 16th century was a watershed in the history of the interpretation of the Bible, for the hermeneutics and exegesis of the Roman Catholic Church in modern times has been continually affected by the debates between it and Protestant theology. In opposition to the Protestant principle of "Scripture alone," the Council of Trent declared that it "receives and venerates with a feeling of piety and reverence all the books both of the Old and New Testaments . . . also the traditions . . . preserved in the Catholic Church in unbroken succession." Proceeding from this, the council decreed that no one relying on his own judgment shall, in matters of faith and morals pertaining to the edification of Christian doctrine, distorting the Holy Scriptures in accordance with his own conceptions, presume to interpret them contrary to that sense which holy mother church, to whom it belongs to judge of their true sense and interpretation, has held and holds, or even contrary to the unanimous teaching of the fathers.

It seems clear from later research that these decrees were intended not to set up Scripture and tradition as two separate sources of knowledge about divine revelation but to assert the hermeneutical interdependence of the Bible and the tradition of the church in the task of exegesis. Nevertheless, both the critics and the defenders of the Council of Trent took it to mean two separate sources, and much of Roman Catholic hermeneutics in the 16th and 17th centuries was based upon this presupposition. This has been called "the golden age of Catholic exegesis," from 1546 to 1660; it has been estimated that more than 350 Roman Catholic theologians busied themselves with exegesis during this period. But the defense of traditional hermeneutics and exegesis against Protestantism demanded such a large share of their attention that most of their work was devoted to conservation rather than to creation.

The same Enlightenment that was responsible for the rise of biblical criticism within Protestant theology affected hermeneutics and exegesis in the Roman Catholic Church as well. Richard Simon worked to interpret the Old Testament by the canons of historical-critical study, but he met with bitter resistance from other Roman Catholic exegetes. Although the cause of biblical criticism was espoused by various scholars like Ernest Renan, who left the church early in his career, the most vigorous discussion of the historical-critical method of exegesis was provoked by the rise of Modernism (*q.v.*) at the end of the 19th and the beginning of the 20th century. Partly under the influence of liberal Protestant thought, the hermeneutics of the Modernists taught that the biblical writers of both the Old and the New Testament were conditioned by the times in which they lived and that there had been evolution in the history of biblical religion. It was only a matter of time until the implications of their radical position clashed with the official teaching of the church, which condemned Modernism as a "synthesis of all heresies" in the encyclical *Pascendi gregis* and the decree *Lamentabili*, both issued by Pope Pius X in 1907. Shorn of its radical excesses by this action, the historical-critical method of biblical exegesis, espoused by Modernism, has continued to play a role in Roman Catholic theology.

Alongside the Modernist movement, opposed to its conclusions but not unaffected by its historical-critical hermeneutics, the main body of Roman Catholic exegesis flourished in the 19th century. Although the ecclesiastical standing of the Latin Vulgate had been reaffirmed at Trent, biblical scholars began turning more and more to the Hebrew and Greek texts. The vast historical research of the 19th century, both inside and outside Roman Catholic scholarship, provided the exegetes of the church with more knowledge about the total context of the Scriptures than had been available

to any of their predecessors. And the growth in lay education meant that the exegesis of the Bible for the entire membership of the church would gradually become an obligation of its teachers. In the religious and secular press it may have been Modernism that received the publicity, as it still does in the textbooks of church history, but the state of Roman Catholic hermeneutics and exegesis in the 20th century (*see* below) was evidence for the deepening consciousness among Roman Catholic theologians of the centrality of biblical exegesis in the theological enterprise.

20th Century.—In both Roman Catholicism and Protestantism the 20th century witnessed a new interest in biblical hermeneutics and exegesis, wide discussion of the fundamental theological issues and the growth of attention to the ecumenical implications of biblical interpretation.

Well equipped with a knowledge of ancient languages and with the help of critical scholarship, let the Catholic exegete approach that task which of all those imposed upon him is the highest, namely, to discover and set forth the true meaning of the Sacred Scriptures. In this work let interpreters keep in mind that their greatest care should be to discern and define what the so-called literal sense of the language of the Bible is.

These words from the encyclical *Divino afflante Spiritu*, issued by Pope Pius XII on Sept. 30, 1943, were the charter of Roman Catholic hermeneutics and exegesis in the 20th century. Under the aegis of this encyclical, Roman Catholic scholarship in many countries, especially in France and in Germany, was employing the historical-critical method of study and the results of modern literary and philological research to translate and interpret the Scriptures both for the theologian and for the lay public. The 20th century saw the proliferation of translations of the Bible as well as the preparation of notes, helps and commentaries for the lay student of Scripture. At the level of scholarly exegesis, the interest in literal, historical exegesis was combined—as, for example, in the hermeneutics of the French Jesuit Jean Daniélou—with a new appreciation and study of the typology of early Christian exegesis. The Pontifical Biblical commission continued to hold the line against any revival of Modernism in the critical study of historical questions, but without transgressing the decisions of the commission Roman Catholic hermeneutics and exegesis showed great freedom and creativity.

Protestant exegesis and hermeneutics in the 20th century were challenged by the biblical theology of Karl Barth. The opening sentences of Barth's epoch-making commentary on *The Epistle to the Romans* of 1919 simultaneously acknowledged the legitimacy and defined the limits of the historical method of interpretation:

Paul, as a child of his age, addressed his contemporaries. It is, however, far more important that, as prophet and apostle of the kingdom of God, he veritably speaks to all men of every age.

Recovering what he once called "the strange new world within the Bible," Barth's exegesis went beyond the hermeneutics of historical criticism to a method of biblical interpretation closer to Luther's. The way for this shift had been prepared, ironically, by some of the results of historical criticism. When Albert Schweitzer applied the historical method to the study of the Gospels, the outcome was not, as his mentors, such as Adolf von Harnack, had claimed, an "essence of Christianity" that could be extracted from its historical setting, nor what Wilhelm Herrmann had hoped for, an "inner life of Jesus" that could serve as a norm. Instead, Schweitzer found in the New Testament an expectation of the speedy end of the world, something far removed from the optimism and ethical idealism of the late 19th century. Yet the question of what to do with the relation between the Bible and its thought-world continued to engage Protestant hermeneutics even after Barth and C. H. Dodd had recovered some of the depth of biblical thought. The question of the connection between biblical language and near eastern, especially Gnostic, mythology was raised again by the work of Rudolf Bultmann, and in the middle of the 20th century hermeneutics had once more become an overriding preoccupation not only of exegesis but of systematic theology and the philosophy of religion as well. For this preoccupation the theologians were drawing also upon general investigations of the problem of interpretation, among which the hermeneutical theories of Wilhelm Dilthey were perhaps the most profound.

In the theological situation of the 20th century within both Roman Catholicism and Protestantism, the problems of hermeneutics and exegesis were displaying their perennial vitality. Theological systems may come and go in the church, philosophical fashions may change. Issues that dominated one age in church history are relegated to the footnotes in the textbooks of a later epoch. But it is certain that so long as there is a Christian church on earth there will also have to be hermeneutics and exegesis, the theory and practice by which the church and its teachers seek to find and to articulate the message of the word of God in the Bible.

BIBLIOGRAPHY.—*General:* Frederic William Farrar, *History of Interpretation* (reprinted 1961); Robert M. Grant, *The Bible in the Church* (1948); Gerhard Ebeling, *Kirchengeschichte als Geschichte der Auslegung der Heiligen Schrift* (1947); Ernst Fuchs, *Hermeneutik* (1954); R. E. Brown, *The Sensus Plenior of Sacred Scripture* (1955); American Council of Learned Societies, *Five Essays on the Bible* (1960); Joachim Wach, *Das Verstehen,* 3 vol. (1926–33).

Ancient and medieval: Jean Daniélou, *From Shadows to Reality: Studies in the Biblical Typology of the Fathers,* Eng. trans. by Wulstan Hibberd (1960); R. P. C. Hanson, *Allegory and Event* (1959); Robert M. Grant, *The Letter and the Spirit* (1957); Rowan A. Greer, *Theodore of Mopsuestia: Exegete and Theologian* (1961); Beryl Smalley, *The Study of the Bible in the Middle Ages,* 2nd ed. (1952); Henri de Lubac, *Exégèse médiévale* (1959).

Reformation and modern: George Tavard, *Holy Writ or Holy Church* (1959); Gerhard Ebeling, *Evangelische Evangelienauslegung* (1942); Jaroslav Pelikan, *Luther the Expositor* (1959); Emil G. Kraeling, *The Old Testament Since the Reformation* (1955); Albert Schweitzer, *The Quest of the Historical Jesus,* Eng. trans. by W. Montgomery, new ed. (1954); James M. Robinson, *A New Quest of the Historical Jesus* (1959); Jean Guitton, *The Problem of Jesus,* Eng. trans. by A. Gordon Smith (1955); Pius XII, "Divino afflante Spiritu," *Acta Apostolicae Sedis,* XXV, 297–326 (1943). (J. J. PE.)

EXETER, EARLS, DUKES AND MARQUESSES OF, English titles which have been borne at different times by members of the Holand (Holland), Beaufort, Courtenay and Cecil families. The earls of Devon of the family of de Redvers (1141–1262) were also sometimes called earls of Exeter.

The 1st duke of Exeter was JOHN HOLAND (d. 1400), a younger son of Thomas Holand (d. 1360), 1st earl of Kent. John's high station at the English court resulted from the fact that he was half brother to Richard II, his widowed mother having married Edward the Black Prince. John married Elizabeth (d. 1425), a daughter of John of Gaunt, duke of Lancaster, and was constantly in Richard's train until 1385, when his murder of Ralph Stafford temporarily disturbed these friendly relations. He was created earl of Huntingdon in 1388, admiral of the fleet in 1389 and chamberlain of England in 1390, and was again high in the king's favour. In 1397, after assisting Richard in the proceedings against the duke of Gloucester and the earl of Arundel (*see* ENGLISH HISTORY), he was created duke of Exeter. After the accession of his brother-in-law Henry IV, Holand was tried for his share in the death of the duke of Gloucester and was reduced to his earlier rank of earl of Huntingdon. His plots against Henry led to his capture and execution at Pleshey, Essex, on Jan. 9 or 10, 1400. A few months later he was attainted and his titles and lands forfeited.

In 1416 THOMAS BEAUFORT (d. 1426), earl of Dorset, was created duke of Exeter for life (*see* BEAUFORT). About the same time JOHN HOLAND (*c.* 1396–1447), son of the former duke of Exeter, was allowed to take his father's earldom of Huntingdon. Holand assisted Henry V in his conquest of France, fighting both on sea and on land. He was appointed deputy marshal of England (1432), admiral of England (1435) and lieutenant of Aquitaine (1439) under Henry VI. He was one of the king's representatives at the conference of Arras in 1435, and in 1444 was created duke of Exeter. When John died on Aug. 5, 1447, his titles passed to his son HENRY (1430–75), who, although married to Anne (d. 1476), daughter of Richard, duke of York, fought for Henry VI during the Wars of the Roses. Imprisoned by York at Pontefract, he was later released and was present at the battle of Towton (1461), after which he was attainted and his dukedom declared forfeited. He sailed with Henry's queen, Margaret of Anjou, to Flanders in 1463, and was wounded at Barnet in 1471. He died without sons, in Sept. 1475.

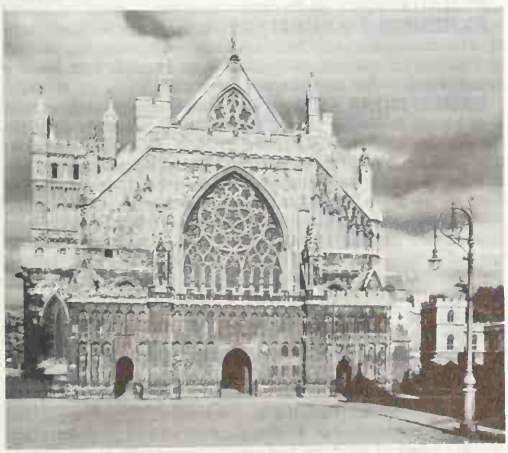
The title of marquess of Exeter was borne by HENRY COURTENAY (*c.* 1498–1539), earl of Devon, who was made a marquess

in 1525. Courtenay was a prominent figure at the court of Henry VIII until Thomas Cromwell rose to power, when his high birth, his great wealth and his independent position made him an object of suspicion. Some slight discontent in the west of England gave the occasion for his arrest, and he was tried and beheaded on Jan. 9, 1539. A few days later he was declared a traitor and his titles were forfeited. Courtenay's only son, EDWARD (1526–56), was imprisoned in 1538 with his parents and only released in 1553, when he was restored to the earldom of Devon. He was a suitor for the hand of Queen Mary. Disappointed, he turned his hopes to Princess Elizabeth and became involved in Sir Thomas Wyat's rebellion. He was exiled in 1555. Edward is sometimes called marquess of Exeter.

The title of earl of Exeter was bestowed upon the Cecils (*see* CECIL) in 1605 when THOMAS (1542–1623), 2nd Lord Burghley, the eldest son of William Cecil, Lord Burghley, was made earl of Exeter by James I. Thomas was a member of parliament during the reign of Queen Elizabeth I, who knighted him in 1575, and he fought under Robert Dudley, earl of Leicester, in the Netherlands in 1585–86. He became president of the council of the north in 1599 and was made a knight of the Garter in 1601. He died on Feb. 8, 1623. His direct descendants continued to bear the title of earl of Exeter and in 1801 HENRY (1754–1804), 10th earl, was advanced to the dignity of marquess of Exeter. DAVID GEORGE BROWNLOW CECIL (1905–), 6th marquess, being his lineal descendant.

EXETER, a city, a county of itself and the county town of Devon, Eng., 168 mi. W.S.W. of London by road. Pop. (1961) 80,321. It lies on the Exe about 10 mi. above its entry into the English channel and derived its early importance from its position at the river crossing. The original town was on the east bank, but the modern city covers both banks. As the cathedral city and county town, Exeter is annually visited by large numbers of tourists. Distributive and service industries employ a large proportion of its inhabitants. Among the manufacturing industries, which are relatively few and small, metal- and leatherworks and the manufacture of paper and agricultural implements are the most important. Exeter suffered air-raid damage in World War II.

Except for its unique twin Norman towers over the north and south transepts, the cathedral of St. Peter is of the Decorated period. Transeptal towers occur elsewhere in England only in the collegiate church of Ottery St. Mary (10 mi. E. of Exeter), for which the cathedral served as a model. The sculptured screen on the west front was added later, probably in the late 14th century. Bishop William Warelwast, nephew of William the Conqueror, started the building of the Norman cathedral which was consecrated in 1133. The present Decorated cathedral of gray limestone was begun around 1275 and finished about 90 years later. In the interior, the long unbroken roof extends through nave and choir and has no central tower or lantern. The elaborately carved bosses and corbels were all refurbished after World War II. A minstrels' gallery projects high up in the north bay of the nave. The window tracery is varied, but each window answers to that on the opposite side of nave or choir; pier answers to pier, aisle to aisle and chapel to chapel, while the transeptal towers complete the balance of parts. The episcopal throne, a sheaf of tabernacle work in wood, was erected by Bishop Walter de Stapeldon and probably made by Thomas de Winton about 1313–17. In the north transept is a large, square clock (1376, with modern works) and a bell, Great Peter, weighing 125 cwt., given by Bishop Peter Courtenay (1478–87). Some important manuscripts, including the famous book of Anglo-Saxon poetry given by Bishop Leofric to his cathedral, are preserved in the chapter house. A complete restoration under Sir George Gilbert Scott was carried out between 1870 and 1877. In 1942 the cathedral was damaged by air raids and St. James's chapel, with the muniment room above, and three bays of the south choir aisle were demolished. The choristers' school, the abbot's lodge and the rear portion of the bishop's palace were also destroyed.



A. F. KERSTING

DECORATED CATHEDRAL OF EXETER, WEST FAÇADE; 14TH CENTURY

Although the destruction of medieval houses and churches in the World War II raids was extensive, many buildings of historic and architectural importance survived. The 14th-century guildhall was rebuilt in 1468–70 but its frontage, carried on arches over the pavement of the main street, was added in 1592. At mid-20th century this guildhall had been longer in continuous civic use than any other building in England. The greater part of the castle, built by William the Conqueror and called Rougemont from the colour of the rock on which it stood, was demolished in 1774 when the assize courts and county offices were built, but the Norman gateway, Aethelstan's tower and the walls were spared. The 1942 bombing exposed many portions of the medieval city walls. Exeter has several churches containing Norman and Saxon work. St. Nicholas' priory, a Benedictine house founded in 1080 and long used as a dwelling house, was restored in 1913. Tuckers' hall, originally the chapel and then the guildhouse of the still-existent Incorporated Company of Tuckers (Fullers), Weavers and Shearmen (incorporated 1489), was built in 1471.

Exeter University college (1922) was reconstituted Exeter university in 1955. The Royal Albert Memorial museum has natural history and archaeological collections and an art gallery.

The river Exe was navigable from the city to the sea until the 13th century when Isabella, countess of Devon, built a weir across the river. In 1564 the citizens built a three-mile ship canal just below the weir—the first in the country to enable seagoing ships to enter an inland port. Later the canal was extended about another two miles to Topsham, at the head of the tidal estuary of the Exe, and an outport was constructed at Exmouth. The port of Exeter is controlled by the corporation.

History.—Exeter was the centre of an early British tribe, the Dumnonii, and was called by the Romans Isca Dumnoniorum. Mosaic pavements, potsherds, coins from the eastern Mediterranean and other relics have been found, as well as portions of the Roman walls. The place is said to have been the Caer Wisc of the Britons, who were driven out by Aethelstan in 926. Its importance as a stronghold is shown by the great earthworks which the Britons threw up to defend it, on the site of which the castle was afterward built, and by the number of roads which converge upon it.

Exeter sustained a number of sieges because of its position as the chief town in the southwest of England. In 877 a Danish force seized it but was blockaded by Alfred the Great and forced to withdraw. About 17 years later the Danes besieged the town again but were once more defeated by Alfred. In 1001 it was again unsuccessfully besieged by the Danes. In the following year, however, it was given by King Aethelred II to Queen Emma; she appointed as reeve Hugh, a Frenchman, by whose treachery it was taken and destroyed by King Sweyn I of Denmark in 1003. In 1050 the Saxon church became a cathedral and Bishop Leofric the first bishop of Exeter, when the united sees of Devon and Cornwall were transferred from Crediton. In 1068, after a siege

of 18 days, Exeter surrendered to the Conqueror. Again in 1137 the town was held for Matilda and Baldwin de Redvers against King Stephen for three months and surrendered at last because of lack of water. Three times subsequently Exeter held out successfully for the king—in 1467 against the Yorkists, in 1497 against Perkin Warbeck and for six weeks in 1549 when the men of Cornwall and Devon rose against the use of the new Book of Common Prayer and other religious innovations.

During the Civil War the city declared for parliament, which garrisoned it in 1642, but in 1643 it was taken by the royalists who held it until 1646. It was one of the places where, in 1688, William of Orange stopped on his first march to London. In 1950 King George VI and his consort, Queen Elizabeth, visited Exeter on the 900th anniversary of the founding of the see.

Exeter was a borough some time before the Conquest, since the burgesses are mentioned in Domesday Book. Its first charter, granted by Henry II, gave the burgesses all the free customs which the citizens of London enjoyed and was confirmed and enlarged by most of the succeeding kings. By 1227 government by a reeve had given place to that by a major and four bailiffs, which continued until the Municipal Reform act of 1835. In 1537 Exeter was created a county of itself by charter. It was represented in parliament by two members from 1295 to 1885, when the number was reduced to one. Numerous trade guilds were incorporated in Exeter, one of the first being the tailors' guild, incorporated in 1466. By 1482 this had become so powerful that it interfered with the government of the town and was dissolved on the petition of the burgesses. Another powerful guild was that of the Merchant Adventurers, incorporated in 1559, which is said to have dictated laws to which the mayor and bailiffs submitted.

Exeter was formerly noted for the manufacture of woolen goods, introduced in Elizabeth I's reign. At the time of the trade's greatest prosperity it was surpassed in value and importance only by that of Leeds.

EXETER, an incorporated town (township) in Rockingham county in southeastern New Hampshire, U.S., 12 mi. S.W. of Portsmouth. It was settled in 1638 by the Rev. John Wheelwright and a group of associates who had been exiled from Massachusetts for harbouring heretical, antinomian doctrines. During its first years the town existed as a commonwealth independent of the English colonies, but an increasingly unfavourable economic situation caused Exeter voluntarily to submit to the jurisdiction of Massachusetts in 1643. Later in the 17th century the area became a part of the colony of New Hampshire. Industry came with the establishment of a sawmill at the falls of the Exeter river, which passes through the unincorporated village of Exeter. Indian attacks inhibited growth for 50 years after 1675, but following the removal of this danger and the beginning of shipbuilding, development accelerated. Exeter was the provincial capital during the period of the American Revolution. Manufacturing grew throughout the 19th century; textiles, shoes and a brassworks continued to constitute the main local industries.

Phillips Exeter academy, founded in 1781, became one of the nation's leading preparatory schools for boys. The academy buildings, constructed in colonial style, are a dominant feature of the town, and many private homes built in the 18th and early 19th centuries further emphasize this architectural theme. The town was the birthplace of Nicholas Gilman, signer of the constitution; Lewis Cass, U.S. secretary of state under Pres. James Buchanan; and the sculptor Daniel Chester French. Exeter has a town-manager form of government, in effect since 1956, and a board of selectmen. For comparative population figures *see* table in NEW HAMPSHIRE: *Population.* (R. C. GI.)

EXETER BOOK, an anthology of Anglo-Saxon poetry in Exeter Cathedral library, which can be identified with the "large English book on various subjects composed in verse" mentioned in the list of donations made to Exeter cathedral by Leofric, who became bishop of Crediton in 1046 and in 1050 moved the see to Exeter where he died in 1072. The first seven folios, containing a late 11th-century version of this list and a number of 12th-century manumissions and other legal transactions, originally formed part of another manuscript also given to the cathedral library by Leofric, the Old English gospels, now in Cambridge University library (MS. Ii 2, 11), but had been bound with the *Exeter Book* by the latter part of the 16th century.

The poetic codex is a handsome volume, though without decoration, written in a beautiful hand of the later 10th century. It is incomplete at the beginning, and perhaps at the end, and at least seven folios are missing from the interior of the volume. Even so, it remains the largest extant collection of Old English poetry, and much the most varied. It begins with some long religious poems, the three parts of *Christ,* the two *Guthlac* poems, *Azarias,* the *Phoenix* and *Juliana,* and is then followed by a number of shorter pieces, religious and secular, in no discernible arrangement.

The religious poems include an *Address of the Soul to the Body,* a *Judgment Day, The Descent into Hell,* catalogue poems on *The Gifts of Men, The Fortunes of Men* and *The Wonders of Creation,* parts of a *Bestiary, The Exile's Prayer* (or *Resignation*) and a few smaller pieces. There are also poems of types whose existence would have been unsuspected if this codex had not survived. All the extant Anglo-Saxon lyrics, or elegies—*The Wanderer, The Seafarer, The Wife's Complaint, The Husband's Message, The Ruin*—are found here, and there are 95 *Riddles,* a genre that would otherwise have been represented only by the so-called *Leiden Riddle,* preserved in a manuscript in the University library at Leiden.

The Rhyming Poem is the only example of its kind, and the *Gnomic Verses* differ in type from those in Cotton Tiberius B.I. *Widsith,* a catalogue of heroes and nations, and *Deor* and the so-called *First Riddle,* which use heroic material and are the only Old English poems to have a refrain, make up the tale. Since any attempt at logical arrangement breaks down after the first few items, and the rest of the book seems to consist of items added in haphazard fashion as they came to hand, it has been plausibly suggested that the scribe who wrote this beautiful manuscript so evenly was copying an earlier collection. For comment on the individual poems in the *Exeter Book, see* the article ENGLISH LITERATURE.

BIBLIOGRAPHY.—A transcript, made in 1831–32 by Robert Chambers, is Additional MS. 9,067 in the British museum, London, and is still useful for its rendering of passages which have deteriorated since that date. There is a facsimile, *The Exeter Book of Old English Poetry,* with introductory chapters by R. W. Chambers, Max Förster and Robin Flower (1933), where all aspects of the subject are treated. The first complete edition is that of B. Thorpe, *Codex Exoniensis* (1842), and later editions include that of the Early English Text Society, *The Exeter Book,* pt. i ed. by I. Gollancz (1895), pt. ii ed. by W. S. Mackie (1934), which includes a translation; and *The Exeter Book,* ed. by G. P. Krapp and E. van K. Dobbie (1936), with full bibliographies. *See* also K. Sisam, "The Exeter Book," in *Studies in the History of Old English Literature* (1953). (D. WK.)

EXHAUSTION: *see* FATIGUE; VACUUM.

EXHIBITION ARCHITECTURE, refers to the design of structures specifically built to house exhibitions, expositions and fairs, in Europe and the United States, from the 18th century to the present. *See* EXHIBITIONS AND FAIRS.

EXHIBITIONS AND FAIRS. An exhibition or exposition is an organized display of works of art, science and industry for the purpose of stimulating public interest, promoting manufactures, expanding trade and illustrating the progress and accomplishments of one or a wide variety of productive activities. Exhibitions are normally held at irregular intervals for periods lasting several months or more, and no distinction is made between business visitors and the general public. Visitors attracted to this type of display are not primarily interested in buying anything but attend because of the exhibition's educational and entertainment value. The main commercial value is found in the publicity it affords the products on exhibition.

A fair is a temporary market where buyers and sellers gather to transact business. It is organized at regular intervals, generally at the same location and period of the year, and usually lasts for several weeks. The primary function of a fair is the promotion of trade. Historically, the trade fair or show has followed the horizontal line of organization with displays of many different kinds of products and services in specific commodity or industry groupings. The vertical type of fair, in which participa-

tion is confined to exhibitors representing one industry or even specialized segments of an industry, gained in popularity during the first half of the 20th century. It became the more common form in the United States and also won some popularity in other countries. In addition to the broad-range, classic type of international trade fair, there are many fairs that are important for international trade because they portray the total industrial offerings of one country. Prominent examples of this type of trade fair are the Swiss Industries fair, German Industries exhibition and Pakistan International fair. Among the more notable examples was the British Industries fair held for many years at London and Birmingham. However, the growing popularity of specialized exhibitions in England led to the dissolution of the London section of this fair in 1956 and Birmingham in 1957. Exhibitor participation in such national fairs is usually restricted to organizations and individuals established in business in the country concerned.

The terms exhibition, fair and exposition are used almost interchangeably in the United States. In Great Britain they have retained more exact meanings, fair being employed for those institutions primarily devoted to trade or commerce, such as local fairs and industrial fairs, and exhibition being reserved for those of national or international scope. In France the usage is even more precise: *foire* ("fair") denotes a display of products of industry, predominantly commercial in scope, and *exposition* ("exposition") denotes an international display. Thus, a world's fair in Great Britain is referred to as an "international exhibition" and in France as an "*exposition internationale*."

EARLY HISTORY OF FAIRS

Historically, fairs were created to solve the early problems of distribution. They offered an opportunity for the demonstration of skills and crafts, for the exchange of ideas and for bartering of goods. They concentrated supply and demand in certain places at certain times. They developed along tribal and cultural boundaries, frequently in a burial area because of the belief that visiting merchants would be less prone to belligerence in a cemetery.

It is generally conceded that the word fair is derived from the Latin *feriae*, meaning "feasts." This derivation has its basis in the fact that many of the earliest fairs were held during the festivals of the warm season or at the beginning of harvest. The Romans introduced markets and fairs into northern Europe to encourage trade with their conquered provinces and to spread political propaganda. When the Roman empire disintegrated late in the 5th century virtually all commerce ceased until the 7th century. Commercial interchange began to grow under the aegis and encouragement of Charlemagne, who promoted the holding of markets everywhere in his domain. However, trade was then largely controlled by Arabs, Jews, Saxons and other foreigners. Exchanges of goods were made in or near churches and were guaranteed protection by the peace of God or by the King's peace.

Fairs evolved from some of the local markets, particularly at points of major caravan route intersections and wherever people congregated for their religious festivals. A time and place were fixed where people of different and even conflicting nations and cultures could come together and trade in peace. Eventually, standards of measure, weight and quality were adopted; wholesale distribution practices developed; standards of value for mediums of exchange were established; and even stock corporations were created. The methods of commerce introduced at fairs became widespread and rules of the fair eventually became the basis of business law.

As commerce began to flourish the great fairs became increasingly important. Soon after it was chartered in 629, the fair of St. Denis near Paris rose to prominence. The Easter fairs at Cologne were equally popular events in the 11th century. The fair of St. Demetrius in Thessalonica flourished in the Byzantine empire until the eve of the Ottoman Turkish conquest in the middle of the 15th century. Antioch and Trebizond also held Byzantine fairs.

For several hundred years, beginning in the middle of the 12th century, the fairs of Champagne reigned supreme among the important fairs of Europe. At these fairs were found the products of all Europe and beyond: furs from Russia; drugs and spices from the orient; cloth from Flanders, and later from England; linens from south Germany. The several annual fairs of Champagne included two at Troyes (from which the term "troy weights" originated) and the others were located at Lagny, Provins, Brie and Bar-sur-Aube.

Because of the higher tolls imposed upon traders and merchants by the French kings and the disturbance from political upheavals, the fairs of Champagne began to lose their importance sometime in the 14th century. Thereafter the important fairs took place in Flanders, largely because of the inauguration of commercial sea routes. Part of the commerce previously centred at Champagne found its way to the fairs at Lyons, France, and Geneva, Switz. German fairs have an origin almost as ancient as those of France. England, during the middle ages, was on the periphery of the trade area and its commerce was not highly developed. The largest English fair was the famous event held several times each year at Stourbridge. Among other fairs of historical interest were those at Kinsai in the Chinese province of Manza that flourished during Europe's dark ages, the great Aztec fair which the Spaniards found on the present site of Mexico City and the Nijni-Novgorod fair in central Russia. Wherever they were held these fairs were in many ways alike. They succeeded because they were located on the great highways of communication and in some cases on the routes of holy pilgrimages.

In every civilization, as commerce became standardized, transportation organized and cities large enough to require daily markets, fairs became less important. Sometimes they resulted in religious festivals, as many Roman fairs did; others became places for amusement or pleasure, as did the English fairs at Stourbridge, Oxford and Stratford-upon-Avon. Often they were ruined by rulers grasping for revenue; some were abolished or forced to move sites elsewhere. The Soviet Union abolished all of its 17,500 fairs in 1930 as being "unsuited to the Soviet economic and political objective."

The commerce of western Europe during the feudal era was largely based on fairs, but with the coming of the Industrial Revolution the vitality of the great fairs was sapped, many perished and the character of others changed. While this was happening, the increase of population gave rise to a greater interest in agriculture and food supply. Societies to promote improved breeding of livestock and plant production were organized, particularly in England. They sponsored fairs to stimulate production, promote the sale of improved breeding stock and to educate the patrons.

By the middle of the 19th century the specialty fair had run its course of usefulness. Its one remaining vestige in the 20th century is the county fair and the local livestock show, still popular in the United States and Europe. From the specialty fair finally evolved the sample fair at which manufacturers display their products in small booths. The Leipzig Trade fair and later the Milan Samples fair and Paris International fair are the best examples of this evolutionary process. The Italian and French fairs are the largest trade fairs in the world.

EUROPEAN EXHIBITIONS

The first exhibitions, in the modern sense of the word, were fine art exhibitions. They appeared in Rome in the beginning of the 17th century. In 1662 the Royal Academy of Painting and Sculpture at Paris presented its first art exhibition. For many years thereafter such exhibitions were held in the court and gallery of the Palais Royal. The first "general" exposition was held in Sept. 1699 in the gallery of the Louvre at Paris; although its theme was primarily artistic, it also included products of manufacture. For this reason some historical accounts credit the French with having originated industrial exhibitions. Actually, an industrial exhibition was opened at London by the Royal Society for the Encouragement of Arts, Manufactures and Commerce (Royal Society of Arts) as early as 1761. The first French industrial exhibition was held in 1791 on the Champ-de-Mars in Paris, the scene of French national celebrations after the fall of the monarchy. It was housed in an arcuated gallery encompassing a large quadrangle reminiscent of a market square. During the

following half-century similar exhibitions were presented at Manchester, Leeds, Birmingham, Dublin, Ghent, Berlin and Vienna. These fairs came into being as a result of the rise of mechanized industry, and in turn had an enormous effect upon the arts and manufactures of their respective countries.

England.—The best-managed and most thoroughly organized exhibitions in England were those of the Cornish Polytechnic society held for a number of years through 1850. Other English cities such as Manchester, Birmingham and Liverpool also held local expositions of substantial importance. Shortly after the French exposition of 1844 the Royal Society of Arts held annual exhibitions in its own hall, the Adelphi. These exhibitions may be considered to be the forerunners of the Crystal Palace exhibition of 1851.

In 1849, Prince Albert, husband of Queen Victoria and president of the Royal Society of Arts, became thoroughly imbued with the idea of inviting exhibitors from all civilized nations to participate in an exposition. Plans were developed and the necessary funds speedily raised with Queen Victoria heading the list of subscribers. The exhibition opened in the Crystal palace in Hyde park, London, on May 1, 1851.

The Crystal palace, designed by Sir Joseph Paxton, was a remarkable construction of prefabricated parts, a long rectangular stepped form intersected in the middle by a vaulted transept, consisting of an intricate network of slender iron rods sustaining walls of clear glass. It was 1,851 ft. long, 456 ft. wide, 66 ft. high and contained a floor area of more than 800,000 sq.ft. On the ground floor and galleries there were over 8 mi. of display tables. The number of exhibitors was over 13,000, nearly half of whom were non-British. France sent 1,760 exhibits and the United States 560. Among the U.S. exhibits were false teeth, artificial legs, Colt's repeating pistol, Goodyear india rubber goods, chewing tobacco, trotting sulkies and McCormick's reaper. More than 6,000,000 visitors attended the exhibition and it was a financial success, closing with a $750,000 surplus. In architecture, the Crystal palace established a standard for later international fairs, likewise housed in glass conservatories, the immediate successors being the Cork exhibition of 1852, the Dublin and New York expositions of 1853 and the Munich exhibition of 1854. Queen Victoria, when describing the opening of the exhibition as "the greatest day in our history," little realized then that it was to be the first of a long line of similar international exhibitions.

Prince Albert attempted to revive the spirit of the Crystal Palace exhibition at South Kensington in 1862, but was unsuccessful. In 1863 an exhibition was held for national Turkish manufacturers and for foreign implements and machinery. The display of jewelry from the imperial Seraglio made it something more than a mere local celebration. An exposition of Dutch industry was held in Amsterdam in 1864, and during the same year industrial shows were held at Malta, in India at Calcutta and Lucknow, and a combined French and Spanish exposition at Bayonne, France. In 1865 several exhibitions, although chiefly agricultural in character, were held in Germany, the Netherlands and Belgium. During this period Europe became so satiated with exhibitions and fairs that sponsors themselves became discouraged. The aim of most of those exhibitions was not so much to promote business directly but to show technical developments and stimulate manufacturers to compete with each other. This was, in the main, encouraged by competitive awards and prizes.

From 1862 to 1923 no significant exhibition was held in England. Then in 1924 the British Empire exhibition was held in Wembley, but poor management and adverse economic circumstances resulted in a substantial deficit. The Festival of Britain of 1951 was devoted to demonstrating the solidarity and industrial might of the British empire and its achievements since 1851. The Royal Festival hall was an outstanding example of architecture and acoustical engineering. As a result of economic reverses and site problems, original plans to build an exhibition of record size were abandoned. Instead many small exhibitions were created, each linked to the key site on the south bank of the Thames river. The exhibition was national in scope and included demonstrations of scientific achievements at South Kensington; the pleasure gardens and the fairgrounds in Battersea; and exhibits in popular modern architecture and a heavy engineering section at Glasgow, Scot.

France.—After the outstanding success of the Crystal Palace exhibition of 1851, Napoleon III gave increased impetus to French expositions by resolving to present to the world an even greater and more brilliant spectacle. Several expositions were held, beginning with the Paris exposition of 1855, which gave birth to the great Palais de l'Industrie constructed on the Champs Élysées. The "cathedral of commerce," as contemporaries called it, represented a more complete realization of metallic construction than did the Crystal palace. It was a rectangular structure with a glass central nave about 150 ft. wide—over twice that of the Crystal palace—surrounded on four sides by a double row of low aisles. The glass vault of the nave rested on a series of arched wrought-iron lattice girders, the outward thrust of which was checked by lead-block abutments. Although showing only glass from the inside, the Palais de l'Industrie was encased in stone façades treated in monumental architectural style. This extraneous grandiosity served as the model for the exhibition buildings at the later fairs in London (1862) and Chicago (1893). Queen Victoria's visit to Paris did much to strengthen the ties of friendship between the two nations.

Fairs and exhibitions of all types exercise an influence that is immediate, continuous and far-reaching. The French expositions demonstrated the rapid progress of French industry, which had for many years lagged behind that of Great Britain. The results of these expositions did much to aid the transformation of a predominantly agricultural country into an industrial one. They exerted a great influence to break down the isolationist and protectionist philosophies among the industrialists of France.

The impressive concentric oval structures used for the Paris exposition were again the scene of the great universal exposition in 1867, followed by the exposition of 1878, which came seven years after the disastrous Franco-German War. It was held on an enlarged site covering approximately 66 ac. in the centre of Paris. Until then the largest exposition in world's fair history, it boasted 52,835 exhibits and drew an attendance of more than 16,000,000 visitors.

The main building of the 1867 exhibition, erected by Krantz and Alexandre Eiffel, was an ellipse measuring 1,608 by 1,266 ft., divided into seven concentric galleries centred on an oblong open court; the outermost ring (Galerie des Machines), by far the largest (height 25 m. [82 ft.], width 35 m. [115 ft.]), was without structural clutter. Numerous small pavilions and towers were scattered about the principal building.

The Paris exposition of 1878 was divided into two sections. One was housed in temporary structures on the old site, the main hall a rectangular composition of domed and vaulted sheet-metal forms connected by lower glass-walled wings. A series of small galleries was flanked by the Halle des Machines, in height, width and construction about equal to that of the 1867 fair, though making use of girders of the improved De Dion type. The other part of this exhibition was located in the Trocadéro, on the far side of the Seine.

The exhibition of 1889, centred around the Eiffel tower, had two long galleries devoted to the fine and liberal arts connected by a third given over to general exhibits, behind which rose the imposing Palais des Machines. This last, designed by Ferdinand Dutert and constructed by M. Contamin, exceeded in size any unobstructed interior built up to that time, its 20 main trusses covering a length of 1,270 ft., each spanning about 480 ft. The trusses formed a pointed arch and curved down to a narrow wedge shape at the base, transmitting their weight through a hinged joint to the ground.

Belgium.—During the 20th century the Belgian expositions, one at Liège in 1930 and the others at Brussels in 1935 and 1958, were splendid examples of the use of exhibitions as vehicles of city planning. The main object of the 1958 Brussels Universal exhibition was to present a synthesis of views on "humanism." This world's fair was held on a 500-ac. site in Heysel park about four miles from the heart of the city of Brussels. It lasted for

six months and was a universal exhibition in which 45 nations participated. The Belgian government is reported to have subsidized approximately half the estimated cost of $400,000,000. It was the most expensive undertaking of its kind in history. Attendance totaled 41,454,412 visitors, a record for European fairs.

At the Brussels fair of 1958 Edward D. Stone erected for the United States exhibition building the world's largest free-span circular structure (diameter 350 ft.), having a system of slender posts and cables under tremendous tension supporting the outer plastic roof. Gilding on these sustaining members and traceried screens enriched the great airy pavilion. Eleven indigenous willow trees left undisturbed inside the building complemented the formalized planting on the paved plaza with its reflecting pool outside.

Diplomatic Convention on World Fairs.—Since the first world exhibition at the Crystal palace in 1851 about 30 major world fairs have been held. Until 1928 any country could organize an activity of this kind whenever it wished to do so. Many were spectacular and significant; others were more like expanded national fairs than universal exhibitions. To remedy this situation, 35 countries signed a diplomatic convention in Paris on Nov. 22, 1928, to regulate the frequency and method of organizing world fairs. The convention later established the Bureau International des Expositions at Paris to deal with the recognition, patronage and authorization of international exhibitions, which were divided into two classes, general and special. A general exhibition includes the products of more than one branch of human activity, or is organized to demonstrate the progress achieved in the whole of a given sphere of activity. A special exhibition deals with only one applied science or activity.

Using another means of classification, exhibitions fall into two categories: in the first, the countries invited to participate are obliged to construct national pavilions; in the second category, countries are not so obliged. For the organization of international exhibitions the world is divided into three zones: Europe, North and South America, and the remainder of the world. In the same country not more than one first-category exhibition may be held during any period of 15 years, and an interval of 10 years must elapse between two general exhibitions of either category. There must be an interval of at least six years between two first-category exhibitions in the same zone or at least two years between exhibitions in different zones. For second-category exhibitions the intervals are two years and one year respectively, and for special exhibitions the intervals are still shorter.

Two international exhibitions in the United States, the Century 21 exposition at Seattle, Wash., 1962, and the New York World's fair, 1964–65, felt the effects of the bureau's power. The former was recognized and approved as a second-category universal general exhibition but the latter was not sanctioned as a universal exhibition. The Seattle fair, contained in an area of 74 ac., for the most part emphasized technical and scientific achievements of the space age. The New York fair, with over 600 ac. of ground space, provided for a more general form of exhibition. A new methodology in the organization of world fairs was introduced in the New York World's fair by Robert Moses, its president. Unlike its predecessors, an exhibit space fee was imposed for national and international participants, who were also required to bear the cost of construction and maintenance of buildings. U.S. world fairs are organized and operated generally by a nonprofit membership corporation which is not directly associated with any governmental agency; financing is, for the most part, accomplished by selling debenture bonds to the public. European world fairs on the other hand are, for the most part, subsidized by the host nation.

UNITED STATES

Most American international expositions, unlike their European counterparts, have searched for a theme based upon the celebration of some historical event of importance or upon the measurement of past progress.

Crystal Palace, 1853.—The history of international expositions in the United States started inauspiciously with the so-called Crystal Palace exposition held in New York city in 1853. Born of a desire to recreate and, if possible, surpass the British Crystal Palace exhibition of 1851, it was promoted by New York businessmen, including Horace Greeley. A Crystal palace modeled after London's impressive structure was somewhat imperfectly executed by local architects as the main feature of the 13 ac. of Reservoir square (now the site of Bryant park behind the New York Public library) on which the exposition was held. The New York Crystal palace was an octagonal form, with Greek-cross superstructure abutting a round dome at the centre and tall minarets at the outer angles. The glass and iron building was erected by Georg Carstensen and Charles Gildemeister for the Association for the Exhibition of the Industries of All Nations, a private concern. The building housed 4,854 exhibitors, of whom about half came from 23 foreign nations. A block north stood the 350-ft.-high wood and iron companion structure, the Latting observatory. Both were burned in the 1850s.

The roof of the ill-fated Crystal palace leaked and heavy rains ruined exhibits and soaked visitors. Delays, poor management and dwindling attendance caused the exposition to close with a substantial loss. An attempt was made to revive the exposition in 1854 under P. T. Barnum. But it had already been overwhelmed by debts and delays and he concluded, after heavy investments of time and money, that "the dead cannot be raised." Though a failure, the New York event did give impetus to U.S. manufacturers by stimulating competition with foreign producers.

There have been other more successful international expositions in the United States since the Crystal Palace exhibition of 1853: the Centennial exposition, Philadelphia, Pa., 1876; the World's Columbian exposition, Chicago, Ill., 1893; the Louisiana Purchase exposition, St. Louis, Mo., 1904; the Panama-Pacific International exposition, San Francisco, Calif., 1915; A Century of Progress exposition at Chicago, 1933–34; the New York World's fair at New York, 1939–40; and the Century 21 exposition at Seattle in 1962.

Philadelphia Centennial, 1876.—As the close of the first century of the independence of the United States approached, it was generally considered desirable to celebrate the centennial of the signing of the Declaration of Independence in a manner worthy of the occasion. The eastern cities (Boston, New York, Baltimore, Washington and Philadelphia) competed for the honour of being host to the celebration. Philadelphia, as the birthplace of the Declaration, was deemed the most appropriate choice and on May 10, 1876, the United States held its first major international exposition on a 236-ac. site in Fairmount park. This "Mighty Cosmos," as it was called, consisted of 167 buildings that together housed over 30,000 exhibitors. Thirty-five foreign governments accepted the invitation to participate issued by the secretary of state on behalf of the president. It was at first planned to erect a fine arts hall and another large exhibition building covering 44 ac., of the Crystal palace type, the largest of which up to this time was the 30-ac. ovoid built for the Paris exposition of 1867. Upon further study, however, the Philadelphia Centennial commission decided to construct separate buildings for the various types of exhibits. The largest structures were the Main Exhibition building and Machinery hall (the first slightly greater and the second somewhat smaller than the London Crystal palace), which flanked the entrance plaza. Next in size was Agricultural hall, a sort of "cardboard-gothic" conservatory, placed on the diagonal a few hundred yards north of the first two. Midway between was the glass Horticultural hall, and the cruciform Ladies' pavilion and U.S. Government building. The art collection was in Memorial hall, on an axis with the main building. Judges' hall faced the entrance plaza; and there were, in addition, smaller foreign and state pavilions, most of them rather ornate in design, and the severely simple Pomological building next to the agricultural conservatory. Memorial hall and Horticultural hall were left standing after the fair.

More than 8,000,000 visitors saw exhibits that included Alexander Graham Bell's telephone, the Westinghouse air brake, Edison's duplex telegraph, the typewriter and the sewing machine. Americans coming to the Centennial exposition with thoughts of the old Liberty bell and august patriots found instead a new era of in-

ventive ferment and the birth of the machine age.

The Philadelphia Centennial exposition established new precedents with its Ladies' pavilion, a federal government building, a royal visitor and a record attendance. But it was doomed to financial failure when repayment of a $1,500,000 grant from the federal government, guaranteed by Philadelphia businessmen, was demanded by congress before stockholders could be reimbursed for the amount of their subscriptions. In all other respects the Centennial exposition of 1876 was considered to be a success. It promoted unification of the nation after the Civil War and stimulated the social and industrial consciousness of the country.

New Orleans, 1884–85.—At New Orleans, La., in 1884–85 was held the World's Industrial and Cotton Centennial exposition, notable for having the most extensive buildings under one continuous roof erected up to that time. The main building, measuring 1,378 by 905 ft., covered 33 ac. Horticultural hall (600 by 194 ft.) was the world's largest greenhouse.

Chicago Columbian, 1893.—The success of the Centennial exposition of 1876 spurred commemorative activities by other cities in the United States. The next major international exposition celebrated the 400th anniversary of the discovery of America by Columbus. There was spirited competition for this exposition among the nation's leading cities. Chicago was chosen in part because it was a railroad centre and in part because it offered a guarantee of $10,000,000.

Continuing the precedent set at Philadelphia of creating a vast gardened layout containing numerous separate buildings rather than a single great hall, the World's Columbian exposition of 1893 was planned to spread over 686 ac., part of which is now Jackson park in Chicago, fronting Lake Michigan. The chief planner was the Chicago architect Daniel H. Burnham; Charles B. Atwood was designer-in-chief; and Olmstead Co. was entrusted with the landscaping. The buildings of the World's Columbian exposition of 1893 were impressive by virtue of their classic façades having a uniform cornice height of 60 ft. The plaster palace fronts bore little functional relationship to the exhibition halls inside; but the grandeur of "The White City," electrically lighted at night, led to a resurgent interest in classical architecture throughout the country.

Behind the calm pillared façades and classical porticoes of the great "White City" the visitor found unexpected excitement and novelty, the first Ferris wheel and a dazzling new wonder: electricity. Electricity had been introduced and exploited at the Paris exposition of 1889, but in 1893 it was still unfamiliar to most Americans. The exposition was opened by a dramatic act when Pres. Grover Cleveland pushed a "magic" button in the White House and set the great Allis engine in motion in Chicago, turning on the power of the exposition. The engine, the dynamo and the alternating-current generator displayed for the first time by George Westinghouse later became the basic tools of the electric power industry.

The Columbian exposition gross outlays amounted to $28,340,700, of which $18,678,000 was spent on grounds and buildings. More than 21,400,000 visitors attended the exposition and the cash balance remaining at closing was $446,832, making it the first U.S. international exposition to close with a profit. The Palace of Fine Arts, a 600,000-sq.ft. building, was rebuilt in permanent limestone in 1928–32 to house the public exhibition of the Museum of Science and Industry.

During the years immediately following Chicago's Columbian exposition many other smaller expositions were held in the United States: San Francisco, 1894; Portland, Ore., and Atlanta, Ga., 1895; Nashville, Tenn., 1897; and Omaha, Neb., 1898.

Buffalo Pan-American, 1901.—The Pan-American exposition held at Buffalo in 1901 differed from its predecessors in not commemorating a historical event. Its purpose was "to promote commercial and social interests among the States and countries of the Western Hemisphere." It was prompted by the Pan-American conference held in Washington, D.C., 1889–90. "The Landscape City," as the exposition was popularly called, boasted as its dominating theme a 375-ft. electric tower surmounted by a statue of the goddess of light—studded with incandescent lamps and powerful searchlights. The waters of Niagara falls, which supplied the power for the exposition's electricity, were illuminated by varicoloured searchlights. For the first time visitors were afforded an opportunity to examine the wireless telegraph system and Thomas A. Edison's new storage battery. The closing month of the exposition was marred by the assassination of President McKinley during a reception at the fair's Temple of Music.

Louisiana Purchase, 1904.—In 1904 the city of St. Louis inaugurated its Louisiana Purchase exposition to commemorate the 100th anniversary of the acquisition of the Louisiana territory from France. Taking as their goal the sum of $15,000,000, which the United States had paid for the territory in 1803, the promoters of the exposition succeeded in raising the sum in equal amounts from the federal government, the city of St. Louis and popular subscription. Congressional appropriation of $5,000,000 for this purpose was an unprecedented legislative act.

Education was declared to be the keynote of the St. Louis exposition, and it was advertised that processes, not things, were on display. The crowning feature of the fair was the display of 100 automobiles, the first sizable exhibition of automobiles ever held anywhere. Most exhibits showed some form of application of science in everyday life. In the New York state building, for example, the cooking was done entirely by electricity. There were chemical exhibits from Germany, thousands of articles from Japan and an immense Philippine reservation to show Americans a part of the world they had recently acquired.

The exposition spread over 1,240 ac. of Forest park in St. Louis and sought bigness at the expense of harmony. The fan-shaped area, largest ever utilized by any exposition, was laid out with cascades, basins, gardened terraces, promenades and drives, radiating from the domed Festival hall, which stood in front of the permanent Fine Arts building. The main halls were wedge-shaped; architecturally they reflected the influence of the World's Columbian exposition. Most of the foreign exhibitors were located in the northwest corner of the park; the states' pavilions were in the southeast corner and bordering the south side of the fan layout.

The most striking feature of the exposition was its demonstration of the progress that had been made in technological production in the 11 years since the Chicago exposition. Culturally and aesthetically its influence was slight, and it was a dismal financial failure.

Panama-Pacific, 1915.—The Panama-Pacific International exposition of 1915 in San Francisco celebrated a contemporary event—the opening of the Panama canal. It was the first international exposition held in the United States that did not stress industry or science. Its great contributions were of a cultural nature. Despite the fact that it opened just after the outbreak of World War I, the exposition enjoyed a moderate success with a small attendance. It closed with a surplus of over $2,000,000 with which a memorial exposition auditorium was financed. It was the first exposition at which the public could take airplane rides and the first to use motion pictures. Architecturally, the classical overtones of Chicago were still in evidence; but they were enhanced by an added note of gaiety achieved through the use of colour and special illumination. One of Burnham's later associates, Edward H. Bennett, engaged as architect of the grounds, produced a novel plan, making use of a series of courts, which trapped the sunlight and sheltered visitors from the winds that blow across the northern tip of the San Francisco peninsula in midsummer. A Mediterranean theme combined mainly Moorish-Spanish, Romanesque and Italian Renaissance styles. Ruins of Bernard Ralph Maybeck's crescent-shaped Palace of Fine Arts, embracing an independent domed octagon surrounded by water, mark the site in Presidio park.

Buildings of the Panama-California exposition held at San Diego, Calif., the same year were all designed by Bertram G. Goodhue in the Spanish Renaissance manner as employed in the colonial southwest. The buildings in Balboa park were reconditioned and used for a second exposition in 1935–36, with a few temporary additions, in simple modern style, and afterward served as city library and university.

Century of Progress, 1933–34.—In 1933–34 Chicago was again

the site of a great international centennial exposition, the central theme of which was "the dramatization of the progress of civilization during the hundred years of Chicago's existence." Appropriately named A Century of Progress International exposition, it focused on the great scientific, technological and engineering developments of the century. Held on 424 ac. reclaimed from the lake since the Columbian exposition of 1893, it had an 8-ac. Hall of Science containing exhibits which illustrated contemporary thought in the pure sciences. Building developments such as prefabricated materials, uniform lighting and air conditioning were utilized at the exposition.

An entirely new concept determined the architecture of A Century of Progress, based upon a forthright functionalism that depended for its effectiveness upon plain, geometric forms and a bold use of pure colour. The design took into account the primary need, that of housing suitably and attractively an exhibition reflecting the scientific and industrial progress of the past century, and considering the movement of crowds of people through the grounds and buildings with comfort and safety. Much of the construction was prefabricated.

Unlike most of its predecessors, the Century of Progress exposition received no government subsidies. Despite the fact that it was held during the depths of the depression, receipts yielded a great financial profit. One factor that contributed to this monetary success was the adoption of a new idea in American expositional procedure—the advance sale of tickets. The managers correctly calculated that such an advance sale would stimulate a substantial early attendance and thus promote the success of the exposition. Another expedient constituted a radical departure from traditional exposition practice but proved extremely effective. This was the selling of exhibit space to exhibitors in place of giving it to them free of charge as all previous fairs had done. Ever since the Columbian exposition big business had grown increasingly dependent on expositions as means of advertising. This dependency was profitably exploited by the 1933 fair to the extent of at least $3,000,000. At the final tally of the books, A Century of Progress had earned the distinction of being the only major world's fair to pay all its debts, reimburse all its bondholders in full, plus 6% interest, and have a surplus in cash at its closing.

While A Century of Progress was being held at Chicago, the material benefits accruing to that city and the enhancement of its prestige did not escape the attention of civic-minded citizens in other metropolitan centres. There can be little question that the California-Pacific exposition in San Diego in 1935, the Texas centennial in Dallas in 1936, and the Great Lakes exposition in Cleveland, Ohio, in 1937, were directly inspired by the example of Chicago.

New York World's Fair, 1939–40.—Despite its financial failure, the New York World's fair of 1939 and 1940 was the most outstanding American international exposition of the first half of the 20th century. Although it commemorated the 150th anniversary of the inauguration of Pres. George Washington in New York, its theme was "The World of Tomorrow." The architecture of the fair was severely modern. Tree-lined avenues leading from each important entrance plaza converged on the focal structures, the Trylon, a tall triangular shaft tapering to a point 728 ft. high, and the Perisphere, a true sphere 180 ft. in diameter, representing the structural principles of the column and the dome and "an expression of the shape of things to come."

The New York fair had the largest foreign participation—63 nations—of any fair in history, but many nations did not participate in its second year because of the outbreak of World War II. Despite disturbing international circumstances and financial reverses, the fair was able to attract 44,932,978 visitors in 354 days, the largest attendance in the history of fairs.

Golden Gate, 1939–40.—The Golden Gate International exposition, celebrating the opening of the Oakland and Golden Gate bridges, was held in San Francisco during the same period as the New York World's fair. It did not achieve the status of an international exposition but nevertheless attracted more than 17,000,000 visitors. (F. P. P.; C. Lr.)

Seattle, 1962.—The "Century 21" exposition held in Seattle, Wash., in 1962 was small compared to other such fairs of the 20th century, covering only 74 ac. The Space Needle, a 600-ft. steel tripod with a revolving restaurant at its top, and an elevated monorail train that moved passengers from downtown Seattle to the fair grounds in 94 seconds were popular with the crowds. After the closing of the fair some of the buildings, including the Space Needle, were left standing as part of the Seattle civic centre.

New York, 1964–65.—This fair, held on the site of the 1939–40 exposition, opened in April 1964. Although not officially sanctioned by the Bureau International des Expositions, the New York World's fair drew exhibitors from more than 60 foreign nations. The Unisphere, an open model of the earth wrought in steel, was the focal point of the fair. (X.)

INFLUENCE ON ART AND ARCHITECTURE

Fairs and expositions, conceived to present advanced design trends and the finest products, visited by multitudes of people, many from great distances, are expected to have an immediate and widespread influence. Japanese elements in interior decoration began to appear on the east coast of the United States after the Philadelphia centennial of 1876, and in the midwest and California following the Chicago and San Francisco fairs of 1893 and 1894, due to the display of buildings sent from Japan. The U.S. enthusiasm for columned buildings after the Columbian exposition has been referred to. This movement away from solving design problems organically was condemned by Louis Sullivan, who rightly predicted: "The damage wrought by the World's Fair will last for half a century from its date, if not longer." The Spanish colonial style ubiquitous in home building in the U.S. about the time of World War I can be accounted for largely by Goodhue's work at the San Diego fair of 1915.

European fairs around the turn of the century acquired a new look through adoption of the *art nouveau* (*q.v.*) style, as seen in Henry van de Velde's lounge at the Dresden, Ger., exposition of 1897 and in his Central hall at the Dresden fair of 1906. It appeared at the Paris exposition of 1900, especially in Pierre Dulong's fantastic design for the Pavillon Bleu restaurant, and in Georges de Feure's studiolike pavilion for the chief Parisian exponent of the *art nouveau* movement, the decorator Samuel Bing.

The Turin, Italy, exposition of 1902 marked the climax of the *art nouveau* style, the most extravagant example being Raimondo d'Avonco's domed central rotunda, a compact mass having undulating eaves, sweeping buttresses and an abundance of sculptured figure groups. These fairs were instrumental in popularizing *art nouveau* throughout Europe.

The Exposition des Arts Décoratifs at Paris in 1925 introduced and popularized a prismatic ornament known as *style moderne,* and within two years this species of decoration was everywhere apparent on commercial buildings throughout Europe and on U.S. skyscrapers.

A Century of Progress International exposition at Chicago, eight years later, similarly created its own abstract ornament applied to clean-cut forms. This exhibition, and the New York World's fair of 1939–40, did much to disengage contemporary U.S. architecture from traditional styles, to simplify its lines and shapes, to integrate planting with building, and to give fresh meaning to the use of colour, texture and new materials. (C. Lr.)

COMMERCIAL FAIRS

Commercial and samples fairs and exhibitions have assumed an increasingly important role in the promotion of international relations and trade during the 20th century. This is particularly evident throughout Europe where nearly every country has at least one major annual international exposition. These exhibitions range in scope from special fairs dealing with one industry or branch of industrial production such as automobiles or machinery, to general exhibits of goods and merchandise of every variety. Grounds and buildings of the smaller and newer fairs are generally temporary. The larger organizations have permanent grounds ranging from 200 to 500 ac. and exhibition halls extending to 100 in number.

Among the larger commercial and samples fairs the most notable

International Exhibitions and Fairs

City	Date	Area (acres)	Number of exhibits	Receipts	Sponsors' cost	Attendance
London	1851	26	13,939	$2,525,000	$1,775,000	6,039,195
New York	1853	13	4,854	$340,000	$640,000	1,250,000
Paris	1855	24⅓	20,839	£128,999	£1,000,000	5,162,330
London	1862	24¾	28,653	£459,631	£460,000	6,211,103
Paris	1867	41	43,217	£420,735	£800,000	6,805,969
Vienna	1873	40	25,760	£206,477	£2,200,000	6,740,000
Philadelphia	1876	236	30,000	$3,666,715	$8,980,000	8,004,274
Paris	1878	66	52,835	23,700,000 fr.	55,400,000 fr.	16,032,725
Sydney	1879	15	9,345	£210,372	£313,987	1,117,536
Melbourne	1880	20	12,792	£331,901	£330,330	1,330,279
London	1886	13	...	£249,861	£215,218	5,550,745
Paris	1889	72	61,722	50,000,000 fr.	44,000,000 fr.	32,350,297
Chicago	1893	686	65,000	$28,787,532	$28,340,700	21,477,212
Paris	1900	336	80,000	£4,578,249	£4,660,000	39,000,000
Buffalo	1901	350	3,500	$5,534,643	$8,860,757	8,304,073
Glasgow	1901	...	...	£404,105	£350,600	11,559,649
St. Louis	1904	1,240	70,000	$12,185,500	$26,564,000	12,804,000
Liège	1905	173	16,119	14,526,930 fr.	14,451,813 fr.	7,000,000
London	1908	140	13,500	£798,771	£783,521	8,396,673
Brussels	1910	200	...	£692,933	£702,933	4,196,939
San Francisco	1915	635	80,000	$27,178,065	$24,690,770	13,127,103
Wembley	1924–25	200	...	£2,814,935	£4,396,840	27,102,498
Philadelphia	1926	450	...	$5,000,000	$26,000,000	5,852,783
Paris	1931	500	12,000	159,000,000 fr.	225,000,000 fr.	33,500,000
Chicago	1933–34	424	...	$43,589,154	$42,900,989	39,052,236
Brussels	1935	375	9,000	...	...	26,000,000
Paris	1937	250	11,000	4,000,000 fr.	17,000,000 fr.	34,000,000
New York	1939–40	1,216⅔	100,000	$48,287,767	$67,010,989	44,932,978
London	1951	...	...	£2,560,000	£11,500,000	18,000,000
Brussels	1958	500	17,000	...	$200,000,000	41,454,412
Seattle	1962	74	...	...	$100,000,000	9,639,969
New York	1964–65	646	...	...	...	...

From Frederick P. Pittera, *The Art and Science of International Fairs and Exhibitions* (1961). Seattle (1962) and New York (1964–65) fairs added.

are the Milan fair, Italy; Zagreb International Trade fair, Yugos.; Paris International fair; Canadian National exhibition; and the several special national exhibitions exchanged between the United States and the Soviet Union.

General trade fairs, for the most part, rely for their revenues not on rent for exhibit space received from exhibitors nor on registration fees paid by trade buyers but on entrance fees paid by the public. Hence the growing popularity of specialized exhibitions, many of which are closed to the public (many trade fairs reserve certain hours or days for buyers and permit the public to attend at other times). Some trade fairs, notably those of communist countries throughout Asia and Africa, tend to group exhibitors by nationality and not by type of product. This practice has become unpopular with trade buyers who prefer to see gathered together all of the exhibits in which they are interested irrespective of national origin.

Trade fairs, expositions and exhibitions held in Europe may be grouped into several classes such as general, samples, technical, industrial and special. Best known and most numerous are the samples and general fairs which include many varieties of raw, semimanufactured and manufactured products.

The United States does not have any exhibitions or commercial fairs comparable to those in Europe, for the trade expositions regularly held in the United States are usually confined to one building and generally to a single industry.

AGRICULTURAL FAIRS

More than 2,000 state, district and county fairs are held each year in the United States. They provide industrial exhibits, demonstrations and displays aimed at livestock, horticultural and agricultural advancement. Special emphasis is placed on educational activities such as the 4-H club and similar institutions and programs. As a result of their value in the improvement of agricultural practices these fairs receive financial support from state and local governments. Money and other prizes are commonly distributed to encourage and stimulate competition in animal husbandry, crop production and home economics. Among the larger U.S. county and state fairs setting attendance records are the Texas State fair, Dallas; Minnesota State fair, St. Paul; Los Angeles County fair, Pomona, Calif.; California State fair, Sacramento; Wisconsin State fair, Milwaukee; and Michigan State fair, Detroit.

Great Britain and the Commonwealth.—Livestock or agricultural exhibits referred to as "fairs" in North America are usually termed "shows" in England. Fairs or shows evolved in Europe and on the American continent from manorial sheepshearings held by the large estate owners in the latter part of the 18th and early 19th centuries.

Prominent among the early agricultural and livestock shows of the British commonwealth was the National show held in 1821 at Aldridges' repository in London. Later such shows were sponsored by the Royal Agricultural society, beginning in 1839 at Oxford. The Royal shows were thereafter held in different parts of the country and eventually embraced all types of farm machinery and implements in addition to livestock and agricultural displays and demonstrations.

In Scotland the Royal Highland society sponsors the chief agricultural show while other societies such as the Royal Ulster Agricultural society (founded 1854) and the Royal Welsh Agricultural society (founded 1904) also hold national shows. The Royal Dublin society's spring show at Ballsbridge, Dublin, is Ireland's leading agricultural and livestock show.

Specialized shows such as the London Dairy show and Smithfield show representing the dairy and beef cattle industries respectively are typical of the many varieties of agricultural and livestock exhibits that take place throughout the commonwealth. The Smithfield show was first held in 1799 at Smithfield, later at Islington and currently at Earl's Court stadium, London. The best known fat-stock shows are those held at Edinburgh, Birmingham and Norwich, with the Smithfield show at London considered as the premier show. The livestock portion of the Smithfield show in 1960 was canceled at the last moment by national restrictions on the movement of livestock owing to outbreaks of foot-and-mouth disease.

Canada's most important agricultural show is the Royal Agricultural Winter fair at Toronto, Ont. The Canadian National exhibition, which is predominantly an international trade fair, also has a large agricultural section featured as part of its 500-ac. site at Toronto. Other smaller shows are held at Calgary, Alta.; Regina, Sask.; Saskatoon, Sask.; Brandon, Man.; and Edmonton, Alta. Other regional shows are the Central Canada exhibition at Ottawa, the Pacific National exhibition at Vancouver and the Western fair at London, Ont.

Typical of the agricultural shows of Australia is the Sydney Royal show which covers 71 ac. and is held for ten days at Easter.

Argentina.—In South America the Argentinian Exposición Nacional de Ganadería, known as the Palermo show, has become one of the leading livestock shows of the world.

United States.—Agricultural fairs in the United States were organized out of a desire to improve American agriculture and to transplant the European agricultural revolution to the newly born republic. The Arlington sheepshearings were perhaps the most popular of American livestock events. They were held on the estate of George Washington Parke Custis in the early 1800s.

Credit for founding the first agricultural society that appealed to the practical farmer is given to Elkanah Watson, a businessman and gentleman farmer of Pittsfield, Mass. He inspired the first of the famous Berkshire Cattle shows, held in 1810, and played a leading part in founding the Berkshire Agricultural society, incorporated in 1811.

Within a few years other societies were formed on the model of the Berkshire society. They also held cattle shows that later came to be known more popularly as agricultural fairs. Though subject to local modifications, these fairs were generally of the same pattern and stimulated great interest in agricultural and livestock improvement by awarding prizes for competitive exhibits. Interest in both the societies and in the fairs waned during the period from 1820 to 1825.

Significant interest in agricultural societies and fairs was stimulated by such technological developments in the late 1830s as the McCormick reaper, Deere's steel plow, threshing machines, corn planters and cultivators. As counties and states were organized, they replaced the practical societies as sponsors of the agricultural fairs.

In Sept. 1841 the first state fair was held at Syracuse, N.Y. After 1841 hundreds of state and local fairs were held throughout the nation and by 1868 there were 1,367 state, county and district fairs each year—a period referred to as "the golden age of the agricultural fair." New York's first state fair was soon followed by one in New Jersey. Other states followed suit—Michigan in 1849; Pennsylvania, Ohio and Wisconsin, 1851; Indiana, 1852; Illinois, 1853; and Iowa, 1854.

Since 1900 the agricultural fair in the United States has grown steadily in importance, with a trend toward expanded entertainment programs and events in which rural youth organizations such as the 4-H clubs, Future Farmers of America and similar organizations compete for prizes. During the 20th century, community and district fairs displaced many of the county fairs while regional fairs replaced many of the state fairs. The best-known regional fairs include the Eastern States exposition held at Springfield, Mass.; the Mid-South fair, Memphis, Tenn.; the Tennessee Valley exposition, at Huntsville, Ala.; and the Atlantic Rural exposition, Richmond, Va.

State fairs have come to be well-established state institutions and are usually conducted by a bureau in the state department of agriculture or by a special state fair board. The early temporary wooden buildings housing exhibits of livestock and industrial products have in many instances given way to well-designed permanent buildings surrounded by landscaped gardens and other decorative features.

Exhibitors' products are more advantageously displayed to the public. New grandstands, amphitheatres and lavish entertainment spectacles have led to increases in admission fees over the years. The best-known livestock expositions in the United States include the American Royal Livestock exposition, held annually in Kansas City, Mo., in October; Southwestern Exposition and Fat Stock show, Fort Worth, Tex.; National Western Livestock show, Denver, Colo.; Ak-Sar-Ben Livestock show, Omaha, Neb.; Grand National Livestock exposition, San Francisco; Pacific International Livestock exposition, Portland; and the most notable of all, the International Livestock exposition, Chicago, which traces its history to the American Fat Stock shows held annually in Chicago from 1878 to 1897. (F. P. P.)

BIBLIOGRAPHY.—Sigfried Giedion, *Space, Time and Architecture*, pp. 178–211 (1943); Georg J. Carstensen and Charles Gildemeister, *New York Crystal Palace* (1854); Hubert Howe Bancroft, *The Book of the Fair* (1893); Frank H. Norton, *Illustrated Historical Register of the Centennial Exhibition, Philadelphia, 1876* (1879); Herbert S. Fairall, *The World's Industrial and Cotton Centennial Exposition, New Orleans, 1884–1885* (1885); John J. Flinn, *Official Guide to the World's Columbian Exposition* (1893); *The Official History of the California Midwinter International Exposition* (1894); Holman D. Waldron, *With Pen and Camera at the Pan-American Exposition* (1901); Walter B. Stevens, *The Forest City* (1904); David R. Francis, *The Universal Exposition of 1904* (1913); James A. Buchanan (ed.), *History of the Panama-Pacific International Exposition* (no date); *Official Guide Book of the Panama-California Exposition* (1915); Jack James and Earle Weller, *Treasure Island "The Magic City"* (1941); Frank Monaghan (ed.), *Official Souvenir Book, New York's World's Fair, 1939* (1939). Frederick P. Pittera, *The Art and Science of International Fairs and Exhibitions* (1961); Wayne Caldwell Neely, *The Agricultural Fair* (1935); P. Boissonnade, *Life and Work in Medieval Europe* (1927); George R. Leighton, "World's Fairs: From Little Egypt to Robert Moses," *Harper's*, pp. 27–37 (July 1960), "The Year St. Louis Enchanted the World," *Harper's*, pp. 38–47 (Aug. 1960). (C. Lr.; F. P. P.)

EXILARCH (PRINCE OF CAPTIVITY), the title of the head of the Jewish communities in Babylonia and adjacent countries in ancient and medieval times. At the height of his career the exilarch was described by the distinguished Muslim scientist al-Biruni as "the master of every Jew in the world; the ruler whom they obey in all countries." The exilarch claimed direct descent from the house of David through King Jehoiachin, exiled to Babylonia before the fall of Jerusalem in 586 B.C. This claim, though supported only by a questionable genealogy recorded in the 8th century, was nevertheless recognized by the exilarchs' main rivals, the Palestinian patriarchs; it has also received some support from 20th-century archaeological discoveries.

Their real authority is documented, however, only for Sasanian Persia (after A.D. 226) and the great caliphate (after 640): in both the exilarchs were recognized as heads of the Jewish communities, serving in the imperial councils as representatives and ministers for Jewish affairs. As such they exercised considerable control over Jewish taxation, the judiciary and communal administration.

Before long, however, their authority began to be contested by the scholarly heads of the academies of Sura and Pumbeditha. At the end of the 3rd century, Nahman, himself a son-in-law of the reigning exilarch, claimed supreme jurisdiction in legal matters. Soon the communal assemblies convened at the seat of the academy rather than at the exilarchic court, Ashi (d. *c.* 427), head of the Sura academy, finally arrogating to himself the title rabban, theretofore reserved for the exilarch. Similarly, under the Muslim regime the early power of the exilarchs was whittled away under rabbinical onslaughts. These reached their climax in the rabbinate's repudiation of 'Anan ben David in favour of his brother, which became the immediate cause of the rise of the Karaite schism (*see* KARAISM), and in Saadia ben Joseph's conflict with Exilarch David ben Zakkai (in the 930s). At one time Exilarch Hezekiah held for 20 years (1038–58) both offices, but after his death the controversies were renewed.

This recurrent struggle, facilitated by the uncertain laws of exilarchic succession, which combined the hereditary principle with elections by leading scholars and royal appointment, resembled only remotely the western conflicts between church and state, for the exilarchic office itself was but an ecclesiastical institution, on a par with that of the Nestorian catholicos (*q.v.*). Within the Jewish community, nonetheless, the exilarch represented the more secular, the head of the academy the more strictly religious forces.

The power of the exilarchs depended upon the rise and fall of the Persian and Muslim imperial fortunes. During the period of anarchy generated by the rise of the semicommunist Mazdakite movement in Persia, Exilarch Huna Mari was executed (in 471). Mar Zutra II, Huna Mari's successor, established a Jewish principality that lasted for seven years, but he was ultimately defeated and killed (*c.* 491). His son, Mar Zutra III, emigrated to Palestine, the exilarchic office remaining vacant until its revival by the Arab conquerors' appointment of Bostanai (*c.* 640). Once again the dissolution of the great caliphate in the mid-10th century marked the end of exilarchic authority. The rise of independent Jewish leaders in various provinces reduced the exilarch to the rank of a local chieftain, despite the pomp and ceremony still witnessed by the traveler Benjamin of Tudela on his visit to Baghdad (*c.* 1170). The Mongol conquest of Baghdad in 1258 marked the end of the office, although some descendants of the exilarchic line roamed through the eastern communities for another century or two, demanding and sometimes receiving princely homage and financial support.

BIBLIOGRAPHY.—Felix Lazarus, *Die Häupter der Vertriebenen* (1890); Alexander D. Goode, "The Exilarchate in the Eastern Caliphate, 637–1258," *Jewish Quarterly Review*, 31:149–169 (1940–41); Salo W. Baron, *The Jewish Community*, 3 vol. (1942), and *A Social and Religious History of the Jews*, 2nd ed., 8 vol. and Index vol. (1952–60), with additional bibliography. (S. W. B.)

EXILE: *see* DEPORTATION AND EXILE.

EXISTENTIALISM is the name of a philosophical tendency or attitude which manifested itself in Germany a few years after World War I and later spread to France and to Italy so that immediately after World War II it was not only influential in academic circles there but was widely talked about in literary cafés and the semipopular press. Since existentialism is a tendency or attitude rather than a philosophical school, there are few doctrines common to all exponents of it. But it may be generally characterized as a protest against views of the world and policies of action in which individual human beings are regarded as the helpless playthings of historical forces or as wholly determined by the regular operation of natural processes. All existentialist writers seek to justify in some way the freedom and importance of human personality. They all emphasize, too, the place of will in human nature by contrast with reason. A consequence of this is that some of them have spread their views by means of novels and plays as well as by the more ordinary ratiocinative methods. Most of the existentialist writers have been consciously influenced by the works of the Danish thinker Søren Kierkegaard (*q.v.;* 1813–55), which, neglected in their author's lifetime, exerted an increasing influence after the publication of the German translation of 1909–14. In particular, it was from Kierkegaard's use of it that the word "existence" gained the significance that it now bears as a technical term in the existentialist philosophy. We shall start, therefore, with an elucidation of this special meaning of the word "existence" and with a brief account of Kierkegaard's main contributions. We shall then give an outline of the range of ideas of later existentialist philosophers and conclude with some comments on the significance of the existentialist movement as a whole.

Kierkegaard as the Originator of Existentialism.—The word "existentialism" was used to translate the German *Existenzphilosophie*. *Existenz* became a technical philosophical term in Germany after World War I. In nontechnical language anything actual may be said to exist, whether it be a man or a fish or a stone, but in the sense with which we are now concerned it is primarily human beings who are said to have "existence." The word is then used to emphasize the claim that each individual person is unique and inexplicable in terms of any metaphysical or scientific system; that he is a being who chooses as well as a being who thinks or contemplates; that he is free and that, because he is free, he suffers; and that since his future depends in part upon his free choices it is not altogether predictable. There are also suggestions, in this special usage, that existence is something genuine or authentic by contrast with insincerity, that a man who merely contemplates the world is failing to make the acts of choice which his situation demands.

Running through all these different though connected suggestions is the fundamental idea that each person exists and chooses in time and has only a limited amount of it at his disposal in which to make the decisions which matter so much to him. Time is short; there are urgent decisions to make; we are free to make them, but the thought of how much depends on our decision makes our freedom a source of anguish, for we cannot know with any certainty what will come of us.

The word "existence" acquired these peculiar meanings as a result of Kierkegaard's use of it in his *Concluding Unscientific Postscript* (1846). In this book Kierkegaard holds that Hegel's attempt to demonstrate that the world is a rational system, that "the real is the rational and the rational is the real," is not only presumptuous but ridiculous, for it rests upon the assumption that a particular part of an as yet uncompleted scheme which is not in process of being created by itself could know what its completed form must be. If the world were a system it would take God Himself, not Hegel, and still less a reader of Hegel's books, to know it. Finite individuals, therefore, behave like figures of comedy when they claim to be able to demonstrate that any individual occupies a specific place in some rational system of the world. It follows from this, according to Kierkegaard, that no one can know his place, that no one can have his duty proved to him, but that each must take his courage in both hands and choose as best he can. Kierkegaard was a profoundly religious man who conceived of himself as having to choose before God with no possibility of knowing whether the outcome would be his salvation or his damnation. The question was not whether Christianity was true, in the sense of being philosophically respectable and fully supported by historical documents. One could spend one's life investigating such matters without reaching an assured conclusion, but then one would, in effect, have chosen *against* Christianity. In order to become a Christian it is necessary to commit oneself by leaping into the darkness. The religious man is like someone who puts his money on an outsider; he acts from faith in "the absurd"—that is, in something for which there is no rational proof. According to Kierkegaard, there is about this religious choice a particularity, a loneliness and an anguish which may be lacking in the moral sphere where a man, in committing himself, acts as he would have other men act in like circumstances. "Sickness," writes Kierkegaard, "is the natural state of the Christian" (*Stages on Life's Way*, 1845). Abraham, in obeying God's command to kill Isaac, went against the moral law, but in doing so he did not lose Isaac but gained him. When a man's immortal soul is at stake, aesthetic enjoyment, subtle speculation and even moral conduct toward one's fellows fall into place as subsidiary to the crucial choices in which one determines what one really is. It was the greatness of Socrates, according to Kierkegaard, that he abandoned his researches into the natural world and, disclaiming any certain knowledge about what would become of him or of any other man, showed the quality of his faith by the life that he led. When, in the *Concluding Unscientific Postscript*, Kierkegaard says that truth is subjectivity, he means not that the nature of the world is revealed in some personal and incommunicable premonition rather than by objective inquiry, but that everything anyone does, including his search for objective truth, gets its value from the way in which it is willed and decided by him.

According to Kierkegaard, then, God is not discovered by abstract demonstration or in nature; religion is a matter of inward choice involving the leap of faith; art, science and history are subordinate in value to the lives of which they form part; these lives are constituted by the decisions of men who necessarily choose in partial ignorance; and men are therefore in a condition of anxiety and must, if they become Christians, pass through despair.

20th-Century Existentialism.—Kierkegaard wrote at a time when organized Christianity was very strong, and his criticisms of formal, worldly Christian conformity—of what he called "Christendom"—were in favour of a genuine renewal of the old religion. In the years after 1918 Christian institutions had lost influence, so that a call for active and honest personal decision was as likely to be in terms of an agnostic or atheistic point of view as in favour of a supposedly "real" or "genuine" Christianity. The moral and social upheavals of the time made Kierkegaard's descriptions of the anguish of choice in face of uncertainty seem unpleasantly apposite, but revealed Christianity called for a sort of "leap" that many considered irrelevant to their situation. In these circumstances Kierkegaard's insights came to be developed in a variety of ways.

Perhaps the most thorough and satisfying and certainly the most ample development of the philosophy of existence in the 20th century is contained in the writings of Karl Jaspers (*q.v.;* 1883–). Like Kierkegaard and many 20th-century philosophers who are not existentialists, Jaspers holds that the day of rational metaphysical systems like that of Hegel has gone, but he sees the chief danger to individuality not in "Christendom" but in our highly organized technological society in which the chief aim is to produce a standardized level of satisfaction for as many people as possible (*Die geistige Situation der Zeit*, 1931; Eng. trans., *Man in the Modern Age*, 1933). Kierkegaard, like Socrates, had depreciated the importance of natural science, but Jaspers, although he admits that the search for more and more facts can be a means of escape from moral decisions, holds that the devoted exploration of regions which can never be completely mapped is one of the most admirable and significant of human activities (*Philosophie*, 3 vol., 1932, vol. i). Jaspers speaks not of God (except in his more popular writings) but of "transcendence," by which he appears to mean that the very incompleteness of our knowledge and

the limitations of our power point to some unknown and unknowable source of being. (This topic is most fully discussed in *Philosophie*, vol. iii.) There is no written or traditional revelation from which we can discover what to do. It is not even possible to prove that we are free, for alleged proofs can refer directly only to what exists objectively, and freedom is not a characteristic of an object at all (*Philosophie*, vol. ii). Jaspers elaborates the part of Kierkegaard's teaching in which the inadequacy of general rules and the necessity for crucial and critical decisions are described. He is very conscious of the danger of shattering the fundamental existentialist theme by turning it into a system, and he therefore regards his own work as a setting before others of possibilities which it is for them to actualize as their experience and circumstances require.

A work which has greatly influenced later existentialist writers is Martin Heidegger's *Sein und Zeit*, volume i (1927). Heidegger (*q.v.*) was a pupil of Edmund Husserl, the founder of phenomenology (*q.v.*). Husserl held that philosophers should turn their attention away from the natural world and toward the inner experiences which are, he alleged, basic for our apprehension of the natural world and our thought about it. Heidegger follows this course to the extent of claiming to describe fundamental experiences which are behind our everyday and scientific knowledge, but his account of what he finds when he exercises this method is very different from and much more bizarre than anything that Husserl succeeded in uncovering. It is as though Kierkegaard had ceased to believe in God without losing his distaste for science. Heidegger thus describes man as cast into an unsympathetic world in which he tries to achieve purposes all of which will inevitably come to nought in death. He may try to evade the thought of his own coming dissolution by living his life in terms of impersonal and conventional generalities, but he can be true to himself only by living constantly with the thought of his eventual death. Heidegger's account of man as inevitably given over to care and subject to a fate to which he can close his eyes but which he cannot evade was intended as the first part of a theory of being as a whole (ontology). Now we have seen that existentialists are skeptical of the possibility of any such theory, and theretore Heidegger prefers not to be regarded as an existentialist. His disclaimers, however, cannot alter the fact that his own sombre assessment of man's condition is rooted in Kierkegaard and has been incorporated in the work of Jean Paul Sartre (1905–), whose *L'Être et le néant* (1943) is an arresting and thorough working out of the atheistic form of existentialism. This book and the novels and plays in which themes from it are independently developed have made some of Sartre's phrases part of the literary currency of the time—such observations, for example, as that "man is condemned to be free"; that his life is "an ineffective passion"; that "hell is other people." The basis of Sartre's work is the distinction between unconscious beings who are said to exist only *in themselves* and conscious beings who, because they transcend the present place and time and exist *for themselves*, are necessarily free. His ethics is an attempt to draw the consequences of ceasing to believe in a God who decrees and reveals the rules that men should live by. That the fundamentals of existentialism are compatible with Christianity is the view of the Catholic existentialist Gabriel Marcel (1889–), who appears to have come to his conclusions through his own criticism of idealism rather than under the influence of Kierkegaard (*Journal métaphysique*, 1928; Eng. trans., 1952). Marcel is especially important for his elaboration of the ethical consequences of generalizing and abstracting and for his description of the motives of so doing. He holds, for example, that wars would be less likely to occur if we ceased to substitute such abstractions as "Fascists" or "Communists" for the really existing particular human beings (*Les Hommes contre l'humain*, 1951; Eng. trans., *Men Against Humanity*, 1952). A prominent representative of existentialist thought in Russian Orthodox Christianity was Nikolai Aleksandrovich Berdyaev (*q.v.*).

Critical Comments.—The criticism most often brought against existentialism is that it is a form of irrationalism. This ambiguous charge may be understood in a number of ways. In the first place it may be preferred by the very rationalists who are the objects of existentialist criticism; if so, it amounts to saying that the world is and may be known to be a rational system. On this point most contemporary philosophers would side with the existentialists as a result of accepting the empiricist principle that human knowledge cannot reach certainty beyond the range of possible experience. Empiricism was a feature in the development of Kierkegaard's own ideas, for he was influenced by the lectures given in Berlin in 1841–42 by F. W. J. Schelling (1775–1854), in which the inexplicability of the particular was stressed, and he quotes approvingly from Kant's friend J. G. Hamann (1730–88), who supported Christianity in terms of Hume's radical empiricism and in opposition to the arguments of rational theology. We may say that the existentialists enunciate the empiricist principle in rather excited tones, exclaiming that particular facts are "absurd" or "*de trop*" instead of merely stating that they just are as they are. The existentialists are also at one with those empiricist philosophers who hold that it is impossible to obtain such a systematic view of human society as would enable firm predictions to be made about its future state. If the existentialists' irrationalism, again, consists in holding that moral rules can give guidance but cannot give certain knowledge of what one ought to do in particular circumstances, there are many philosophers who are not existentialists who would agree with them on this. What seems quite unjustified is a suggestion in some of Sartre's writings that human dignity is displayed on occasion in acts of singular atrocity. It is easy, too, to exaggerate the amount of uncertainty in moral decision, as for most men most of the time their duty is clear enough. Kierkegaard's attitude toward science may justly be criticized as irrationalistic, and his view of religious faith deserves that description too. For even if it be granted that religious faith involves a leap beyond what reason guarantees, this does not mean that some leaps are not more reasonable to take than others. It is clear that existentialist theologians set too little store on natural theology for them to receive approval from the Roman Church. In his encyclical *Humani generis* (1950) Pius XII accused existentialists of believing that "a man who cannot make up his mind what is the true answer to some intellectual problem need only have recourse to his will" and exclaimed: "A strange confusion, here, between the provinces of thought and of volition!" But this sort of criticism does not touch such a writer as Jaspers, who certainly would not override this distinction. Marxist critics of existentialism regard it as the ideological reflection of capitalism's inability to prevent crises and wars, but existentialists reply that the Marxist ideal of controlling nature and society is unrealistic. According to existentialists, moreover, there is an element of pride and presumption in claiming to know what future generations of free men will or ought to do. Whereas the Marxist honours Prometheus for his inventive and hopeful courage, the existentialist looks upon Sisyphus as the human exemplar because of his disillusioned, hopeless, and therefore still more courageous, striving. With its emphasis on individuality, existentialism is undoubtedly a predominantly liberal philosophy, but it is a liberalism with no belief in a natural social harmony and no hope in man's perfectibility.

See also ETHICS, HISTORY OF: *Existentialism;* FREE WILL.

BIBLIOGRAPHY.—R. Bretall (ed.), *A Kierkegaard Anthology* (1947); J. Wahl, *Etudes kierkegaardiennes* (1938), and *Les Philosophies de l'existence* (1954); K. Löwith, *Von Hegel zu Nietzsche*, 2nd ed. (1941); K. Jaspers, *Vernunft und Existenz* (1935, new ed., 1947; Eng. trans., *Reason and Existenz*, 1955); H. J. Blackham, *Six Existentialist Thinkers* (1952); H. Knittermeyer, *Die Philosophie der Existenz* (1952); F. H. Heinemann, *Existentialism and the Modern Predicament*, 2nd ed. (1954). (H. B. A.)

EXMOOR, a high moorland in northwest Somerset and north Devon, Eng., forms a national park 265 sq.mi. in extent. It borders the Bristol channel on the north and has an exceptionally beautiful coast line of rugged headlands interspersed with narrow, wooded valleys, or coombs. Inland, beyond a fringe of farms, lies the bleak misty plateau of heather moors, mostly more than 1,000 ft. above the sea, with Dunkery Beacon (1,708 ft.) as the highest feature.

Exmoor was owned by Alfred the Great and was a favourite hunting ground of King John in about 1200. Much of it remained a royal forest until 1815, when 19,000 ac. were legally disafforested

and sold to private developers. The moors, proving unsuited to agriculture, remain grazing grounds for hardy Exmoor ponies and Exmoor horned sheep. Wild red deer are hunted there.

The rivers East and West Lyn and Badgworthy water have short turbulent courses northward, whereas the Exe and the Barle flow more gently to the south; all hold trout. The Badgworthy valley, on the Devon-Somerset border, is the Doone valley of fiction, immortalized by R. D. Blackmore in his novel of moorland outlaws, *Lorna Doone*. Lynton, Lynmouth and Porlock are the main tourist centres. The roads are exceptionally hilly. (H. L. En.)

EXMOUTH, EDWARD PELLEW, 1st Viscount (1757–1833), English admiral, chiefly remembered for his great battle in 1816 when he defeated the dey of Algiers and forced him to abolish Christian slavery, was born at Dover on April 19, 1757. Both he and his brother Israel (who also became an admiral) entered the navy in 1770. Pellew saw service in the American Revolutionary War and was promoted captain in 1782. The next year he married Susan Froude, by whom he had six children, of whom his second son, Fleetwood, became an admiral. In 1793, when commanding the frigate "Nymphe," he captured the "Cléopâtre," the first French frigate to be taken in the war, an action for which he was knighted. In 1795 he became captain of the frigate "Indefatigable" attached to the western squadron, the only ship to catch sight of the French expedition to Bantry bay in Dec. 1796 when, by firing false signals, he was able to confuse the French armament. On Jan. 13, 1797, the frigates "Indefatigable" and "Amazon" drove ashore the French line-of-battle ship "Droits de l'Homme."

Pellew became member of parliament for Barnstaple in 1802 as a follower of Lord St. Vincent, whom he defended against the attacks of William Pitt. From 1804 till 1809 he commanded the East Indies station, where he defeated a Dutch squadron off Java and protected trade against French cruisers. He took over the command of the Mediterranean fleet in 1811 and was created Baron Exmouth in 1814. After the war he negotiated with the dey of Algiers over the ransom of slaves. A massacre of 200 Christian fishermen had violated the truce imposed by the congress of Vienna, and Exmouth was sent to Algiers with a force of 17 ships, to which the Dutch added 6 frigates. The strong fortifications of the town were bombarded on Aug. 27, 1816, and after the expenditure of more ammunition than in any single battle in the war just ended, the batteries were silenced and the Algerian fleet destroyed. This victory enabled the European powers, and also the Americans, who had been acting independently, to dictate terms to the dey and to abolish Christian slavery. Algerian corsairs did not disappear, however, until the French occupied the town in 1830. Exmouth was created a viscount after the attack and retired from the sea. He died at Teignmouth on Jan. 23, 1833. Though never a popular admiral, he was a fine seaman and a good tactician.

Bibliography.—E. Osler, *Life of Admiral Viscount Exmouth*, rev. ed. (1854); C. N. Parkinson, *Edward Pellew, Viscount Exmouth* (1934), *War in the Eastern Seas, 1793–1815* (1954). (C. C. L.)

EXMOUTH, a seaside market town and urban district of Devon, Eng., on the east bank at the mouth of the Exe, 11 mi. S.S.E. of Exeter by road. Pop. (1961) 19,753. Its sandy shores and mild climate and the beauty of its coastal and moorland scenery attract many visitors in summer. A promenade runs along the sea wall and there are golf links, public gardens and a pavilion. The port is a yachting centre, a regatta being held annually. A steam ferry connects Exmouth with Starcross on the opposite shore.

Exmouth was early a place of maritime importance and in 1346 contributed ten vessels and 193 seamen to the fleet sent to attack Calais. Sir Walter Raleigh was born at Hayes Barton, west of the town. In 1646 the fort, which commanded the estuary of the Exe, was captured by the parliamentarians, and during the 16th and 17th centuries the town was raided by Barbary pirates. A curious house in Summer lane, called "A la Ronde," was built in 1798—a 16-sided building with shell decorations, said to be modeled on the church of San Vitale at Ravenna, Italy. At that time Exmouth was a fishing village, but in the 19th century shipbuilding was carried on and docks for river shipping were constructed (1869). Terraces were built up the sides of Beacon hill and the town became one of the leading seaside resorts in Devon. Fishing on a considerable scale succeeded shipbuilding and was in turn succeeded by the tourist industry.

EXOCHORDA, commonly known as pearlbush, a small genus of ornamental Asiatic shrubs of the rose family, consisting of hardy deciduous, spiraealike plants grown chiefly for their showy racemes of pure white flowers. *E. racemosa*, a slender spreading shrub often reaching 10 ft. high, and *E. giraldi* and its variety *wilsoni*, both vigorous growers, are generally considered the best of the group. They grow best in a sunny location in a well-drained loamy soil. *See* Rosaceae. (J. M. Bl.)

EXODUS (in the Hebrew Bible, Shemoth), the second book of the Old Testament and of the five books of the Law (the Pentateuch), traditionally known as "the second book of Moses." Its English title is the equivalent of that in the Septuagint, the Greek version of the Old Testament. The word connotes a going out or departure and refers to the escape of the early Israelites from their slavery in Egypt, the account of which is related in the first 15 chapters of the book.

Contents.—The Book of Exodus may be outlined as follows:

i 1–xv, 21:	the deliverance from Egypt
i, 1–vii, 13:	God prepares for the deliverance
i, 1–22:	the situation of the enslaved
ii, 1–25:	the birth and training of Moses
iii, 1–vi, 1:	the call and commission of Moses
vi, 2–vii, 13:	the appointment of Aaron
vii, 14–xv, 21:	the great acts of God for Israel's deliverance
vii, 14–xi, 10:	the plagues (signs) in Egypt
xii, 1–xiii, 16:	the Lord takes possession of Israel
xiii, 17–xv, 21:	God's final triumph and Israel's joy
xv, 22–xviii, 27:	Israel's desert pilgrimage
xv, 22–xvii, 7:	God gives water and manna
xvii, 8–16:	the Amalekites
xviii, 1–27:	the visit of Jethro, Moses' father-in-law
xix, 1–xxiv, 18:	the sealing of the covenant at Sinai
xix, 1–25:	the theophany; God's presence manifest to Israel
xx, 1–xxiii, 19:	the obligations of the people under the covenant
xx, 1–17:	the Decalogue
xxi, 1–xxiii, 19:	the social code
xxiii, 20–33:	a closing exhortation
xxiv, 1–18:	the cultic ratification of the covenant
xxv, 1–xxxi, 18:	instructions for the building and furnishing of the Tabernacle
xxxii, 1–xxxiv, 35:	the golden calf, the intercession of Moses and the restoration of the covenant
xxxv, 1–xl, 38:	the execution of the blueprints for the Tabernacle and its furnishings

Literary History.—Since the Book of Exodus is integrally a part of the literary unit known as the Pentateuch, its literary history is originally related to that of the Pentateuch and cannot be discussed apart from it. A full treatment of this history, and especially of the various stages and results of the critical research that has aimed at its reconstruction, is given in the article Pentateuch.

The process of critical research in the history of the Old Testament has made it increasingly evident that many parts of it had a long history as oral tradition before becoming fixed in literary forms. This is especially so in the case of the Pentateuch. Israel was a community of faith. As such it was persistently concerned to confess its faith in God and to bear witness to his deeds and his character. This concern expressed itself especially in communal observances and acts of worship. Scripture had its inception in this communal "response of faith." While all sorts of individual voices and authors contribute to the growth of the Old Testament, as in the books of the prophets, it is also true that in some parts of it the communal testimony predominates over the contributions of individual authors in the shaping of the literature. This is probably so in the case of Exodus, especially the first 24 chapters.

Since, as suggested, the growth of a book such as Exodus was motivated by Israel's concern to worship—*i.e.*, to recite and celebrate the great acts of God's redemption communally and ritu-

ally—the book is a testimony of faith, an account of what is spontaneously assumed and celebrated. When it deals with history its main concern is not with a scientific documentation of fact but with faith's interpretation of the significance of an event. Similarly, its account of the deeds and character of God takes the form of a spontaneous declaration rather than a critical and reflective analysis. Moreover, since faith assumes that the destiny of all is provided by the divine action, the community's recital is wholly theocentric. Perhaps the closest contemporary literary equivalent for a book such as Exodus would be a church liturgy, complete with rubrics, or, in a more limited sense, a hymnal. Thus the concern and subject matter of Exodus, like that of the Pentateuch as a whole, is antecedent to the history of the literary development that issued in its present form.

Purpose and Themes.—The general purpose of Scripture is the recital of the great redeeming action of God, which is the central theme of the community of faith that produced the Bible. The specific purpose of Exodus, in this general context, is to relate the deliverance and establishment of Israel as the special people of God who are to function both as the primary illustration and as the instrument of his power and his goodness. The main headings in the outline of the book given above make this clear. The purpose and content of Exodus may be examined by examining each of these four topics in detail.

Deliverance From Egypt.—The stories of the patriarchs in Genesis had set the stage for the drama of the exodus. The call of Abraham (Gen. xii, 1–9) and the migration of his descendants to Egypt are recalled in Ex. i, 1–6 to stress the relation of the exodus to preceding events. And the notice (i, 7) of the rapid increase of the tribes in number is intended to show that this book, too, is to deal with events that will fulfill the ancient promises to Abraham that his numerous offspring would inherit the land of Canaan (Gen. xiii, 14–18). The events in this book, thus, are seen as a new phase in an ongoing drama.

Verifiable facts relating to the escape of the slaves from Egypt are scanty. Beginning with the violently anti-Jewish Egyptian historian Manetho (250 B.C.) the event was traditionally placed in the 16th century B.C. and associated with the expulsion of the Hyksos power from Egypt. But this association lacks evidence. Modern scholars are increasingly unanimous in the view that the exodus occurred in the 13th century B.C. Ramses II (1304–1237) is commonly referred to as the Pharaoh of the oppression; and the escape probably occurred in his reign or in that of his son, Merneptah (1236–23). Ex. i, 11 states that the Hebrews worked as slaves on the building of the "store cities, Pithom and Raamses." Recovered Egyptian records and archaeological research have established the fact that these cities were located in the Nile delta and were founded by Ramses, who was the first of the post-Hyksos rulers to establish his capital in lower Egypt. Both the sites and time of Israel's enslavement are thus quite precisely established. The Merneptah stele raises some questions about the time. This monument, raised by the Pharaoh to record his victories in Canaan, lists Israel among the peoples he destroyed in this campaign, implying that Israel was in Canaan during his reign. This led some scholars to place the exodus rather early in the reign of Ramses II. However, there is increasing evidence that the Hebrews in Egypt did not include all clans known by the title Israel and that, therefore, the inscription of Merneptah may not be a decisive factor in fixing the date of the exodus.

The migration of seminomads from Asia, Semites driven by famine into the sown lands of both Egypt and Mesopotamia, was a constant feature of ancient oriental life, abundantly illustrated in Egyptian records. That such migrants maintained their distinctive tribal customs and organization amid a settled society, often for several generations, and sometimes returned to their nomadic places of origin is also a matter of record. There are even extant Egyptian portrayals of Semites engaged in bricklaying under Egyptian oversight. The biblical account of the Hebrew settlement in Egypt and of the exodus is thus also in harmony with the general pattern of ancient life in the near east.

Israel's crossing of the "Sea of Reeds" was the climax of the exodus. Considerable uncertainty remains about the precise location of this crossing, though there is general agreement that it was not at what is now known as the Red sea, but farther north. There can, however, be no doubt about the fact of the occurrence or about the critical role it played in the escape. The so-called Song of Miriam (Ex. xv, 21), the composition of which literary historians unanimously associate immediately with the events of the exodus, makes the crossing the decisive stage of the escape, and the cultus of Israel is largely built on this.

The general result of scientific research relating to the material facts of the exodus is twofold. In the first place, it establishes the event; everything confirms the biblical report that there really was an escape of Israel from Egypt. In the second place, it leads to the conclusion that this event was much less obtrusive in the ancient near eastern scene in which it occurred than the biblical account seems to imply. In fact, it hardly affected the socioeconomic or political processes or structures of power. As in the case of the death of Jesus 12 centuries later, it was not immediately noted or recorded by outsiders. The movements of Semitic migrants being what they were, the report of successful departure of Hebrew slaves fit naturally into the scheme of things, especially since, as seems probable on many counts, the numbers involved were very much smaller than a later era imagined (Ex. xii, 37). But for Israel this very ordinary occurrence is the event of revelation. It is the event that is occasioned by and occasions Israel's faith in God as the Lord of all nature and history; and in Exodus it tells the story not to record a series of material facts but to proclaim its faith in the power and goodness of God.

Ex. i–xv probably developed as a liturgical recital of faith confessing the great acts of God and using the incidents of Israel's escape as illustrations. Johannes Pedersen proposed that this section of the book may have developed as a Passover liturgy. In any case it is a liturgical dramatization of Israel's exodus from slavery designed to serve as a symbol of its faith.

The central theme of this section is the authority of God as creator and ruler of nature and as Lord of history. From i, 1 to vii, 13 the narrative sets the stage on which the action of God as deliverer will occur. The famine-driven migrants living in Egypt are oppressed by a "new king . . . who did not know Joseph" (i, 8). The rescue and training of Moses, humanly speaking an accidental by-product of the oppression, is treated as a special aspect of God's utilization of all human and natural situations, however recalcitrant, as serviceable for his plans. This overruling control involves individual lives as well as the comprehensive processes of human existence.

The narrative of the ten plagues declares God's rule of nature. The disclosure of his power and authority is presented in dramatic fashion, each plague illustrating the divine power more fully. Thus the magicians were reportedly (vii, 11; *22*, viii, 7) able initially to match the acts of God. But at the third of the ten plagues they acknowledge defeat (viii, 18 ff.); and the death of the first-born illustrates the faith that God's authority is universal and absolute. The institution of the feast of Passover and, especially, of the rite of the first-born (xii, 1–xiii, 16) is set in the context of the account of the slaying of the first-born of Egypt to emphasize that God's authority includes his possession of Israel also. Israel's offering of the first-born is its confession that God is the Lord of all life. While Israel makes this confession willingly, the same truth was brought home to Egypt as a divine judgment.

The exodus narrative declares the power of God by citing "signs" of his authority over nature, but its good news is the announcement of God's defeat of the Pharaoh. For a book such as Exodus, the theocentric perspective stresses the continuity between nature and history; the processes of nature and the world of man's making are not set over against one another. But God's revelation of himself is in the world of human events, specifically in his defeat of Egypt as illustrated by the exodus of the slaves. For faith this episode becomes the warrant for the affirmation that God is the Lord of history always and everywhere and that no tyrant can withstand him. Thus Exodus presents the history of Israel's escape as an act of God; and, on the basis of this single very particular episode, Israel bears its faith about the meaning of all history.

For Israel the exodus was an escape and a return to freedom.

Therefore, in crediting God with it, Israel acknowledges him as saviour and redeemer. He is the one who in his goodness came to Israel's rescue when it could not help itself. This is a theme in the narrative that becomes decisive for the whole biblical tradition; *i.e.*, the power and rule of God are always confessed on the basis of an act or event that features his grace and goodness; God is affirmed as creator in the light of his revelation as redeemer.

Israel's Desert Pilgrimage.—Following their escape the Israelites set out for Canaan, the land of promise. But according to tradition they spent 40 years in the wilderness, mostly on the southern and eastern borders of Canaan, before finally entering it under Joshua. As in the case of the exodus, the scientifically verifiable details relating to this period are few. Modern research confirms the biblical report that the centre of Israelite life during most of this period was at Kadesh, south of the Negev. According to the tradition (Num. x, 11–12) Israel set out for "the wilderness of Paran" and Kadesh in the second year after the exodus, following the sealing of the covenant at Sinai; and the departure from Kadesh, to travel around Edom to the east side of Jordan, marked the beginning of the end of the nomadic period (Num. xx, 14–22).

The Book of Exodus confines its treatment of the wilderness period to that brief portion which begins with the crossing of the sea and ends at Sinai, reportedly a period of three months (Ex. xix, 1). The further narratives on the wandering begin at Num. x. Since the location of the places on the journey cannot be verified, the direction of the route is uncertain, especially since the identification of Mt. Sinai is disputed. Since at least as early as the 4th century A.D. Jebel Musa, the southern point of the Sinai peninsula, has been accepted as Sinai; but modern scholarship is more inclined to associate the mountain of the Law with the wilderness of Paran in the neighbourhood of Kadesh, and there is persuasive biblical evidence for this (Num. x, 12; Deut. xxxiii, 2; Judg. v, 4–5).

As in the case of the exodus, the stories of the wandering are told to bear witness to the benevolent power and leadership of God in Israel's history. They represent a further instance of the habit of biblical faith to take a particular historical episode or natural occurrence as an occasion for confessing God as Creator and Lord. What the sections on the water, manna and quails (xv, 22–xvii, 7) have as their real theme is that human hunger and thirst are satisfied by God the Creator. The spiritualization of this theme, stressing the hunger and thirst of man's soul, begins almost immediately, and is featured in the New Testament; but the affirmation of God as providing the same physical well-being is never obliterated. The tendency to generalize marks the leadership of God; as he led the wandering Israelites, he leads all men in the pilgrimage of life.

Woven into the episode of the provision of water and food are the themes of faith and disobedience. Israel's complaints against Moses, its fears of death by thirst and starvation and its longing to return to Egypt are all treated as signs of its lacks of full trust in the adequacy of God. As such they are acts of disobedience against him. In the subsequent narratives of the wandering, particularly in the account of the spies (Num. xiii–xiv), this lack of faith explains the long delay in the conquest of Canaan. And in the Old Testament as a whole, lack of faith in God becomes the fundamental sin that causes postponement in the fulfillment of all his promises.

Sealing of the Covenant at Sinai.—Ex. xx, 2–17 consists of the Ten Commandments. In their original form they were all equally brief, though some now have added explanatory detail. In form they differ from the law code given in xxi–xxiii, 19. While the provisions of the latter are given in the form of "case law" (*e.g.*, "when you buy a Hebrew slave . . ."), the Decalogue consists of absolute commands unconditioned by a situational context (*e.g.*, "You shall have no other gods . . ."), a form known as apodictic law. For an account of this form and the history of the Ten Commandments *see* DECALOGUE.

Modern scholarship has popularized the view that the most solemn rite in ancient Israel was a "covenant renewal ceremony." The thesis is that the institution of the covenant was not simply solemnized at a single initial rite but was celebrated annually or, perhaps, every seven years, at a central shrine to which pilgrimage was made. It is held, for example, that Shechem was for long the site for this pilgrimage (Josh. xxiv). An important feature of such a covenant renewal ceremony was the reading of the Law, upon the observance of which Israel's enjoyment of the blessings of the covenant depended.

In this section of Exodus is found what purports to be the initial covenant ceremony, the prototype for the ceremonial renewal of the covenant. There are three main sections. Ch. xix consists of a report of preparation for the manifestation of God, culminating in a theophany (*i.e.*, an appearance of God in visible form; xix, 16–18) indicated by disturbances in nature; scholarly opinion has never agreed as to whether meteorological or seismic phenomena are alluded to. In all probability the metaphors are evoked by Israel's traditional cultic affirmations about the power of God over nature and not just by the physical scene at Sinai, whatever that may have been. Similar reports of theophanies elsewhere (*e.g.*, Judg. v, 4–5; Hab. iii, 3–15) support this view of the matter.

The theophany is the symbol of the dynamic presence of God acting in the absolute authority of his power; it is the cultic equivalent of the event of the exodus and is the real source of meaning for the covenant. The Decalogue and the social code which follow sum up Israel's responsibility in this relationship. The preface to the former—"I am the Lord your God who brought you out of the land of Egypt . . ."—refers specifically to the antecedent divine action as that which gives substance and vitality to the covenant, even though, practically speaking, Israel forfeited its benefits if it failed to observe the commandments.

The Decalogue and code are described as having been spoken by God to Moses (xx, 1; xxi, 1) for communication to the people. In xxiv, 12 and xxxiv, 1, God himself is depicted as inscribing the Decalogue. Both must be understood as ways of saying that Israel's laws, however culturally conditioned, are for the faithful an adequate form for and a sign of the divine will.

In ch. xxiv the initial ratification of the covenant is portrayed by means of two features from Israel's cultic life; the sacrificial ceremony featuring the sprinkling of blood both against the altar and upon the people; and the communal meal featuring the Lord as host.

Supplement.—The sections dealing with the instructions for the building and furnishing of the Tabernacle and with the execution of these tasks are in a sense a supplement to the rest of the book. To be sure, the Tabernacle is associated with that period of the wandering that follows the sealing of the covenant at Sinai. However, the instructions about it, given to Moses, are not set in the context of the covenant ceremonial but following its completion. There is wide agreement that the sections are the work of the late priestly school in Israel. While this school exalted Moses as prophet and lawgiver, it tended to feature Abraham rather than Moses as the real mediator of the covenant. This school felt that in the priesthood of Aaron, in the Temple and its furnishings and in the rites in which these were utilized Israel possessed a divinely appointed means by which its covenant with God was maintained and through which the grace of God was mediated to his people.

The sacred clothing of the priests, the ark with its golden "mercy seat," the veil separating the holy of holies and the great altar stand at the focus of attention. It is assumed that this ritual and the appurtenances for it were in use in Israel from the days of Moses. Therefore the writers take great effort to describe in detail the portable wilderness Tabernacle which is treated as the prototype and model for the Temple of their own day. Especially since it follows the account of the establishment of the covenant, the thematic significance of this unit is most appropriately described as relating to its maintenance.

Exodus in the Bible.—Since it deals directly with the decisive event of God's revelation in history, the role of Exodus in the Old Testament has its counterpart in that of the Gospels in the New Testament. It deals with that moment in history and with those episodes in terms of which all of history is to be assessed and given its meaning. Therefore the historical particulars of Exodus quickly give form to the themes in which faith describes God's relation to

Israel. The 8th-century prophet Hosea describes the imminent divine judgment as a return to Egypt (Hos. viii, 13; ix. 3); and in St. Matthew (Matt. ii, 13–15) the point of the flight into Egypt is the saving action of God illustrated by the exodus and now to be re-enacted in Jesus. In the light of the wandering the wilderness becomes the metaphor for the renewal of the covenant (Hos. ii, 14 ff.). According to reports, Elijah, John the Baptist, Jesus and Paul all went to the desert to receive the confirmation of their vocation; and the Dead sea community, a community of renewal whose ruins and legacy were recovered beginning in the 1940s, was planted in the wilderness for symbolic as well as practical reasons. In the Book of Isaiah (xl, 3–5; xliii, 18–21) the protection and guidance of God are couched in the metaphors of the wilderness wandering. In the New Testament, especially in the Epistle to the Hebrews, the period of Israel's wandering has been transformed into a metaphor that is applied to every individual life. Human life on earth is a pilgrimage; and men are on the march not to a promised land on earth but to a "rest" (Heb. iv, 9 ff.). The whole story of Christian piety has amplified and filled in this pattern begun in the Bible, by which the material events of the founding of Israel as a people in history, recorded in Exodus, became the metaphors for the redemption and spiritual pilgrimage of Everyman.

BIBLIOGRAPHY.—Martin Buber, *Moses* (1946); Jean Daniélou, *Sacramentum Futuri* (1950); Martin Noth, *Das zweite Buch Mose, Exodus*, in *Das alte Testament Deutsch*, vol. v; H. H. Rowley, *From Joseph to Joshua* (1950); J. Coert Rylaarsdam, *The Book of Exodus; Introduction and Exegesis*, in *Interpreter's Bible*, vol. i (1952). (J. C. Rv.)

EXOGAMY, the practice of marrying outside one's own social group, in contrast to endogamy (*q.v.*), which requires marriage within the group. (*See* MARRIAGE, PRIMITIVE.)

In every human society there are certain regulations which control the relation of the sexes and the selection of a mate. Intercourse between close blood relations such as brother and sister, father and daughter, and mother and son is almost everywhere condemned. But among very many peoples it is forbidden both on grounds of consanguinity and because two individuals are members of the same social group. This prohibition, based upon common membership in a social group, is known as the law of exogamy. If the group is a territorial unit, *e.g.*, a village, the exogamy is local. More commonly, membership in the group concerned is determined by kinship, real or fictitious, as in the clan (*see* KINSHIP; CLAN). Exogamy is often used to indicate clan exogamy. That exogamy prevents the marriage of all near relatives is true only if membership in the exogamous group is determined by descent reckoned through both parents, but this is very rare. Normally, descent is traced through only one parent, and it is therefore inevitable that certain close blood relations will not belong to the same group and will therefore be possible mates for each other. Thus, if a tribe is patrilineal a man can select wives from among his mother's sisters, her brother's daughters and her brother's son's daughters, and, were exogamy the only marriage prohibition, even his mother would be available to him. On the other hand, clan exogamy does prevent unions between people bearing no relationship to each other, since membership in a clan is dependent upon fictive, not blood relationship.

The rigidity with which the law of exogamy is observed varies considerably. Among some peoples a breach of it is regarded as incest; among others, though marriage is forbidden, extramarital relations between clan members are tolerated; while in some cases even marriage can be condoned. J. F. McLennan (*q.v.*), who first coined the word, regarded it as the outcome of female infanticide which, by limiting the number of women available within the group, forced tribesmen to capture their wives from their neighbours, but increased knowledge of the facts has made this theory untenable. Another suggestion is that the horror of incest, supposedly innate, has extended to all those women whom, under the classificatory system of relationship, a man addresses by the term for "sister."

On another view the original form was local exogamy arising from a natural distaste on the part of those who have been reared together to cohabit. Others believe it originated with totemism (*q.v.*), while the diffusionists believe that it developed under special conditions in one place and then spread throughout the world.

See also references under "Exogamy" in the Index.

(C. H. W.)

EXORCISM, an adjuration addressed to evil spirits to force them to abandon an object, place or person which they have more or less in their power (*see* POSSESSION). The universal belief of primitive peoples in unseen spirits—some favourable, others unfavourable to them—produced certain rites and practices to propitiate the good, and to ward off or to expel the evil spirits. Though such phenomena might broadly be included under the term exorcism, they belong rather to the field of magic and witchcraft (*q.v.*).

In a more technical sense, exorcism is a ceremony used by the Christian church to expel demons from persons who have come under their power. Jesus himself expelled demons by a word of power, and stated that this act was a sign of the coming of God's Kingdom. His followers, and others as well, drove out demons "in his name." In the first two centuries of the Christian era the power of exorcism was considered a special gift which might be bestowed on anyone, lay or cleric. About the year 250, however, there appeared a special class of the lower clergy, called exorcists, to whom was entrusted this special function. About the same time, exorcism became one of the ceremonies preparatory to the baptism of neophytes, and it has remained an integral part of the Roman Catholic baptismal service. Its presence there does not mean that the church considers candidates for baptism as victims of demoniacal possession; rather exorcism is regarded as a means to remove impediments to grace resulting from the effects of original sin and the power of Satan over fallen nature. In an analogous fashion, inanimate things such as water, salt and oil, destined for use in divine service, receive an exorcism before being blessed or consecrated.

The exorcism of persons possessed by demons is carefully regulated by the canon law of the Roman Catholic Church (can. 1151–52), and the extremely elaborate rite is contained in the *Rituale Romanum*. (The class of exorcists still exists, as the second of the minor orders of clergy, but its duty to "cast out devils" has become obsolete.) A priest may not perform an exorcism without special authorization of his bishop, and such authorization is to be granted only to a cleric noted for his piety, prudence and moral integrity. Before proceeding with an exorcism, the priest must make a careful and thorough investigation to determine whether or not he has to deal with a case of real possession.

BIBLIOGRAPHY.—*Dictionnaire de théologie catholique*, vol. v, col. 1762–86 (1913); J. Tambornino, *De antiquorum daemonismo* (1909); Joseph Coppens, *L'Imposition des mains* (1925); Franz Joseph Dölger, *Der Exorzismus im altchristlichen Taufritual* (1909), *Die Sonne der Gerechtigkeit und der Schwartze* (1918); Celestine Kapsner, *Begone Satan* (1935); S. V. McCasland, *By the Finger of God* (1951); R. M. Woolley, *Exorcism and the Healing of the Sick* (1932).

(V. L. K.; R. McQ. G.)

EXOSKELETON: *see* SKELETON, INVERTEBRATE.

EXPECTANCY, LIFE: *see* LIFE EXPECTANCY.

EXPLANATION. In its widest sense an explanation of any fact is any true answer to the question "Why is the fact what it is?" which provides any sort of intellectual satisfaction. The theoretical function of science is that of providing explanations of natural phenomena by discovering relationships between these phenomena and other events; these relationships fall under general laws, and the intellectual satisfaction provided by science is given by knowledge of the laws according to which the events are "governed." Knowledge of these laws enables us to make predictions as to what events to expect in particular circumstances and sometimes, by controlling the circumstances, to ensure that these events will occur. The practical function of science, that of enabling us to adjust our lives to nature and, sometimes, nature to our lives, thus derives from its intellectual function—that of explaining phenomena by means of scientific laws.

Scientific explanation frequently proceeds in stages. If the fact to be explained is a particular fact, its first-stage explanation consists in the assertion of another fact or set of facts with which the fact to be explained is associated according to a general law.

If this fact or set of facts is about events which precede in time the event in the fact to be explained, these events will be regarded as making up the cause of the later event (and they will sometimes be regarded as making up the cause if the events are all simultaneous). In causal language the first-stage explanation of a fact about a particular event is the assignment of a cause of which the event is an effect, together with the (explicit or implicit) assertion of the causal law governing the cause-effect relationship. Whether it is the discovery of the particular cause or that of the causal law which provides the intellectual satisfaction depends upon the previous knowledge of the person asking the "Why?" question. In the historical parts of a science such as geology, the general causal laws involved are frequently already known as laws of physics; what is then demanded is the particular preceding event which is the cause, according to these known laws, of the event to be explained. But in most of the sciences it is the causal law itself which is the important part of the explanation.

The general laws which occur explicitly or implicitly in explanations of the occurrence of particular events are themselves facts about which "Why?" questions can be asked; and the second-stage explanation of a particular fact consists in an explanation of the law which is involved in its first-stage explanation. Explanation of a law is effected by finding a more general law or set of laws from which it can be deduced. In an advanced science such as physics the various laws are of many different levels of generality, and they are arranged in a hierarchical system in which the lowest-level laws are logical consequences of sets of laws at a higher level, and these again are logical consequences of laws at a higher level still. Thus Newton explained Kepler's laws of the motion of the planets by deducing them from his general laws of motion and law of universal gravitation, and Einstein explained the law of gravitation by showing that it is a close approximation to a consequence of a law on a still higher level concerning the structure of physical space and time.

Theoretical Concepts.—An important feature in nearly all these hierarchical systems is that the concepts ("theoretical concepts") which occur in the higher-level laws of the system are not directly related to observation; they are not objects or properties or relations of observable things as are the concepts with which the lowest-level laws of the system are concerned. The lowest-level laws are general propositions whose instances are directly observable facts; frequently these laws have been established by a simple inductive generalization of such observed facts. Higher-level laws involving theoretical concepts are not simple generalizations of observed facts, since the theoretical concepts which occur in them are not directly observable. The reason for accepting such higher-level laws is that lowest-level laws, which are generalizations of observed facts, can be deduced from them.

Although there is general agreement about this account of the way in which higher-level laws involving theoretical concepts function in a hierarchical scientific system, the status of these theoretical concepts is a matter of controversy among philosophers of science. One school holds that electrons, for example, are real in exactly the same sense as that in which large-scale material objects are real. Explanation of physical phenomena by means of electron theory is then taken to be analogous to explanation of the behaviour of a whole in terms of the behaviour of parts whose nature is similar to that of the whole. These parts are in principle open to observation; that they are in fact not directly observable is a result of deficiencies in our powers of observation. But most philosophers of science would wish to ascribe a status to theoretical concepts different from that of directly observable things and qualities. They would maintain that the status of theoretical concepts is given by the role which these play in scientific systems. The reason for making use of theoretical concepts—namely, that they enter into the higher-level laws from which directly testable lowest-level laws may be deduced—is itself a description of their status. Electrons do not exist independently of the laboratory observations which confirm electron theory; their existence consists in the truth of the theory in which they occur. As F. P. Ramsey said, the statement that there is the quality *mass* is nonsense unless it means merely to affirm the consequences of a mechanical theory. Theoretical concepts are posited, constructed, invented for the purpose of higher-level explanation. The "reality" of electrons derives from the observable facts which they are posited to explain. (This, however, does not imply that it is improper to stretch the meaning of "see" so that the observation of tracks of condensation in a Wilson cloud chamber may be described as seeing electrons.)

That theoretical concepts are in some way constructions out of the data of observation is implicit in the writings of philosophers of science such as Ernst Mach and Karl Pearson; it was made explicit by Bertrand Russell in his doctrine of "logical constructions." Russell held that "the supreme maxim in scientific philosophising is this: Wherever possible, logical constructions are to be substituted for inferred entities" (*Mysticism and Logic and Other Essays*, Longmans, Green & Co. Limited, London, 1918). To say that an electron is a logical construction out of the observed events and objects by which its presence is detectable is, for Russell, equivalent to saying that the word "electron" can be explicitly defined in terms of such observations. On this view every sentence containing the word "electron" is translatable, without loss of meaning, into a sentence in which there occur only words which denote entities (objects, events, properties, relations) which are directly observable. A philosopher of physics should be able to make this translation and thus to exhibit the way in which electrons are logical constructions. Russell's doctrine is similar to the "operationalism" of P. W. Bridgman, according to which theoretical terms must be defined by means of the empirical "operations" involved in their detection and measurement.

There is, however, a very serious objection to the logical-construction doctrine first pointed out by F. P. Ramsey. An explicit definition of a theoretical term by reference to the phenomena it is posited to explain makes it impossible to use the theoretical concept in the future in a wider system in which new phenomena are explained. The great value of most of the historic theoretical concepts of physics (*e.g.*, electric charge, kinetic energy) is that not only did they explain, by their place in higher-level laws, lowest-level laws which were known at the time of their introduction, but also use was made of them in systems propounded at a later date (*e.g.*, electromagnetic theory, kinetic theory of gases) to explain a wider range of phenomena. To treat theoretical concepts as logical constructions out of a limited range of phenomena is to ossify the scientific system in which they occur; such treatment would from the outset abandon any hope of extending the system to cover different sorts of phenomena, or even different generalizations about the same sort of phenomena. If a scientific system is to be capable of growth, more freedom of play must be allowed to its theoretical concepts than is permitted by the logical-construction doctrine.

The difficulty is avoided if a looser kind of definition is given of a theoretical term. Instead of defining it explicitly in terms of observable entities, it must be "contextually" defined by the way in which it functions in the language (the "calculus") by which the scientific system is expressed. The calculus representing the system consists of a series of sentences or formulas in which the first few formulas (the "initial formulas") stand for the highest-level laws containing the theoretical concepts; formulas occurring later in the series stand for laws at a lower level, the final formulas standing for the lowest-level laws which are empirically testable. The rules by which later formulas are derived from the initial formulas within the calculus correspond to the logical principles in accordance with which the lower-level laws expressed by later formulas are deduced from the highest-level laws expressed by the initial formulas. Thus the deductive logic of the system of laws is reflected in the rules of manipulation of the calculus representing the system, and the status of theoretical concepts is given by the way in which the calculus containing the theoretical terms is interpreted.

Calculi representing deductive systems of pure mathematics are interpreted all of a piece. Meanings are attached to the symbols or words occurring in the initial formulas, and the same meanings are then attached to these symbols throughout the system. Euclid, whose system of geometry originated the notion of deductive sys-

tem in European thought, gave explicit definitions of the terms ("point," "straight line," etc.) occurring at the beginning of his system, so that it is not necessary to understand the theorems of the later books of Euclid before embarking on his first book. But in calculi which represent scientific deductive systems containing theoretical concepts the interpretation of the final formulas comes before that of the initial formulas containing theoretical terms. Meanings are first attached to the symbols denoting the directly observable properties and relations which occur in the lowest-level laws, the laws which are directly testable; meanings are subsequently attached to the symbols which are to denote the theoretical concepts by virtue of the fact that these symbols occur in the initial formulas of a calculus whose final formulas have been interpreted as expressing lowest-level laws. This procedure does not give an explicit definition of a theoretical term such as "electron"; instead it gives a contextual definition of it by displaying its place within the calculus representing the whole scientific system. The "context" for the contextual definition of "electron" is not only the sentences in which it occurs, but the whole system of sentences (the calculus) which is being taken to represent electron theory. If the context alters (*e.g.*, by electron theory's being found to account for new phenomena), the meaning of the term "electron" will automatically be adjusted to its way of functioning within the calculus representing the extended theory. Thus this account of the status of theoretical concepts enables them to retain their place in a growing science, a place which has to be abandoned only when the new scientific system is radically different from the old one.

The Use of Models in Explanation.—Explicit attention to the calculus used to represent a scientific system throws light upon the use of models in scientific thinking. Any calculus may be interpreted in many different ways, and a model for a scientific system is a second interpretation for the calculus representing the system. In this second interpretation the initial formulas of the calculus are taken to stand for propositions the understanding of which is independent of the scientific system in question. These propositions will either be concerned with directly observable entities or, if they contain theoretical concepts, these will be those of a less sophisticated scientific system (*e.g.*, that of mechanics) whose functioning is already understood. Later formulas in the calculus will then represent deduced propositions in the model system. Since both the model system and the scientific system are interpretations of the same calculus, the deductive structure of the model will reproduce that of the scientific system. The logical difference between the model and the system is that in the former the initial propositions can be understood before any deductions are made from them, whereas in the latter the initial propositions (the highest-level laws) are comprehensible only by virtue of what can be deduced from them.

This fact it is that makes models so useful in understanding scientific systems which contain theoretical concepts. For the place of these concepts in the system may be thought of by analogy with the place of their analogues in the model, and this will avoid the self-consciousness required in order to think about the calculus which represents indifferently both the system and the model. Thinking of the model may therefore be substituted for thinking explicitly about the calculus; the theoretical concepts of the system may be regarded *as if* they were elements occurring in the model for the purpose of considering their place in the system. To the extent that it is true to say of scientific explanation that it "describes the unfamiliar in terms of the familiar," it is this use of models that is envisaged.

There are two errors, however, that must be avoided in thinking of a scientific system by means of a model. The first is that of supposing that there must be features in the system which correspond to all the features of the model; *e.g.*, if a "solar system" model of an atom is used, of supposing that the spatial properties and relations of the electrons and the nucleus must be analogous to the spatial relations of planets to the sun. This error can be avoided by always bearing in mind that it is only the deductive structure of the scientific system which is reproduced in the model. A more subtle error arises from the fact that, if a geometrical or other mathematical model is used (which is frequently very convenient), propositions in the model system which correspond to highest-level laws may well be logically necessary theorems of pure mathematics. There will then be a strong temptation illicitly to project the logical necessity of these propositions onto the highest-level laws and to conclude that some or all of the highest-level laws of the system are logically necessary instead of empirical and contingent. But logically necessary propositions occur only among the highest-level laws of a hierarchical scientific system when an excessive number of theoretical concepts is being used; if redundancies among theoretical concepts are removed, all the laws of the system at all levels will be contingent propositions, since a *scientific* system is an empirical system grounded upon observation. The user of a mathematical model must therefore never forget that he, like the user of a "familiar object" model, is engaging in *as-if* thinking. The only feature of the model which is relevant to the understanding of the scientific system of which it is a model is its deductive structure; everything else—the logical status of its initial propositions as much as the nature of the familiar objects entering into these propositions—is irrelevant.

Types of Explanatory Laws.—The view here put forward of scientific explanation in terms of an ascending hierarchy of laws of increasing generality imposes no restriction upon the type of general law which may serve as an explanation. Highest-level laws may, for instance, be statistical laws asserting not that all A's are B's but that, by and large, a certain proportion of A's are B's; *i.e.*, that the probability of an A being a B is a certain fraction. Lower-level statistical laws came into physics with the kinetic theory of gases; when they were introduced it was thought that they required to be explained by higher-level nonstatistical ("deterministic") laws. The irreducibly statistical ("indeterministic") character of the highest-level laws is one of the most striking features of quantum physics; since this system is a closely integrated hierarchical system, it would seem to be mere prejudice to deny that its highest-level laws were explanatory on the ground that they were not of the deterministic form of the laws of classical physics.

A type of law presenting more difficulty for the view presented here are the laws frequently put forward in sciences less advanced than physics, laws expressed in the form of "tendency statements," which assert that so-and-so has a tendency to act in such-and-such a way. (An example is the psychological law that the strength of an association in a person's mind tends to increase the more frequently he has experienced it.) Such a tendency statement does not assert a general law ("every A is B") which would be refutable by the discovery of a contrary instance (a case of A which is not B); what it asserts is that, under certain conditions which are not specified, every A is B. Since the conditions are unspecified, the statement, properly speaking, does not express a general law. Nevertheless quasi laws expressed by tendency statements can be arranged in a hierarchical system with more general quasi laws (which may contain theoretical concepts) at the highest level, the whole hierarchical system being understood as being subject to the same unspecified conditions. These conditions may become more specified (and the specified parts can then be transferred into the body of the system) as relevant observational evidence increases. A tendency quasi law in isolation gives little explanatory satisfaction; but a large number of quasi laws, all subject to the same unspecified conditions and arranged in a coherent system, provide a partial explanation in a science in which nontendency laws have not yet been found.

Teleological Explanation.—A "Why?" question asked about a particular event is sometimes answered by the specification of a goal or end toward the attainment of which the event is a means. Objections are frequently made to such teleological explanations on the ground that they explain an event by reference to an event later in time (*e.g.*, the nest-building activity of a bird by reference to its later laying eggs when the nest is built); and it has been held that an explanation of behaviour as being "goal-directed" is inadequate unless the goal is being consciously pursued—unless, in Aristotelian language, the idea of the "final cause" is functioning as the "efficient cause." Since it would be rash to assume such

conscious intentions in the lower animals, it is frequently held that, whatever may be the case for psychology, teleological explanations are useless in the other biological sciences.

There is no doubt, however, that a teleological explanation provides intellectual satisfaction, even when no conscious intentions can be assumed. And since a first-stage teleological explanation of a particular fact can frequently be succeeded by a second-stage explanation of the teleological law used in the first-stage explanation, teleological laws can frequently be arranged, like nonteleological laws, in a hierarchical system. The teleological law involved in a teleological explanation of a particular nest building—that nest building is directed toward the goal of providing a place for laying eggs—can be subsumed under a more general teleological law relating both these activities to the end of propagating the species. When a biologist says that all organic activities are related to the biological ends of maintenance, development and reproduction, he is asserting highest-level teleological explanations which stand at the summit of a hierarchy of teleological explanations at a lower level. The allegation that teleological laws are mere descriptions and not proper explanations has equally well been made against the nonteleological laws of physics. But, since scientific laws of all sorts enable us to organize our thinking into coherent systems as well as to make predictions about the future, to call any scientific laws merely descriptive, with the implication that they merely describe in summary form the past observations upon which they have been based, would be grossly misleading.

Alternative Explanations.—One reason for the belief that teleological explanations should be rejected in favour of nonteleological ones is the belief that different types of explanation are mutually exclusive. But if the process of successive explanation is viewed as the erection of a hierarchy of laws of increasing generality, there is no reason to prevent different hierarchies from being constructed in different ways to explain the same phenomena.

These different hierarchies may employ different sorts of theoretical concepts. A slip of the tongue, for instance, may be explicable both in a psychological system containing theoretical concepts such as "unconscious wish" and in a physicochemical system in which neurological laws are brought under higher-level laws containing theoretical concepts of physics. Alternative explanatory systems are mutually incompatible only if they deduce incompatible lowest-level laws from their highest-level laws. When two different systems explain the same phenomena, our desire for unity will probably induce an attempt to relate together the different theoretical concepts and ways of thinking of the two systems; for instance, by considering teleological laws as being concerned with processes in which the same end result is attained under varying conditions and comparing such processes with mechanical processes in which such "plasticity" is obtained by purely mechanical means (as by feedback devices in pilotless aeroplanes). But even if the two systems cannot be synthesized, each will properly serve as an explanation, each of its own sort.

Different sorts of explanations have been favoured at different times in the history of thought. Before the scientific renascence of the 17th century, theological explanations in terms of the intentions of a Creator were dominant; in modern times the great success of explanatory systems whose highest-level laws are mechanical or (later) the most general laws of physics has given enormous prestige to explanatory systems of this sort. Part of this prestige is attributable to the fact that, to a large extent, the theoretical concepts used in physical systems are the notions of molecule, atom, electron, in general of "elementary particle"; the systems explain lowest-level laws by means of laws about the combinations and inter-relations of these elementary particles. Systems with atomistic theoretical concepts both lend themselves to "familiar object" models and satisfy the desire for persistence and permanence of the explanatory fundamental objects, a feature to which philosophers of science such as Émile Meyerson attached great importance.

The Alleged Defectiveness of Scientific Explanation.—"Why?" questions asked of lower-level scientific laws, which are answered by finding higher-level laws from which they can be deduced, can prima facie be asked of the highest-level laws themselves. The fact that such questions are unanswerable (except by incorporating the system in a more general system whose highest-level laws would seem to invite similar unanswerable questions) has appeared to many philosophers to show that scientific explanation by an ascending hierarchy of laws is inherently defective. Explanation, it is said, must proceed in a descending hierarchy from "first principles" whose "rationality" is not dependent upon their place in the system itself. But to demand this species of explanation from science is to require that scientific explanation should be regarded as analogous to the justification of legal or moral laws or to the proof of theorems of pure mathematics. We justify a legal or moral law by showing that it falls under a general legal or moral principle the authority for which is already accepted, and we prove a mathematical theorem by deducing it from accepted axioms. But scientific laws are not legal or moral or mathematical; they are empirical. The reason for believing higher-level laws in science is not that they are deducible from highest-level "first principles" but that lowest-level laws, directly testable by experience, are deducible from them. The difference of scientific explanation from other species of explanation arises from the empirical nature of science, whose authority lies not in any "rational" first principles at the summit of a descending system but in the observable facts upon which the system of ascending explanations is based. *See* SCIENTIFIC METHOD.

BIBLIOGRAPHY.—R. B. Braithwaite, *Scientific Explanation* (1953); E. Meyerson, *De l'explication dans les sciences*, 2 vol. (1921); F. P. Ramsey, *The Foundations of Mathematics and Other Logical Essays* (1931). (R. B. B.)

EXPLOSIVES are substances capable of exerting sudden pressure on their surroundings as a result of a rapid conversion of the substance into hot gases or volatile substances that have a larger volume at the surrounding pressure than the volume of the original explosive. Since in the instant of their formation the gases occupy only the volume of the explosive, they are at higher than the normal pressure. Their pressure, which is raised by the generation of heat in the course of the explosion, overbalances the restraining pressure of the surrounding matter. Rapid expansion follows, and this constitutes the explosion.

The study of heat changes accompanying chemical reactions, or the science of thermochemistry (*q.v.*), shows that practically every chemical change is accompanied either by evolution or by absorption of heat. Consequently, if the exchange of heat between a reacting substance and its environment is delayed, the temperature of the reacting substance either increases or decreases. The former type of chemical change is called exothermic; the latter, endothermic. In general, only the former type includes explosively fast processes, because chemical reactions are nearly always accelerated by heating and delayed by cooling. Thus, an explosive must fulfill the following basic requirements: (1) the explosive must contain a substance, or a mixture of substances, which remains unchanged under ordinary conditions but undergoes a fast chemical change upon a proper stimulation; (2) this reaction must yield gases whose volume under normal pressure but at the high temperature resulting from explosion is much greater than that of the original substance; (3) the change must be exothermic in order to heat the products of the reaction and thus to increase their pressure. All types of chemical reactions may give explosions. Thus, acetylene (*q.v.*) decomposes explosively into its constituent elements, carbon and hydrogen. The elements carbon and oxygen, in the form of charcoal saturated with liquid oxygen, may be explosively combined into carbon dioxide. The explosion of nitrated organic compounds, as for instance of TNT or of nitrocellulose, consists in the transformation of the original compound into several new ones. In the nuclear explosions of atomic bombs, the events are not limited to the interaction of chemical molecules. Here the nuclei of the atoms may be ruptured, in a process called fission, to form atoms of lighter elements, or nuclei may combine, in a process known as fusion, to form heavier elements. Such processes are accompanied by the evolution of heat many millions of times greater than that characteristic of chemical processes; hence, the nuclear weapons have immensely greater destructive power than any chemical explo-

sives. Nevertheless, all explosive processes have the common characteristics of suddenly producing heat and gases which exert intense pressure on the surroundings. The more heat and gases an explosive produces per unit volume, the more powerful it is.

The examples cited above suffice to show that the physical state of the original explosive is not decisive: explosives may be gases, liquids or solids. However, dense explosives (*i.e.*, solids and liquids) are in general more violent because they may evolve more heat and produce a larger amount of gas per unit volume.

Classification of Explosives and Regulation of Their Uses.—Explosives may be conveniently divided into propellent explosives and detonating explosives. The detonating explosives may be further classified as high explosives and initiating or primary explosives. The distinction is not sharp; it is that primary explosives, upon ignition, explode or detonate almost instantly, whereas high explosives may be burned without explosion. To detonate the latter instantly it is necessary to subject them to a violent shock; for instance, a mechanical blow or the explosion of an adjacent charge. High explosives may be usefully subdivided into military high explosives and blasting explosives for commercial uses, the former being usually pure or mixed organic nitrated compounds and the latter being less violent mixtures of explosive and nonexplosive ingredients (dynamites).

In the United States the manufacture, transportation and storage of explosives is regulated by the Interstate Commerce commission with the co-operation of the bureau of explosives of the bureau of mines and of the Institute of Makers of Explosives. For legal purposes explosives are classified as follows: blasting (black) powders, dynamites, other high explosives, initiating explosives (detonators), military explosives, shotgun, rifle and common powders. It is unlawful to ship and store explosives without a permit. In Great Britain the handling of explosives has been regulated under the Explosives act of 1875, the home office having the regulatory powers. The Explosives act classifies explosives in seven groups, according to possibility of ignition by fire, friction, chemical decomposition, etc.

Some Common Properties of Explosives.—All explosives undergo chemical change when heated to sufficiently high temperatures. In practice an explosive which undergoes a rapid decomposition at 100° C. is considered rather unstable thermally and unsafe; whereas a material which may be heated to 200° C. without the occurrence of rapid reaction is considered very stable. The character of the reaction that occurs on heating depends on the conditions of heating and on the substance tested. In general, more rapid heating causes more violent chemical reactions. If the rate of heating is in excess of something like 1° per minute, primary explosives, such as mercury fulminate (*q.v.*) and lead azide, remain substantially unchanged until a temperature is reached at which they explode suddenly and violently. Other substances under the same conditions of heating may either decompose gradually (TNT) or ignite without violent explosion (nitrocellulose propellants). The gradual decomposition of most, but not all, explosives is marked by the evolution of brown fumes (nitrogen dioxide). When this evolution is observed, the explosive should be immediately destroyed by an expert since it has become wholly unsafe to store and handle.

All explosives may be exploded instantly by a sufficiently violent mechanical blow. When a small pinch or drop of an explosive is placed on a steel anvil, a gentle tapping with a steel hammer may explode mercury fulminate and even lead azide. On the other extreme, TNT and particularly ammonium picrate require a violent blow with a sledge hammer to explode.

Explosives may be exploded or ignited by an electric spark. Such substances as mercury fulminate, dry nitrocellulose and black powder may be exploded even by the feeble spark from a charged human body, but TNT and ammonium picrate require a strong electric discharge to be ignited. Between these extremes range again the other common explosives.

An open flame instantly explodes primary explosives, but only ignites high explosives. However, all of these explode, if present in sufficiently large quantities, after burning for some time with increasing violence.

The common tests on explosives include the determination of the following properties: thermal stability; sensitivity to mechanical impact; ease of initiation (expressed as the serial number of the smallest detonator which ensures consistent initiation); power (measured by the so-called Trauzl lead-block expansion or by the ballistic mortar); heat of explosion (measured in a calorimeter; the common values range from about 700 cal. per gram for black powder, low-potential gun propellants and TNT to more than 1,500 cal. per gram for PETN or nitroglycerin); volume of gas formed (reduced to 0° C. and 760 mm. pressure, the values ranging from about 270 c.c. per gram of black powder to more than 1,000 c.c. per gram of the very powerful organic explosives like Cyclonite).

Deflagration and Detonation.—As noted before, explosives are capable both of burning (or deflagration) and of explosion (or detonation). By the first is meant a reaction which is accompanied by a flame, and which may be violent and rapid, but during which the pressure within the (granular) mass of the explosive remains substantially equal to the pressure outside. A typical deflagration is that of smokeless powders in guns. Each grain of the powder burns simultaneously with other grains, and the pressure within the cartridge is substantially uniform. The deflagration of an explosive may change into detonation. The most characteristic property of detonations is that the chemical reaction does not occur simultaneously over the entire mass of an explosive charge, but spreads like a wave from the point of initiation. While this is going on, the explosive forward of the detonation wave is wholly unchanged, whereas the regions already passed by the wave contain substantially converted material. Modern studies have shown that the reaction of an explosive in the detonation wave occurs so rapidly that the change is completed in a few millionths of a second. Certain secondary chemical reactions occur for a considerable time after the passage of the detonation wave, but they are not essential to the process of detonation, being caused by the expansion and cooling of the products of explosion. Different explosives vary greatly in the ease with which burning changes into detonation. In primary explosives the change is almost immeasurably fast, this fact being the reason for the use of such substances in detonators. Military high explosives and dynamites may burn in the open for several minutes before deflagration changes into detonation. Smokeless powders, finally, are extremely difficult to bring to detonation; this property is intentionally accomplished by the special methods of their preparation (colloiding) from readily detonating ingredients, nitrocellulose and nitroglycerin.

Although superficially the deflagration and the detonation differ markedly, the chemical processes involved in both are substantially the same. The rate of deflagration of all explosives that have been tested in this respect is accelerated by rising pressure. Detonation may be said to be a deflagration in which the pressure in the locus of the wave is so high that the chemical reaction completes itself before adjacent layers of the substance are set into reaction. The accelerating effect of pressure causes a confined explosive to detonate more readily than one in the open. Thus, an explosive trapped between the hammer and the anvil detonates on impact, although the same material may burn on the application of a lighted match. In both cases the initial event is a thermally induced ignition, but in the former the gases thus formed cannot escape; the pressure, therefore, rises, and the deflagration accelerates to such an extent that it changes into a detonation within a minute fraction of a second. On ignition in the open, the gases escape as soon as they are formed and so the material continues deflagrating at a relatively slow rate.

The detonation of typical pure organic nitrated explosives, such as are used largely for military purposes, obeys rather simple laws. Quickly after the initiation, the detonation rate reaches a definite value which is maintained as long as the detonation travels through a uniform charge. Except in small charges, this stationary detonation rate depends only on the chemical composition and the bulk density of the explosive. In granular TNT, for instance, the rate is 5,030 m. per second at a density of 1 g. per cubic centimetre, and 6,630 m. per second at 1.5 g. per cubic centimetre;

it depends linearly on the density in the entire density range investigated. In other military explosives the detonation rates are of the same order of magnitude, Cyclonite having probably the highest known velocity of about 8,100 m. per second at 1.6 g. per cubic centimetre density.

The detonation of gaseous mixtures obeys even simpler laws, being dependent only on the chemical composition and not on pressure. The rates are somewhat slower, ranging from about 1,000 to 3,000 m. per second.

Detonations of solid mixtures of the types known as dynamites (*i.e.*, of mixtures containing essentially nonexplosive ingredients like wood meal, sodium nitrate, etc.) show more complex behaviour. The rates are, in general, slower than those of pure nitrated explosives and depend on additional factors, such as the particle size of the ingredients (being usually slower in coarser materials), the confinement of the charge and its size (being higher in bigger and more confined charges). Such effects are observed with pure nitrated explosives only in very small charges.

Theoretical and experimental studies have shown the detonations to be special cases of shock waves. The latter are acoustic waves of finite amplitude with a steep front; because of these characteristics, the laws of their propagation differ somewhat from the laws governing the propagation of ordinary acoustic waves. In particular, shock waves travel faster than acoustic waves, the more so the higher their amplitude is. The intensity of a shock wave decreases faster as it passes through a medium (such as air) than does that of an acoustic wave, because the energy of the shock wave is being employed to heat the medium. Thus, for instance, the amplitude of a spherical acoustic wave decreases inversely as the distance of travel, but the amplitude of a strong shock wave, such as is created in air by an explosive charge, decreases almost as the inverse square of the distance until the wave has become so weak that its laws of propagation agree with those of acoustic waves. To maintain a shock wave at constant amplitude and constant velocity it is necessary, therefore, constantly to feed new energy into it. Detonation waves are shock waves of constant amplitude and velocity. The energy which supports them is the energy released by the chemical reaction of the explosive.

The general mathematical theory of detonation and shock waves was founded by researches of H. Hugoniot, D. L. Chapman and E. Jouguet in the 19th and early 20th centuries, but only during World War II was it brought to a final quantitative formulation. The detonation velocities can now be calculated theoretically and their dependence on such factors as the chemical composition of the explosive, its density, etc., predicted in good agreement with experimental data. This constitutes strong evidence that the theory is essentially correct. The theory reveals that, in detonation waves of dense explosives, the temperatures and pressures reach values as high as 5,000° C. and 200,000 kg. per square centimetre, respectively. This explains why even the strongest materials are easily ruptured by explosives (the ultimate tensile strength of even the best steels is of the order of 15,000 kg. per square centimetre). The basic theory of detonation waves assumes that the chemical reaction occurs instantly as the wave passes through the explosive. The excellent agreement of the theory with experimental observations on gaseous mixtures and on pure nitrated (military) explosives shows that in these the chemical reaction is rapid. The more complex behaviour of detonation waves in such mixed explosives as dynamites is attributable to slower chemical reactions of the nonexplosive ingredients of these mixtures. When, finally, an explosive is made by mixing several stable and nonexplosive ingredients only, as in black powder, the chemical reactions become too slow for a real detonation to be possible, although by fine granulation, etc., black powder can be made to explode fairly violently.

COMPOSITION AND PROPERTIES OF SOME EXPLOSIVES

Initiating or Primary Explosives.—Besides mercury fulminate (*q.v.*) and lead azide, lead trinitroresorcinate [$C_6H(NO_2)_3(O_2Pb)$] and diazodinitrophenol (DDNP, $C_6H_2N_4O_5$) also find extensive uses as initiating explosives. All of them are easily detonated by impact and quickly change an ignition into a detonation. Because of these characteristics they are not suitable for use as the main blasting or propelling charge.

Military High Explosives.—Explosives of this type must satisfy the requirements of great safety in use with exceptionally high performance, almost regardless of cost. Experience has shown that pure and mixed organic nitrated explosives, with occasional additions of nonexplosive ingredients for special purposes, satisfy best the military requirements. The most common military explosives are trinitrotoluene (TNT), trinitrophenylmethylnitramine (tetryl), trinitrophenol (picric acid, *q.v.*), ammonium picrate, pentaerythritoltetranitrate (PETN) and cyclotrimethylenetrinitramine (Cyclonite). Most of these materials were discovered in the 19th century, but their explosive behaviour was utilized only later when the technique of initiating the detonation of high explosives by means of detonators became well understood. The first extensive use of these materials (*i.e.*, picric acid and wet nitrocellulose) took place in the Russo-Japanese War of 1905, black powder being the accepted military bursting charge prior to that time. In World War I, TNT became the standard explosive but, because of its shortage, extensive use was made also of less satisfactory compounds and mixtures. In particular, the mixtures of TNT with ammonium nitrate, known as amatols, were widely used. In World War II mixtures of TNT with PETN and Cyclonite, plus in some instances aluminum powder, were introduced. The resulting mixtures have much higher performance than pure TNT; in particular, the large amount of heat accompanying the oxidation of aluminum improved these explosives. From the economic point of view, loading of munitions by pouring a molten explosive is advantageous and has replaced other techniques except in special cases.

Dynamites.—For a long time gunpowder was the only explosive applied to industrial blasting operations, but toward the middle of the 19th century liquid nitroglycerin began to be used on a limited scale. However, the considerable hazards accompanying its manufacture and use, as well as the imperfect method of initiation by ignition then practised, sharply limited its usefulness until the important discoveries by Alfred Nobel in the 1860s of dynamites and of the blasting cap for the initiation of high explosives. In the period following Nobel's pioneering investigations, a great many explosive mixtures suitable for industrial blasting work were patented. The following survey will indicate only the main types of these explosives. Most of them contain nitroglycerin or other similarly behaving nitric esters, which are the sensitizers ensuring the detonation of the mixtures. Some mixed explosives, particularly those containing chlorates or ammonium nitrate and aluminum powder, detonate satisfactorily without nitroglycerin. In general, the detonation rates of dynamites range between about 2,000 and 4,000 m. per second, but, as noted before, they depend upon the size of the charge and upon the confinement. The ability of dynamites to detonate appears to depend in a large degree on their porous, granular structure. When dynamites are highly compressed (as may happen in a wet borehole because of the hydrostatic pressure of water on the dynamite cartridge) the detonation is likely to fail. The detonation rates of most dynamites, particularly those containing much ammonium nitrate, can be extensively varied by changing the particle size of the ingredients. Coarse-grained dynamites detonate more slowly and show a less shattering or less *brisant* action. Thus, by proper compounding and granulation, it is possible to prepare explosives either to shatter hard rock or to break even soft minerals such as coal into large lumps.

Guhr Dynamite.—The first practical dynamite patented by Nobel consisted of a porous nonexplosive absorbent, such as charcoal or kieselguhr, with which nitroglycerin was mixed in such proportions that the resultant material was essentially dry and granular in structure. Being relatively insensitive to shock, dynamite made the commercial use of nitroglycerin practicable. Guhr dynamite is not manufactured in the U.S., but is still extensively used in Europe.

Straight Dynamites.—A few years later, Nobel patented dyna-

mites in which nitroglycerin was mixed with other substances which, although not explosive separately, formed together an active base capable of sustaining explosion. Typical mixtures included sodium nitrate; a carbonaceous material such as wood meal, charcoal, etc.; and nitroglycerin, with a small addition of calcium carbonate to act as an antiacid. Such combinations became known in the U.S. as straight dynamites. In later years the use of pure nitroglycerin was largely discontinued because of its relatively high freezing temperature and the resultant adverse effects on dynamites of storage in the cold. Instead, the nitrated mixtures of glycerin and closely related polyhydric alcohols, including sugars, were adopted; in this manner, low-freezing dynamites are manufactured. The standard 40% dynamite used by the U.S. bureau of mines in its comparative tests contains nitroglycerin 40%, sodium nitrate 44%, wood pulp 15% and calcium carbonate 1%. Other suitably modified compositions make 20% or 60% dynamites. Still further compositions are classified by an indication of the standard composition with the same performance; *e.g.*, 40% dynamite defines any composition which has the same performance as the standard 40% dynamite, although it may contain a different percentage of nitric esters and a different active base.

Blasting Gelatin.—In 1876 Nobel patented blasting gelatin and gelatin dynamites in the U.S. Seven to eight per cent of collodion nitrocellulose dissolved in nitroglycerin converts the latter into a rather stiff jelly capable of detonation upon a suitable initiation by a blasting cap. It may be noted that, as the percentage of nitrocellulose and hence the stiffness of the gel increases, the ease of initiation decreases, until in double-base gun propellants it is lost almost altogether. Blasting gelatin is the most powerful and *brisant* of industrial explosives, detonating with rates reaching 8,000 m. per second. It is not widely used in the U.S., partly because of its relatively high cost and partly because of its *brisance* and sensitivity to ignition by flame.

Gelatin Dynamites.—The substitution of blasting gelatin for nitroglycerin in dynamites leads to a series of gelatin dynamites,

TABLE I.—*Gelatin Dynamites*

	Great Britain		United States	
			Gelatin dynamites	
	Gelatin dynamite	Gelignite	40%	60%
Nitroglycerin	74.1	61.1	33.0	50.0
Nitrocellulose	5.9	3.8	1.0	1.9
Potassium nitrate	15.7	27.6	..	..
Sodium nitrate	..	..	52.0	38.1
Wood meal	4.0	7.2	..	..
Wood pulp, flour, etc.	..	..	13.0	9.0
Chalk	0.6	0.3	1.0	1.0

widely used in the U.S. and Great Britain. For these preparations a lesser proportion of nitrocellulose is employed to make a softer gel than that used in blasting gelatin. Table I indicates some typical compositions. The main advantage of gelatin dynamites over straight dynamites is a much smaller tendency to exude the liquid nitric esters, a feature which contributes to safety in their use.

Ammonia Dynamites.—Industrial explosives containing ammonium nitrate were invented also in the 1860s and have since become probably the most popular type. They are known under a great variety of names, such as extra dynamite of Nobel, ammonia dynamites, monobels, etc. Table II indicates some of the typical compositions.

TABLE II.—*Ammonia Dynamites*

	United States			Great Britain		France	
	Ammonia dynamites		Mono-bel	Rex pow-der	Robu-rite	Ammonia dynamites	
	40%	60%				40%	60%
Nitroglycerin	21	33	10	12	..	40	20
Ammonium nitrate	20	30	70	59.5	60.5	45	75
Sodium nitrate	45	27	..	..	..	5	..
Combustible material	13	9	10	7.7	..	10	5
Sodium chloride	..	..	..	19.5	..	..	..
Chalk	1	1	..	..	..	..	..
TNT	..	..	..	..	16.5	..	..
Ammonium chloride	..	..	10	..	22.5	..	..

Other Explosives.—*Ammonals.*—The large amount of heat released by the oxidation of aluminum led to the introduction of aluminum powder mixed with ammonium nitrate as an industrial explosive. Typical compositions are found in Table III.

TABLE III.—*Ammonals*

	I	II	III	IV
Ammonium nitrate	72	72	47	94
Aluminum powder	16	25	22	3
Charcoal	..	3	1	3
TNT	12	..	30	..

Chlorate and Perchlorate Explosives.—Many explosives containing alkali chlorates and perchlorates have been patented. The chlorate explosives are usually highly sensitive to impact and friction. Some typical compositions are shown in Table IV.

TABLE IV.—*Chlorate and Perchlorate Explosives*

	Cheddite 60 No. 4	Perko-ronit B	Dynobel	Per-chlorate cheddite
Potassium chlorate	70	..	..	..
Sodium nitrate	..	..	..	30
Potassium perchlorate	..	59	27	..
Ammonium perchlorate	..	..	..	50
Wood meal	..	..	9.5	..
Nitronaphtbalene	1	..	..	..
Dinitrotoluene	15	31	..	15
Castor oil	5	..	..	5
Ammonium nitrate	..	10	..	..
Nitroglycerin	..	..	32.5	..
Ammonium oxalate	..	..	29.5	..

Sprengel Explosives.—These explosives embody the principle of mixing the oxidizing agent and the combustible *in situ* or, at any rate, immediately prior to use. Probably the best-known one is rackarock, which was used to blast out the Hell Gate channel in the East river in New York city, and which was made by pouring nitrobenzene over potassium chlorate contained in watertight cartridges. Other oxidizing agents, such as nitrogen dioxide or concentrated nitric acid, have also been recommended and occasionally used.

Liquid-Oxygen Explosives (LOX).—These explosives, which form a special case of Sprengel explosives, were invented by H. Linde in 1895. They consist of charcoal, carbon black or some other porous combustible material soaked in liquid oxygen just prior to use. Their great advantage, as that of other Sprengel explosives, is that explosive hazards begin only just prior to use. The liquid-oxygen explosives have the additional virtue of becoming safe again after a few hours, when the liquid oxygen has evaporated. They suffer from their extreme sensitivity to ignition, but a pretreatment of the carbonaceous absorbents with phosphoric acid appears to reduce this hazard considerably.

Blasting Powders.—Black powder, usually containing sodium rather than potassium nitrate and a large proportion of sulfur, is extensively used for mining operations. The inability of black powder to transmit real detonation has been ingeniously overcome by passing a strand of Primacord through the centre of a long charge of blasting powder. This ensures a rapid propagation of explosion along the charge, with a speed characteristic of Primacord, but avoids the shattering action of high explosives.

Permissible Explosives.—So many disastrous explosions in coal mines have been caused by the explosions of firedamp (methane mixed with air) and of coal dust, initiated by blasting charges, that governmental action to regulate the use of explosives in coal mines was long ago taken in most countries. Extensive investigations, carried out by special commissions in some countries and in the United States by the U.S. geological survey and the U.S. bureau of mines, led to the formulation of certain rules to which explosives must conform in order to be permitted for general use in coal mines. Such explosives are called *permissible* in the U.S., *permitted* in Great Britain, *antigrisouteux* in France, etc. It appears that an explosive producing a hot and long-lasting flame is much more likely to ignite firedamp and coal dust than one with a cool flame of short duration. Blasting powders,

explosives containing large proportions of aluminum powder, and hot explosives rich in nitroglycerin are particularly hazardous from this point of view. Large proportions of ammonium nitrate tend to make the explosive cool; this effect is particularly enhanced by additions of inert salts, such as sodium chloride, ammonium oxalate, etc. Among the explosives listed above Monobel, Rex powder, Ammonal No. 4 and Dynobel are characteristic of permissible explosives. A liquid carbon-dioxide explosive has been introduced for use in the U.S. where extreme hazards of firedamp explosion exist. Liquid carbon dioxide is compressed in a cylindrical steel shell provided with a shear disk at one end. A cartridge containing sodium chlorate, aluminum powder and charcoal, with means for firing it electrically, is located in the centre of the shell. The firing of the cartridge volatilizes carbon dioxide and raises the internal pressure in the shell to about 20,000 lb. per square inch, whereupon the disk is sheared and the gases burst out.

See AMMUNITION, ARTILLERY; BLASTING; PROPELLANTS; *see* also references under "Explosives" in the Index.

BIBLIOGRAPHY.—A. Marshall, *Explosives*, 3 vol., 2nd ed. (1917–32); A. Stettbacher, *Die Schiess-und Sprengstoffe* (1919); P. Naoûm, *Nitroglycerine and Nitroglycerine Explosives*, trans. by E. M. Symmes (1928); T. L. Davis, *The Chemistry of Powder and Explosives*, 2 vol. (1941–43); Sir Thomas E. Thorpe, *Dictionary of Applied Chemistry*, 4th ed., 8 vol. (1937–47). (G. B. K.; P. M. FE.)

EXPONENT, in algebra, indicates the number of times an expression is to be taken as a factor to form a power; the exponent is written above and to the right of the expression. For example, in the monomial $5x^4$, the exponent of x is 4 and it states that x is to be taken four times as a factor to form the fourth power; that is, $5x^4$ means $5xxxx$.

In general, if n is a positive integer the meaning of the nth power of the number a is given by:

$$a^n = a \cdot a \cdot a \cdot \ldots \text{ to } n \text{ factors (where } a \cdot a = a \times a)$$

The exponent of the power a^n is n and the base of the power is a. The fundamental laws of exponents that follow from this definition are:

1. $a^m \cdot a^n = a^{m+n}$
2. $(a^m)^n = a^{mn}$
3. $(ab)^n = a^n \cdot b^n$

From 1 and 3 respectively are derived the two additional laws:

4. $\dfrac{a^m}{a^n} = a^{m-n}$

5. $\left(\dfrac{b}{a}\right)^n = \dfrac{b^n}{a^n}$

in which a is not zero.

As algebra developed, meanings were gradually found for such forms as a^0, $a^{\frac{1}{2}}$, $a^{\sqrt{2}}$ and finally for a^r where r is any real number. As new types of exponents were suggested, the question arose as to what meanings were to be assigned to them so that the fundamental laws should be retained. Thus, if the first law is to hold for $m = 0$, then $a^0 \cdot a^n = a^{0+n} = a^n$. Evidently this requires that a^0 should equal 1 for all numbers a except 0. Again, if the second law is to hold when $m = \frac{1}{2}$ and $n = 2$, then $(a^{\frac{1}{2}})^2 = a^1 = a$. Thus $a^{\frac{1}{2}}$ must be a square root of a for all real numbers a that are not negative. Usually $a^{\frac{1}{2}}$ is taken to be the positive square root ($\sqrt{a}$) of a. Similarly $a^{\frac{p}{q}} = \sqrt[q]{a^p}$ where p and q are positive integers and a is any real value when q is odd but is not negative if q is even. Finally, if a is a nonnegative real value and 1, 1.4, 1.41, 1.414, ... is a sequence of decimals that converges to $\sqrt{2}$, then $a^{\sqrt{2}}$ is defined to be the limit of the sequence

$$a^1, a^{1.4}, a^{1.41}, a^{1.414}, \ldots$$

In general, if $x_1, x_2, x_3, \ldots$ is a sequence of positive rational numbers which has the irrational number r as limit, then a^r is the limit of the sequence

$$a^{x_1}, a^{x_2}, a^{x_3}, \ldots$$

for all positive real numbers a. Once a^r has been defined for a positive real number r, a meaning is immediately given to a^{-r} by the relation $a^{-r} = \dfrac{1}{a^r}$ when a is not zero. Thus a^r is defined as a real number for all real numbers r when a is positive and for all rational numbers r with odd denominator when a is negative, the definitions being so formulated that the laws 1 to 5 remain valid.

Although the idea of the power of a number is very old, and the use of the exponent (but in a form unlike that just given) is found as early as the 14th century, the actual form a^3 first appears in René Descartes's *La Géométrie* (1637). Moreover, some of the facts concerning rational exponents were known before the introduction of Descartes's notations. However, the first algebraist to fully explain the significance of negative and fractional exponents was John Wallis (1655), his work being supplemented soon after (1669) by Sir Isaac Newton.

When z is real, the definition of exponent leads to the equation

$$e^z = 1 + z + \frac{z^2}{2!} + \ldots + \frac{z^n}{n!} + \ldots \qquad (1)$$

where e is the base of natural logarithms (2.71828 ...). Since the series on the right of (1) converges for all complex values of z, this relation may be used to define e^z when z is a complex number. The law $e^u e^v = e^{u+v}$ is then readily verified when u and v are real or complex numbers. Also, if i is the imaginary unit ($i^2 = -1$), the relation

$$e^{\varphi i} = \cos\varphi + i \sin\varphi \qquad (2)$$

due to Rogert Cotes (1722) is easily derived. This relation shows that if π is the ratio of the circumference to the diameter of a circle, then $e^{\pi i} = -1$ and $e^{2\pi i} = 1$ and hence the exponential function e^z is periodic with the imaginary period $2\pi i$; its inverse, the logarithmic function $\ln z$ (natural logarithm of z), has infinitely many values, at most one of which is real. A real value occurs if and only if z is a positive real number. Equation (2) may be used to show that every complex number not zero is a value of the exponential function. In fact,

$$a + bi = e^{x+iy} \qquad (3)$$

where $x = \frac{1}{2} \ln (a^2 + b^2)$ and y is the radian measure of an angle for which $\cos y = \dfrac{a}{\sqrt{a^2+b^2}}$, $\sin y = \dfrac{b}{\sqrt{a^2+b^2}}$. Thus the value of x is uniquely determined, but two values of y may differ by any integral multiple of 2π.

Equation (3) suggests the following general definition of exponent:

$$(a + bi)^{c+di} = e^{(x+iy)(c+di)} \qquad (4)$$

where x and y have the same values as in (3). Because of the lack of uniqueness in y, this general power has infinitely many values except when c is rational and $d = 0$. For this exception, the number of values is the denominator of the fraction c when written in lowest terms. If c is irrational, there is at most one real value among the values of the power. If c is rational and $d \neq 0$, there are either infinitely many or no real values for the power, whereas if $d = 0$ there are at most two real values. For various applications of exponents, *see* ARITHMETIC; ALGEBRAIC GEOMETRY; LOGARITHMS; NUMBERS, THEORY OF. (R. W. BD.)

EXPORT-IMPORT BANK OF WASHINGTON. This bank was established on Feb. 12, 1934, as the principal agency of the U.S. government in international finance. Its primary purpose was to help create better world markets for U.S. products. In the pursuit of this aim it emerged as a pioneer in the U.S. effort to promote the economic development of underdeveloped countries. During the first 25 years of its existence the bank authorized a total of $10,000,000,000 in credits for 1,600 projects in 69 countries. The bank's principal customer in this period was Latin America.

The bank has had a checkered history. Nothing ever came of its original purpose—to finance trade with the Soviet Union—because the Soviet government refused to discuss outstanding debts in the United States. The bank then turned its resources, initially only $11,000,000, to stimulating U.S. exports as a means of supplementing the antidepression measures of Pres. Franklin D. Roosevelt's "New Deal." During World War II the bank helped meet the capital needs of Latin-American countries that were cut off from their normal commercial outlets in Europe. A reorganization of the bank in 1945 provided it with increased re-

sources to assist in European reconstruction. The launching of the Marshall plan (*q.v.*) for European aid in 1948 enabled the bank to give more attention to other areas, particularly Asia, Africa and the Middle East.

The Export-Import Bank act of 1945 vested the management of the bank in a board composed of a president, vice president and three directors appointed by the president of the United States with the advice and consent of the senate. The act requires the board of directors to make semiannual reports to congress on the bank's operations. The bank's total lending authority was fixed on May 22, 1958, at $7,000,000,000. This sum consisted of $1,000,000,000 of capital stock subscribed by the U.S. treasury plus a maximum of $6,000,000,000 in additional credits from the U.S. treasury, repayable at interest. Unlike the International Bank for Reconstruction and Development and the International Finance corporation (IFC), the Export-Import bank has not been free to borrow in the open market.

The act of 1945 requires that the bank supplement and encourage but not compete with private capital. Before the bank extends its own credit, therefore, it has to be satisfied that no private funds can be found on reasonable terms. The bank has extended loans of three broad types: (1) development loans to foreign governments and enterprises for the purchase of capital goods; (2) exporter credits created by purchasing from U.S. exporters of capital goods some portion of their claims against customers abroad; and (3) commodity loans to finance the disposal abroad of U.S. agricultural surpluses. The bank invites private capital to participate in its ventures and, whenever feasible, seeks to make sales from its portfolio to private investors. In addition to direct credit transactions the bank has also issued war risk and expropriation insurance.

The Export-Import bank extended loans ranging from $550 to $100,000,000 to individual private enterprises for periods of time ranging up to 20 years—without government guarantee. These loans were repayable in dollars, a currency that was often hard to obtain in many areas of the world. Yet over the first quarter century of its existence, the Export-Import bank had to write off as uncollectible loans amounting to only one twenty-fourth of one per cent of its total disbursements. The bank was able to pay $218,000,000 in dividends and to accumulate $550,000,000 in reserves against future emergencies. Because the dollar repayment clause limited the bank's scope as a development agency, the Development Loan Fund was created in 1957. At the outset the fund was authorized to accept repayment in "soft" currencies.

The bank's loans are made with the stipulation that the recipient will purchase U.S. products. This stipulation reveals a sharp contrast between the Export-Import bank and its international counterparts, the International Bank for Reconstruction and Development and the International Finance corporation. The difference reflects the fact that the Export-Import bank was originally conceived as an agency to combat depression in one country, rather than as an international agency for joint action to promote economic development. From the point of view of receiving countries, the necessity of purchasing U.S. products may lead to inefficient use of funds when purchases can be made at lower prices in other countries. In attempting to meet this criticism, the United States has not changed the character of the Export-Import bank but has increased its participation in international agencies, including the Inter-American Development bank, founded in 1960.

As required by the Bretton Woods Agreement act of 1945, the lending operations of the Export-Import bank have been co-ordinated with those of other government institutions involved in international finance through the National Advisory Council on International Monetary and Financial Problems, of which the president of the bank is a member. Within the scope of co-ordination, the Export-Import bank has been entrusted with various supplementary tasks. It has, for example, made stabilization loans to countries with balance-of-payments difficulties. Under the Agricultural Trade Development and Assistance act of 1954 the bank can receive, for lending, up to 25% of the local currencies paid for the purchase of U.S. agricultural commodities abroad. The bank furthermore administers loans and guarantees negotiated or authorized by other agencies of the government. *See* also BANK FOR RECONSTRUCTION AND DEVELOPMENT, INTERNATIONAL; FOREIGN AID PROGRAMS; INTERNATIONAL FINANCE CORPORATION.

Current statistics are summarized annually in *Britannica Book of the Year*.

(H. O. Sc.)

EXPORTS are usually defined as sales abroad of a country's goods and services. The term "sales" includes barter as well as the exchange of goods and services for money.

Historical Development.—The ancient Greek and Phoenician export merchants were not parting with their goods for the currencies of other countries. What they were interested in was goods, and to them precious metals were goods rather than money. Not until about the middle of the 2nd century B.C. did Rome introduce gold and silver coins usable in foreign trade. When its empire spread over the greater part of the known world, Rome's money virtually became the world's currency. No great exporters of their own products, the Romans depended on large imports of grain, metals, textile fibres, dyestuffs, spices and, particularly after the end of the republican era, luxuries and works of art. Many of these exports to Rome were not paid for in either money or goods. They were the spoils of war or tribute exacted from conquered or protected nations, and these involuntary exports often included large crowds of slaves. In a certain measure, however, the Romans paid indirectly by performing abroad important services on a vast scale. The efficient administration, public works and communications by land and by sea which they established and maintained in most parts of the empire considerably helped the expansion of many other countries' exports. When Roman power collapsed in the west, the Byzantine empire remained a great market for goods of other countries as far away as India and China.

In medieval Europe new centres of export trade developed, in Italy, the Netherlands, France and Germany. Early in the 13th century England became the main exporter of wool and woolen cloth to Europe; this was one of the most important trades at that time. The "staple," where English export commodities were collected and stored under public control, became an institution of more than commercial importance, since on the customs duties levied on English exports depended the king's finances. Discovery of new sea routes and the rapid growth of navigation and sea power during Queen Elizabeth I's reign opened new prospects for exports. The preponderance of Venice was over. Interest in distant markets, both east and west, diverted the main stream of England's trade away from the Netherlands and France. Wealthy and politically influential export merchants gave strong backing to Sir Francis Drake, Sir Martin Frobisher, Sir John Hawkins and other adventurous seamen who led the way to the colonial markets that were to offset the loss of the wool markets in the Netherlands, Hamburg and the Hanse towns in the 17th century. Trading companies—one of them the famous East India company—were set up in London for doing business with the Baltic, Russia, the Levant and farther east. In France the Compagnie des Indes Orientales, established by Jean Baptiste Colbert under Louis XIV for similar purposes, was much more of a government institution.

The organization of trade, based since the early middle ages on a rigid municipal and guild system, was no longer suitable for the new methods and scope of the export trade. Merchant capitalism, already apparent in the 14th century, received a great impetus under Elizabeth I in England and Louis XIV in France, and became a potent factor in government policy at home and abroad. Since the trading companies were granted monopoly rights within the geographical areas assigned to them, they undertook to protect their interests from "interlopers" and also to defend with armed force their establishments in foreign parts and their ships on the open seas. Special navigation laws regulated access to ports and use of sea routes, while import prohibitions and customs duties helped the change-over from guild and handicraft production to capitalist organization of the industries which supplied the goods for export. These developments led, in the 17th and 18th centuries, to the firm establishment of the mercantilist system, according to which exports were the only way of attracting precious metals to a country and building a "favourable" balance of trade. Governments saw great economic and political advantages in the accumulation of precious metals, and grave dangers in any

failure of exports to outrun imports. National policy was consistently, often ruthlessly, pursued on these principles, with active intervention of governments in trade and exchanges. Later on, the Industrial Revolution and the paramount importance it gave Great Britain in international trade provided the ground on which free trade and *laissez faire* were to thrive for many years. In complete opposition to mercantilism, the free play of supply and demand in the world markets, which at that time converged on London, became the main object of trade policy, and competitive struggle for individual profit was relied upon to ensure the flow of goods, services and precious metals that led to a favourable balance of payments. The working basis of that system was the gold standard.

In the early 1930s the system collapsed under the strain of great accumulated changes in economic conditions brought about by the Industrial Revolution itself. Probably more than anything else, heavy and lasting industrial unemployment contributed to replacing the automatic checks of the gold standard by government controls of international trade and payments. The mobilization of all economic resources of the great industrial countries in World War II extended the scope of government intervention in foreign economic relations. Free trade, largely in abeyance after 1930, ceased to be an object of practical commercial policy; *laissez faire* became a term of reproach. This return to mercantilism had been endorsed by J. M. Keynes in *The General Theory of Employment, Interest and Money*, published in 1936. After World War II, it found effective support in the wide swing of political opinion in favour of the welfare state, in Europe and elsewhere.

These changes in the political climate had considerable bearing on exports. Many governments adopted the "national income approach" to economic problems, with particular emphasis on high employment. In this context, rising exports were thought desirable not only as earners of foreign exchange but as a source of additional employment and also as the only means of securing the raw materials and other goods and services needed for maintaining full employment. Export promotion became one of the major aims of economic policy in most parts of the world. Since, however, the currencies of very few countries other than the United States were convertible, the geographical redirection of exports, away from weak (or soft) currencies and toward the dollar area, had to be enforced by various methods. Subsidies in the form of tax rebates to producers and exporters; differential exchange rates favouring exports to certain "desirable" markets; export premiums and licences; discriminatory allocation of building licences and scarce raw materials; partial exemption to some exporters from the obligation to surrender their dollar earnings to the exchange control; bulk sales of exports through government organizations; and many other devices were being used in these attempts at central management of many countries' exports.

Regulations and Restrictions.—Much more harm was done to international trade by the even more intricate regulations and restrictions applied to imports, not only for balance of payments purposes, but also for protecting domestic price levels, safeguarding the home industries and avoiding falls in employment. Several efforts were made for the removal by agreement of the worst impediments to foreign trade. The most comprehensive experiment in this field, the General Agreement on Tariffs and Trade (G.A.T.T.), which became provisionally effective in 1948, allowed many and important exceptions to its provisions. These were used so readily and persistently by some of the parties to the agreement that there seemed to be little ground for expecting decisive beneficial effects on exports in the near future.

More immediate results were achieved by the European Payments union (E.P.U.), set up in 1950. During its first three years of activity, about four-fifths of the gross trade payment deficits and surpluses submitted by the members of the union were cleared by setting off credit against debit items, and only the balance had to be dealt with partly by transfer of gold and partly by overdraft facilities within the framework of the system. Still more important was the "liberalization" policy to which E.P.U. member states were committed. They had to remove, as far as practicable, all quantitative restrictions (by quotas) on imports of goods handled by private (*i.e.*, nongovernment) trade. The European Monetary agreement superseded the E.P.U. in Dec. 1958. At the end of 1959 the average degree of liberalization was 92% for intra-European trade and 86% for trade including the United States and Canada. (*See* European Payments Union.)

Invisible Exports.—A major factor to be taken into account when considering exports as a whole is the part played by receipts from services like shipping, other forms of transport and communications, banking, insurance, royalties, income from investment abroad, emigrants' remittances and spending by foreign tourists and travelers. These are often called invisible exports. The ratio of invisible to merchandise exports changed considerably in some of the important trading countries after World War II. The change was particularly striking in the United Kingdom, largely as a result of the heavy reduction in income from overseas investment. Invisible exports are of vital importance to countries whose limited resources in foodstuffs and raw materials make them dependent on large imports.

In estimating invisible exports every care is taken to leave out all transactions on capital account, such as loans or investment by foreigners, transfer of gold and exchange reserves. Only current earnings from investment abroad are included. There is a great deal of confusion in the use of the term "export of capital." If it applies to money sent abroad to be invested in a foreign country, this is not export but an addition to the exporting country's total investment. If what is meant is the shifting of bank balances from one country to another, no export is involved but only transfer of claims on money. When stocks and shares are sold abroad, this is a case of disinvestment, not of export.

In the 1930s the term "unrequited exports" crept into use, mostly in newspaper writing and political discussion. It applied to exports which earned no foreign currency but were paid for by the foreign buyers' drawing on their credit balances in the exporting country. After World War II unrequited exports became an awkward problem for the United Kingdom where large sterling balances were held by members of the sterling area and also by other overseas countries.

CHANGES AFTER WORLD WAR II

Although World War II had caused destruction and losses on a scale never seen before, many countries emerged from the trial with greatly increased purchasing power more widely spread among their population. Even some of the belligerents who had suffered extensive physical damage had been producing at a high rate while modernizing their industrial and agricultural equipment under the hard pressure of wartime necessity. Vast areas, unaffected by actual warfare, found ready buyers for the raw materials they were producing and experienced for a while conditions approaching prosperity. Other parts of the world, including highly industrialized countries in Europe and enormous underdeveloped regions in Asia, came out of the war economically crippled, unable to resume international trade for several years. In practically every country, whether richer or poorer at the end of World War II, prolonged scarcity of supplies had brought about a great accumulation of demand for consumer goods and for capital equipment of all kinds. This demand was considerably intensified by the trend toward industrialization in underdeveloped countries and the urgent need for repairing war-damaged plants, buildings and communications elsewhere. Everyone was prepared to pay high prices in order to get served without too much delay. This pressure of universal demand began to ease up about mid-1949, but the effects of the recession that followed were quickly overtaken by the outbreak of hostilities in Korea in 1950. Anticipating another period of shortages and high spending on war materials, governments, traders and industrialists joined in the rush to buy and order ahead goods far in excess of current needs. Export licensing was reintroduced in many parts of the trading world, and some governments imposed heavy taxes on the exports they were still permitting. As the expected return of wartime conditions did not take place, the surge of demand broke down in 1951.

Prices began to sag. Overstocked markets, large deficits in balances of payments and fears of unemployment led to drastic and extensive cuts in imports, particularly in Great Britain, France, Australia, New Zealand, South Africa and many Latin-American countries.

After declining for about 12 months, the volume of world trade resumed its upward trend in the autumn of 1952 and, in spite of temporary local setbacks, continued to rise. A significant feature in the postwar position was the faster rate of increase in manufactured exports than in total exports. This was a reversal of the prewar relation between these two rates. Of even greater importance was the splitting up of the world market into more or less exclusive compartments, the most sharply defined separation being that between the centrally planned economies of the U.S.S.R., eastern Europe and Communist China, on the one hand, and the rest of the world, where private enterprise remained the main basis of economic organization, on the other. Trade between these two parts of the world shrank considerably while exports and imports inside the two groups expanded steadily. Within the free enterprise sector four main areas became clearly outlined, though not rigidly fenced off against each other: (1) the dollar area; (2) the sterling area; (3) the E.P.U. group; (4) Latin America. Japan was the only important trading country unattached to any area.

The Dollar Area.—Consisting of the United States, Canada and the dollar countries in Latin America (Cuba, Mexico and Venezuela), the dollar area became the largest exporting region in the world. After providing the major part of the financial and industrial resources that won World War II, the United States remained the largest supplier of the goods and services so urgently needed in the years that followed. Since no currency apart from the dollar was of much use for international payments, the United States had, in many cases, to supply also the means of payment for these goods and services. The steady outflow of U.S. resources, in the form of loans or grants of help, saved the world from economic collapse by enabling foreign importers to pay for the great expansion of U.S. exports, both visible and invisible. U.S. exports to the sterling area increased about twice; to Canada, more than six times; to Latin America, about seven times; to the countries of the E.P.U. (United Kingdom excluded), about four times; and to the rest of the world, about three times. After allowance for price increases, it is clear that there was a general shift toward North and Latin America. The disappearance of exports to Communist China ($35,000,000 in 1938) was offset by much higher exports to Japan, Korea and Indonesia. By groups of commodities, the largest export increases were in machinery and transport materials, motor vehicles, iron and steel, coal, petroleum products, chemicals and drugs, animal and vegetable fats, paper and cotton cloth. On the other hand, exports of corn (maize), wheat, flour, crude petroleum, preserved fruit and some other agricultural produce declined substantially and those of raw cotton, after a steady increase between 1946 and 1950, gradually fell back. Restrictions on exports, introduced in the United States after the outbreak of hostilities in Korea, slowed down the expansion, while the persistent shortage of dollars in most foreign markets, and import cuts imposed on that ground, undoubtedly prevented demand for U.S. goods from playing its full part in international trade. In the rest of the dollar area, Venezuela's exports made a spectacular advance in the period 1938 to 1952. This was mainly caused by a more than threefold increase in exports of crude petroleum. Canada's sales abroad increased fivefold between 1938 and 1954. Western Europe's steady economic expansion, particularly after 1955, reduced the demand for U.S. goods while western European exports to the United States continued to rise. As a result, the combined gold and dollar reserves of the countries of the Organization for European Economic Co-operation (O.E.E.C.) increased quite substantially, and the "dollar gap" in international payments disappeared, at least as far as trade with western Europe was concerned.

The Sterling Area.—One of the measures of World War II that had considerable effect on the external trade of this informal but fairly closely integrated group was the introduction of exchange control. While payments and transfers remained free within the area, transactions in foreign currencies were strictly regulated and all dollar earnings of the individual members (except those of South Africa) were concentrated into a common pool. Limits were set by agreement to the dollar outlay of every individual member. From the point of view of payments on current trading account between members, the sterling area was a vast market little hampered by exchange restrictions, where the members' exports enjoyed some preferential treatment in terms of customs tariffs, supplemented by massive U.K. purchases of primary products at prices arranged by agreements often spreading over several years. As a result of large wartime supplies to the western Allies and of the sterling area countries' own contributions to the war effort, they accumulated large sterling balances in London. After the war these funds were used for sustained imports of U.K. goods which were needed for consumption, reconstruction and new capital equipment.

Meanwhile, structural changes distorted the pattern of triangular trade which, in prewar times, had enabled the United Kingdom to square its dollar accounts by setting off the sterling deficits of the rest of the area with London against the regular and very substantial dollar earnings of the area's producers of raw materials. After World War II these dollar earnings were not high enough to make up the losses in convertible earnings of sterling area goods from other parts of the world. As economic development and industrialization advanced in many parts of the sterling area, a growing proportion of the foodstuffs and raw materials they produced was consumed by their increasing population or was used for manufacturing goods, some of which they tried to export. A few of these countries became net importers of the foodstuffs and raw materials they used to export before the war. The general trend in the independent countries within the sterling area was toward importing from the United Kingdom less of consumer goods and more of industrial and agricultural machinery and heavy engineering goods. Together with a similar shift of demand outside the area, this led to higher British exports of engineering products, chemicals and goods for industrial use, while exports of some consumer goods, particularly in the textile group, declined.

In the turnover of world trade the sterling area lost ground after the early 1950s, mainly to continental Europe and Japan. The formation of the European Economic Community and its counterweight, the European Free Trade area, confronted the sterling area with structural problems whose solutions were likely to involve changes of considerable importance for the currents and distribution of world trade.

European Groups.—In the post-World War II period the western European nations formed two principal trading groups: (1) the European Economic Community (E.E.C.), originally consisting of Belgium, Luxembourg, France, the Federal Republic of Germany, Italy and the Netherlands (also called "the Six"), who, by the 1957 treaty of Rome, formed a customs union and undertook to integrate their economic and financial policies; and (2) the European Free Trade association (E.F.T.A.), including Austria, Denmark, Norway, Portugal, Sweden, Switzerland and the United Kingdom ("the Seven"), set up in Dec. 1959 for eliminating customs and quota barriers between the members, who remained free in their trade relations with countries outside the area. For the purposes of this summary of developments after World War II, the United Kingdom, although a member of E.F.T.A., is included in the sterling area. Up to the return to currency convertibility at the end of 1958 the European nations were settling payments to each other through the European Payments union (E.P.U.), which performed its clearing functions with great success. Although convertibility led to less extensive use of a mechanism of this kind and to the replacement of E.P.U. by the somewhat looser European Monetary agreement (E.M.A.), the regional character of the monetary relations between the leading continental countries in western Europe became even more pronounced as a result of the co-ordinating functions of the E.E.C. By the end of 1948 western Europe's trade had almost recovered its prewar volume in imports and exports taken together. To a great extent this resulted from a considerable rise in American exports to Europe, making up for the slow advances in European

exports to the world and especially to the dollar area. From 1949 onward, however, the European countries' exports rose steadily, were almost doubled by 1955 and continued to grow, though at a slower rate, after 1960. The rising trend was particularly marked in exports within the area itself, concentrated chiefly on manufactured products. Another significant feature was the substantial increase in the area's exports to the United States: between 1950 and 1959 western Europe's share in U.S. imports rose from 15% to 30%.

Latin America (Nondollar Countries).—Up to World War II the principal export market for this area had been western Europe. This completely changed after the war. In 1938 the United States bought about 20% of Latin America's exports and Europe 66%; in 1951 about 37% went to the United States and only 41% to Europe. One reason was the long interruption of sea transports between 1939 and 1946; another was western Europe's inability to supply quickly enough the engineering goods and other manufactures urgently needed by all South American countries. With the exception of Switzerland and possibly Belgium, no European country could pay in convertible currency for its imports from any source, so the dollar markets, where demand was buoyant, became a magnet for Latin-American exports.

The trouble was that less and less could be exported because of increasing consumption by the local population and even more, perhaps, because of unsuccessful experimenting by governments in export marketing, industrialization, prices and exchange control.

The Communist Area.—A common political ideology and economic organization molded the U.S.S.R., the east European and Asian Communist republics and China into a vast, largely self-contained trading area characterized by state ownership and management of centrally planned industry and foreign trade. Exports and imports were not only administered but actually made by government institutions with monopoly rights. As a result of the high priority given to industrial development in all these countries their agriculture failed to recover after the war, and their exportable surpluses of foodstuffs and raw materials dwindled, in many cases to vanishing point. The area's enormous demand for manufactured goods found little response in the rest of the world, partly because of the embargo after 1949 on the supply of strategically important goods to Communist governments. While trade with outside markets declined heavily after the end of the war, the exchanges of goods and services within the area expanded year by year. From the very low level reached in 1946, the Communist area's exports to the rest of the world rose above prewar level (in current dollar value) by 1950, and its trade then grew faster than trade in the western world, the steepest rise being in exchanges within the area. The U.S.S.R. emerged in the 1950s as a large exporter of oil to the outside world's markets and made efforts to supply heavy industrial equipment, mostly on credit, to underdeveloped countries outside the Communist area.

Of the trading countries lying outside the areas just outlined, Japan was the largest and most efficient exporter. In spite of heavy losses in producing capacity and outlets in foreign markets, exports after 1947 showed remarkable recovery, greatly assisted by the United States, especially in the earliest postwar period.

EXPORTS IN PRACTICE

Methods of Selling Abroad.—Methods of exporting vary according to the kind of goods, the laws and practices in the exporter's and the importer's market, the risks exporters are willing to bear and the expense they can afford for setting up and maintaining their own selling organization abroad. In the course of time export techniques had to be adjusted to changes in technology and political and economic conditions. By the middle of the 20th century government controls and restrictions were inducing or compelling exporters to alter their usual methods, sometimes at heavy cost in time and money. Raw materials and nonprocessed foodstuffs are generally exported by merchants who buy for their own account, or on commission by agents acting on behalf of foreign buyers. In some cases agricultural produce is exported by co-operative societies formed by the growers. With the increase in government trading considerable amounts of exports were sold directly by governments or by marketing boards operating under their auspices. In the U.S.S.R., China and the other countries within the Communist group, all exporting is done by government agencies. Producer goods (industrial equipment, ships, railway material, power-generating plant, etc.) are usually exported by the makers, sometimes on contracts awarded after competitive tendering. Manufacturers of such goods have their own agents in the importing countries and submit offers through these agents who are responsible for steering every transaction through the intricate and changing local regulations. In exports of heavy factory equipment to underdeveloped countries the supplier sometimes has to carry out substantial construction work on the spot, even to operate the plant over a period of time and train a staff to take over from its technicians. Merchants are of little use in these exports. The ability and personal standing of the exporter's agent are of major importance.

For the enormous variety of consumer goods the easiest method is to sell to specialized export merchants who buy either for their own account or on customers' orders from abroad. Some merchants and export agents merely confirm to the manufacturer their foreign customers' orders and take responsibility for payment. The manufacturer delivers the goods to a port, railway station or airport, in accordance with the merchant's or agent's instructions, and receives payment exactly in the same way as for goods sold in the home market. Risks and complications begin to pile up when the purchases from the manufacturers are to be paid for after the goods have left the exporting country. Although many agents give a guarantee (called del credere) for ultimate payment, it may be difficult to ascertain the agent's total commitments and his ability to fulfill them in a falling market. In the United States a large amount of export business is done through commission selling houses acting on behalf of groups of manufacturers and doing the combined work of export merchants and sales agents. For goods exported in a great number of small shipments to different buyers, the mail-order technique, widely used before World War II, came into its own again after the war. In the United States, the United Kingdom, the Federal Republic of Germany and some other countries there are many specialized, highly efficient businesses in the mail-order trade. They usually pay the manufacturer in cash but are hard bargainers.

Export-minded manufacturers try to set up their own selling organizations abroad, and sometimes succeed. They appoint one or several agents in each of the export markets with reasonably good prospects, and expect them to collect and pass on a steady flow of orders. When goods are shipped on such orders the agents are advised, sometimes supplied with a copy of the shipping documents, and instructed to give assistance in case of difficulties. To agents of sufficient standing and reliability goods may be shipped out on consignment, with instructions to sell them and remit the proceeds. For the efficient selling of certain durable goods the manufacturer's agents have to carry adequate stocks of spare parts, to ensure servicing of the goods after sale, to give technical advice and to organize advertising and sales promotion. Careful choice of agents is as essential as keeping in close touch with them through frequent visits by senior members of the manufacturer's managing staff. Firms of modest size and resources may often find this method of exporting beyond their means. Some large companies send out members of their own staff to reside in export markets; others take a shareholding interest in existing local concerns or make licensing agreements with them for the manufacture or assembly of the goods previously exported. In such cases visible exports are replaced by invisible receipts of royalties, other fees and income from investment abroad. For the automobile and some other industries, restrictions in importing countries have made it necessary to export "knocked down" vehicles and machines, and to have them assembled by subsidiary or associated companies. Some importing countries insist that locally produced parts and components must be used in the highest proportion practicable.

Transport, Insurance, Shipping Documents.—Export goods are usually sold for delivery "ex-factory" (mill, warehouse,

etc.), or f.o.b. (free on board), or c.i.f. (cost, insurance and freight). In the first case the buyer or his agent has to make all arrangements for transport and insurance. In the second the exporter's responsibility extends to the moment when the goods have been delivered on board the ship, vehicle or aircraft that will be carrying them in accordance with the conditions of sale; a variant frequently used in Canada is f.a.s. (free alongside ship). In c.i.f. contracts the exporter must arrange for shipping and insurance as required in the contract and pay in advance the freight, insurance and other charges to destination. Transport is a highly specialized business, requiring great care and much experience. Unless the exporter is selling to merchants or agents in his own country, or makes appropriate arrangements with responsible forwarding agents, shippers or shipping companies, he will have to run a fully staffed and equipped shipping department. Only firms doing large amounts of exporting can afford the cost.

Whether he does his own shipping or employs the services of specialists, the exporter must be thoroughly conversant with shipping documents: bills of lading for sea transport and the corresponding documents for rail or air transport; insurance policies; consular invoices; import and export licences; foreign exchange permits; certificates of origin and of analysis; etc. Since payment for the goods depends on strict compliance with conditions relating to these documents, considerable loss and trouble can arise from insufficient knowledge of these matters. Definitions and interpretations of the various terms used in shipping documents were agreed upon in the Uniform Customs and Practice Commercial for Documentary Credits at the 13th congress of the International Chamber of Commerce (Lisbon, June 1951). Letters of credit issued by banks in the United States are subject to this agreement. Government restrictions and controls have created more complications, not only by adding new formalities and documents but also by frequent changes of regulations at short notice or even without previous notice. Apart from the well-known insurable risks attaching to exports, new hazards have become all too frequent; *e.g.*, long delays in the transfer of export proceeds, cancellation or refusal of import licences, sudden alterations of exchange rates or reshuffling of priorities for exchange allocations, etc. A number of exporting countries devised special insurance schemes (export credits insurance) for providing cover against these risks. In the United Kingdom, Federal Republic of Germany, Italy, France, Belgium, the Netherlands, Sweden and Switzerland the ultimate loss is wholly or partly underwritten by the government.

Financing Exports.—No special financing is required when export goods are sold to merchants or agents in the seller's own country, for payment in that country. In many cases, however, the manufacturer has to deliver the goods in a foreign market, to pay in advance freight, insurance and other charges and to extend credit to his customers abroad. In actual practice few exporters can tie up large sums of money and wait until the proceeds of their exports are collected and remitted to them. There are several methods of providing the banking facilities required:

1. The foreign importer arranges through his own bankers for a bank in the exporter's country to issue a letter of credit; *i.e.*, to undertake in writing that payment for the goods plus freight and insurance will be made to the exporter against delivery, by a specified date, of the full set of shipping documents. Letters of credit can be revocable or irrevocable; in neither case can the conditions of the credit be altered by the banks without the written consent of the importer.

2. A bank in the exporter's country merely informs him, without any commitment, that an irrevocable credit has been opened in his favour by a foreign bank. If, in conveying this information, the bank confirms the credit, this is equivalent in effect to an irrevocable letter of credit under (1) above.

3. At the request of a foreign bank and for its account, a bank in the exporter's country buys his drafts on the foreign bank, with shipping documents attached. This does not offer the advantage of a letter of credit, because the bank buying the drafts does not usually commit itself in advance to do so, and the drafts are bought "with recourse against the exporter," unless the foreign bank authorizes purchase of the drafts without recourse.

4. Instead of paying cash, the bank in the exporter's country can undertake to accept bills drawn by the exporter at 30, 60 or 90 days sight, with full set of shipping documents attached to the drafts. Alternatively the bank can advise the exporter in writing that, if the shipping documents are delivered to the bank by a fixed date, bills attached to the documents and drawn on the foreign buyer (or on a bank in the buyer's country) will be accepted on presentation. Such advice does not commit the advising banker unless he explicitly confirms the arrangement. Bills accepted by a bank in the exporter's own country will be discounted without difficulty at a fine rate and he will get the price of his goods (c.i.f. or f.o.b. as the case may be) at once, less interest up to the maturity of the bills. Acceptances by foreign banks, and particularly by trading firms abroad, may be much more difficult to turn into hard cash. Such bills cannot be sold in the London discount market, but in many other markets this can be done, although at higher cost than for bills bearing home signature. In the United States most exporters prefer to sell abroad on a sight draft basis; *i.e.*, for payment in dollar cash on delivery of the documents.

5. Exports against letters of credit are not always possible, but the exporter can lodge the shipping documents as collateral security (through a "letter of hypothecation") with his banker who lends him the value of the shipment, less interest up to the collection of the proceeds and less a certain security margin.

This is a bare outline of the most usual methods of financing exports under normal conditions of trade. There are many variants and combinations in actual practice. The tangle of impediments to international trade after World War II greatly upset the smooth working of this highly developed mechanism. In practically all cases where the ultimate payment of export proceeds and their transfer to the exporter were dependent on foreign governments' controls and regulations, neither the usual shipping documents nor bills accepted in the importing country offered sufficient guarantees to the exporter or the bank financing exports.

Some of the risks involved can be covered only by special export credit insurance. In countries where such schemes are in operation, export credit insurance policies are a most useful addition to the documents required for financing exports. This device is particularly helpful when manufacturers have to tie up resources for many months in the production and installation of producer goods, and to extend several years' credit to their foreign customers. In the United States, where there is no special institution dealing with this type of insurance, the Export-Import bank plays an effective part in providing medium-term dollar facilities and guarantees, without which the export of some producer goods could hardly be financed.

Export Statistics.—International comparisons of export statistics are difficult because of different systems of recording, pricing and classification. For recording, two main systems are in common use: (1) special trade, under which exports of goods wholly or partly manufactured in the exporting country are shown together with goods re-exported after having been included in the country's imports; and (2) general trade, where to special exports as defined above are added goods withdrawn from bonded storage for re-export. Differences exist also in the valuation of imports and exports; some countries (including the United States) value both their imports and their exports at f.o.b. prices. Others (including the United Kingdom) show imports at c.i.f. prices and exports at f.o.b. prices. The League of Nations, and later on the United Nations, tried hard to remove some of the discrepancies. In 1953 the UN Economic and Social council recommended valuation of imports at c.i.f. and of exports at f.o.b. prices. From 1950 the same council urged member governments to adopt a single-standard international trade classification of all goods moving into international trade or, at least, to make available statistical data rearranged in accordance with that standard.

Detailed information on exports, by commodity and destination, can be found only in the national statistics of the different countries. These figures often need adjustment, and fairly thorough knowledge of the methods of compilation, pricing, etc., is required in using them. For purposes of international comparison

the United Nations' publications offer the best statistical material, reduced to a comparable basis as far as this is technically possible. The first comprehensive and reliable analysis of world trade as a whole, of its various crosscurrents and the interdependence of markets in the mechanism of multilateral trade, was produced by the Economic Intelligence service of the League of Nations in the volume *The Network of World Trade* (1942), a standard book of reference. *See* INTERNATIONAL PAYMENTS; TARIFFS; TRADE, INTERNATIONAL; *see* also references under "Exports" in the Index volume. (N. I. M.)

BIBLIOGRAPHY.—W. F. Wyman, *Export Merchandising* (1922); A. Calvert, *Shipping Office Organization, Management, and Accounts* (1926); W. F. Spalding, *Bankers' Credits and All That Appertains to Them* (1930); G. Haberler, *Theory of International Trade With Its Applications to Commercial Policy* (1937); W. Syrett *The Finance of Overseas Trade,* 2nd ed. (1949); C. M. Schmitthoff, *The Export Trade,* 2nd ed. (1950); Guaranty Trust Company of New York, *A Review of Export and Import Procedure* (1957); Harold J. Heck, *Foreign Commerce* (1953); J. E. Meade, *Theory of International Economic Policy,* vol. i (1951), vol. ii (1955); L. E. Walton, *Foreign Trade and Foreign Exchange* (1956); United Nations, *World Economic Survey 1955* (1956); Economic Commission for Europe, *Annual Surveys,* particularly "Economic Survey of Europe Since the War" (1953); J. Viner, *International Trade and Economic Development* (1957); C. Segré, "Medium Term Export Finance," *Quarterly Review of Banca Nazionale del Lavoro* (June 1958).

EX POST FACTO LAW. The United States constitution forbids both congress (art. 1, sec. 9) and the states (art. 1, sec. 10) to "pass any . . . ex post facto Law." It was settled at an early date that this prohibition (literally, against laws passed "from after the fact") applies only to criminal laws and is not a general restriction on retroactive legislation (*Calder* v. *Bull,* 3 Dall. 386 [1798]). An ex post facto law, therefore, is one which purports to make certain conduct criminal which was not a crime when done, increases the punishment for crimes already committed, or changes the procedure in force at the time an alleged crime was committed in a way substantially disadvantageous to the accused. An example of the last-mentioned type of ex post facto law was a statute reducing the size of the jury in criminal cases from 12 to 8 (*Thompson* v. *Utah,* 170 U.S. 343 [1898]). Implicit in the prohibition is the notion that fairness to individuals requires that they be punished only in accordance with standards of conduct which they might have ascertained before acting. The clause also serves, in conjunction with the bill-of-attainder clause, as an added safeguard against the historic practice of legislative punishment of particular individuals because of political enmity (*see* ATTAINDER). The loyalty test oaths passed after the American Civil War, although not really criminal laws, were condemned both as bills of attainder and as ex post facto laws (*Cummins* v. *Missouri,* 4 Wall. 277 [1867]; *Ex parte Garland,* 4 Wall. 333 [1867]).

The policies underlying the ex post facto clause have been recognized in most developed legal systems. They are reflected in the maxim of the civil law *nulla poena sine lege* ("no punishment without law"), a principle whose roots can be found in Roman law. In England, where parliament is not prohibited from passing ex post facto laws, the judges, in accord with common-law tradition, have refused to interpret legislation retroactively unless parliament has clearly expressed such an intention. Prosecution of the Nazi leaders following World War II for the crime of aggressive war, a crime specifically defined for the first time in the Allied charter creating the International Military tribunal for the trial of the war criminals, provoked extensive discussion over the scope and applicability of the principle against retroactive criminal laws (*see* WAR CRIMES). (P. C. N.)

EXPRESS. In transportation terminology this term denotes the speedy conveyance of small parcels. The service is commonly performed by railways, either directly or through subsidiaries, or by trucking companies. The term is best known in the United States, where it is often associated with the colourful events of frontier life.

Express service began in the U.S. during the early railroad era in the mid-19th century as a method of transmitting currency and valuable papers, a function that had earlier been performed by stagecoach drivers. Express companies were usually independent enterprises, often holding railway concessions, although after the American Civil War some western railways organized express subsidiaries. Early express service was provided by a messenger carrying a carpetbag and traveling by boat, train or stagecoach. It later was extended to include the transportation of valuables and other commodities in larger quantities. At times letters were carried in competition with the post office or under contract to it. The conveyance of money led the express companies into banking activities.

In 1918 the several express companies in the U.S. were required, as a wartime measure, to consolidate their transportation functions into a single subsidiary, the American Railway Express company. That company was bought by a group of 66 railways in 1929, and renamed Railway Express Agency, Inc. Air express was eventually handled by the Railway Express agency after experiments with various independent arrangements as early as 1919. In 1959 a new contract was negotiated between the railways and the Railway Express agency, giving more autonomy and incentive to express management.

Early History.—In 1839 William Frederick Harnden (1813–45), a conductor on the Boston and Worcester railway, undertook on his own account the carrying of small parcels and the performance of small commissions. Obliged to leave the company's service or abandon his enterprise, he started an express service between Boston and New York city, carrying parcels, executing commissions and collecting drafts and bills. Alvin Adams followed Harnden's lead in 1840, also between Boston and New York. From 1840 to 1845 similar services between the more important towns were set up by many others. An attempt to include the carrying of letters was stopped by the government as interfering with the post office. In 1854 began the amalgamation of many of the companies. Thus the services started by Harnden and Adams were consolidated under the name of the Adams Express company. The lines connecting the west and east by way of Albany, Buffalo and the Great Lakes were consolidated in the American Express company, under the direction of William G. Fargo (*q.v.*), Henry Wells and Johnston Livingston.

American Express directors could not agree on transcontinental expansion. In 1852 Wells, Fargo and others established the separate California Express, operating west of the Missouri river in competition with Adams. Co-operation between Wells, Fargo and American Express continued close; the latter was headed by a member of the Wells or Fargo families until 1914. The Adams Express company became a closed end investment trust.

The famous western expresses, Butterfield's and Holladay's Overland Mails, the pony express and others, were founded as mail carriers, with passenger and parcel service essentially by-products. In 1858 Butterfield inaugurated his Overland Mail stages, traveling between St. Louis and San Francisco via the southern deserts in 25 days. Money and valuables were handled by Wells, Fargo and Co. Service over the Butterfield route was disrupted in 1861 by the Civil War, and was re-established, but with less regularity, on the central route via Salt Lake city. Holladay obtained control in 1862.

The celebrated pony express was started in 1860 between Sacramento, Calif., and St. Joseph, Mo., the time scheduled being eight days. The service was carried on by relays of horses, with stations 10 to 15 mi. apart. The charge made for the service was $5 per one-half ounce. The completion of the Pacific Telegraph company line in 1862 led to the discontinuance of the regular service.

In 1866 Holladay's Overland Mail and Express became Wells, Fargo & Co. (Colorado). Upon completion of the transcontinental railroad in 1869, the stages were sold and the Wells, Fargo company limited itself to banking and express. By 1925 it had passed under the control of American Express. The 1852 Wells, Fargo & Co. continued its separate existence as a San Francisco bank; in 1960 it merged with the American Trust company under the new name, Wells Fargo Bank American Trust company, and in 1962 changed its name to Wells Fargo bank.

American Express began to take its modern form in 1882 with the sale of money orders, and began the sale of travelers' checks

in 1891. Expansion into Europe, first with express parcels and later with travelers, started in 1895. International banking and travel service eventually became its principal businesses. By mid-20th century the surviving express companies had all become financial institutions. *See* also PONY EXPRESS; SOUTHERN OVERLAND MAIL. (J. P. CA.)

EXPRESSION, MUSICAL, in performance, that which is added by the performer in his rendering of the notated music. In improvised or aurally transmitted music this definition would have to be modified, and the distinction between composer and performer becomes less clear. It is in fact never a precise one since the division of responsibility has varied at different times and places.

In European music before the 19th century, as in oriental music and mid-20th-century jazz music, the performer's share of the responsibility has included not only the nuances but also, to a greater or lesser extent, the notes themselves. Thus, for example, in the course of the baroque period (*i.e.*, from Claudio Monteverdi to J. S. Bach) it became usual in slow movements, both vocal and instrumental, for the composer to notate only the main structural notes of the melody, leaving it to the performer to invent more or less impromptu the ornamental figuration. In all movements he had to introduce specific ornaments such as trills, appoggiaturas, etc., and, in some cases, to modify substantially the notated rhythms. In contemporary accounts of performance the soloist is always praised for his playing of slow movements; *i.e.*, for his invention. Finally an accompanist was provided only with a figured bass line on which to construct a fully worked-out accompaniment (*see* THOROUGH BASS). None of these responsibilities would now be regarded as falling within the performer's province, but at the time they were taken for granted and regarded as customary aspects of expression. Tempo, phrasing and dynamics have always been an essential part of musical expression and, while today the composer may include some approximate indication of them in his notation, the final responsibility for carrying them out remains with the performer. The tendency of composers to include in their notation certain marks of expression increased very rapidly during the 19th century, reached the extreme of what is practicable early in the 20th century and has since declined to a more sensible level.

Tempo (the Italian for "time") is the speed at which the music is performed. At no period has it been possible for more than a hint about the approximate tempo to be inferred from the notation itself. A few written indications of tempo occur in 16th- and 17th-century music, increasing in number later in the 17th and through the 18th century, after which they become relatively standardized. The words used to denote tempo—usually Italian, like so many musical terms—are more strictly indications of mood than of tempo, *e.g.*, *allegro*, "cheerful," or *adagio*, "at ease," while more precise indications of tempo are today given by metronome markings. In no case can even these metronome markings in practice be taken as binding—Brahms wrote in a letter: "As far as my experience goes every composer who has given metronome markings has sooner or later withdrawn them." In fact the "right" tempo of any piece of music depends to a large extent on the interpretation, the performer, the instrument and the place of performance. *Rubato*, or *tempo rubato* is a flexibility of tempo and rhythm within the musical phrase. Contemporary accounts of Chopin's piano playing describe his *rubato*—the accompaniment in the left hand proceeding regularly while the melody in the right hand was rhythmically varied. But *rubato* may also affect the music as a whole, especially as *ritardandos* at cadences.

Phrasing may be described as grouping the flow of notes into natural units and subunits by means of punctuation and accentuation. The indications that the composer can give the performer through the notation are again very limited but, insofar as these can be given, they are more specific than with regard to tempo. The ligatures of medieval and renaissance notation are partly indications of phrasing (*see* MUSICAL NOTATION); so are the commas and strokes used by François Couperin (but not generally adopted at the time); so are the lengthy slurs now customary in printing keyboard music. Slurs in string parts are indications of bowing but, like many other such technical instructions, they also effect the phrasing. The articulation of individual notes is an important part of phrasing and may range from very detached (*staccato*) to very smooth (*legato*) playing. Marks of articulation are today often, but far from always, notated.

Dynamics are a modulation of the volume of the sound in accordance with the expressive requirements of the music. The voice rises with excitement and falls with relaxation; this natural principle influences all musical dynamics. Insofar as dynamics are indicated by the composer, they appear in the notation either by words such as *pianissimo* (very soft), *piano* (soft), *mezzo piano* (moderately soft), *mezzo forte* (moderately loud), *forte* (loud), *fortissimo* (very loud), *crescendo* (getting louder), *diminuendo* (getting softer), etc., or their abbreviations *pp.*, *p.*, *mp.*, *mf.*, *f.*, *ff.*, *cresc.*, *dim.*, etc., or signs such as $<$ for *cresc.* and $>$ for *dim.* The use of words, abbreviations and signs to indicate dynamics was rare during the early 17th century and still remained comparatively so until the 19th century. Dynamic variation may be employed in large-scale contrasts over considerable passages of the music, in smaller nuances within the larger contrasts and in the modification of the volume of a single note (*messa di voce*). The dynamic accentuation of strong beats, of syncopated notes and of strong discords is closely related to phrasing and forms an important part of musical expression, as does the achievement of a dynamic balance between the different parts or voices of a composition.

See Thurston Dart, *The Interpretation of Music* (1954). (R. DO.)

EXPRESSIONISM, a term first used as a generalization distinguishing the early phases of modern painting from Impressionism. It has since come to be applied explicitly to pictures and sculpture informed with a highly charged emotional content, a meaning which has been extended to describe other forms of art, especially German drama about 1914–24 (*see* DRAMA). The approach is personal and intuitive. Technique and subject matter are subordinate to the artist's emotions; it is the art of an individual rather than of a movement. The problem of finding graphic equivalents for passionate feelings outweighs other considerations; form and colour are exploited, distorted and exaggerated to achieve intensity. Although such tendencies have always existed—they are clearly exemplified in the tense, idiosyncratic work of Mathis Gothardt Neithardt (called Grünewald), Albrecht Altdorfer, El Greco and countless others—the egocentric trends of modern painting, assisted by its repudiation of realism, have encouraged the spread of an expressionistic manner that belongs unmistakably to the modern period.

The frenzied canvases of Vincent Van Gogh and, early in the 20th century, the neurotic paintings of Edvard Munch and the mask-haunted world of James Ensor made a considerable impact on European art in general and German art in particular. The emotive influence of Van Gogh, discernible in certain aspects of Fauvism (*q.v.*), was also revealed vividly in the concurrent German movement, *die Brücke*, together with that of Munch. The northern propensity for brooding on the underlying pathos of the human situation, for detecting sinister shadows behind actuality, is manifest in the work of Erich Heckel, Ernst Ludwig Kirchner, Emil Nolde and Karl Schmidt-Rottluff. By 1912 the Expressionism of *die Brücke* had been absorbed into *der Blaue Reiter*, a far larger group in which the Russian expatriate Wassily Kandinsky was a dominant figure. Kandinsky's experiments in abstract forms, prompted by what he called an "inner necessity," anticipated the development of abstract forms of Expressionism.

An overriding concern with aesthetic values has largely insulated the French schools from Expressionism. Distinguished exceptions are Marc Chagall, Georges Rouault, Chaim Soutine and Pablo Picasso (for instance, in his Blue Period and in the stark horror of "Guernica"). Among other artists who have made outstanding individual contributions to Expressionism are Ferdinand Hodler, Oskar Kokoschka, Max Beckmann, Ernst Barlach; the Belgians Constant Permeke and Gustave de Smet; the Americans Ben Shahn and Willem de Kooning; and a notable Latin-American group which includes Diego Rivera, Rufino Tamayo and José Clemente Orozco. Among sculptors who have shown marked Expres-

sionist tendencies are the Germans Ernst Barlach and Wilhelm Lehmbruck and the Italians Marino Marini and Alberto Giacometti. *See* also LITHOGRAPH: *History;* WOODCUT AND WOOD ENGRAVING: *History;* PAINTING: *Modern Painting (Late 19th and 20th Centuries): Expressionism and the German School.*

BIBLIOGRAPHY.—W. R. Worringer, *Form in Gothic* (1927); M. Raynal, *Modern Painting* (1953); B. S. Myers, *Expressionism* (1958).
(F. W. W.-S.)

EXPROPRIATION, the taking away or depriving of property or proprietary rights. The term formerly applied to any compulsory deprivation of property, particularly by a public agency, but now pertains primarily to government takings where compensation is rendered, as in exercising the right of eminent domain (*q.v.*). *See* CONFISCATION AND EXPROPRIATION. (F. H. HE.)

EXPULSION, the act of removing a person from the membership of a body, or the holding of an office, or of depriving him of the right to attend a meeting. In England the house of commons can expel a member by resolution, and such a resolution cannot be questioned in any court of law. In colonial America the legislatures generally looked to the house of commons for precedents, though they were often in a struggle with the governor and had varied success in asserting these legislative prerogatives. Most of them, at least, were able to establish themselves as final judges of the election and qualifications of their members, and some exercised a power to expel. Control over the election and qualifications of members was much the more important issue, and after the American Revolution specific sections protecting this control found their way into most of the state constitutions. The Pennsylvania and Delaware constitutions of 1776 included a power of expulsion as well, both requiring the vote of two-thirds of the house and both providing further that a member might not be expelled a second time for the same offense. The latter may well stem from reactions to the celebrated case of John Wilkes (*q.v.*), whose five-year battle with the house of commons ended when he was seated after the election of 1774. By the 1960s most of the state constitutions provided for the expulsion of members, generally requiring a two-thirds vote.

The federal constitution provides that "Each house may . . . punish its Members for disorderly Behaviour, and, with the Concurrence of two thirds, expel a Member" (art. i, sec. 5). In the same section each house is made judge of the election, returns and qualifications of its members. As a matter of history, the power to refuse to seat a member has been given a broad interpretation, and the grounds have not been limited to the age, citizenship and residence qualifications set out in the constitution. Both houses have at times refused to seat on grounds of morals, notably polygamy and bribery. The difference between the power to refuse to seat and the power to expel is an important one because exclusion can be accomplished upon majority vote, while expulsion requires the concurrence of two-thirds. Since the grounds for neither are anywhere set out, and since the action of congress is not subject to review by the courts, both are theoretically subject to abuse. In fact, however, in both the U.S. and England all cases have involved charges of illegal or immoral conduct.

Municipal corporations have no inherent powers, and in the absence of statutory authority a municipal governing body has no power of expulsion. Both the power to judge the election and qualifications of members and the power to expel are frequently given to such bodies, though it is common to explicitly limit the power of expulsion to cases of malfeasance and corruption. In the United States the courts are not in agreement on the availability of a proceeding of quo warranto (*q.v.*) to test election contests where these powers are vested in the council.

Expulsion from private groups is generally covered by their charters or bylaws. Under certain circumstances, expulsion from a private group will be reviewed by the courts. (J. D. Ls.)

EXTEMPORIZATION: *see* IMPROVISATION.

EXTENSION UNIVERSITY: *see* UNIVERSITY EXTENSION.

EXTENUATING CIRCUMSTANCES, circumstances which diminish the culpability of one who has committed an offense and so can be considered to mitigate the punishment. Many Anglo-American legal systems, including the law of England, do not prescribe minimum punishments for crime. Since the judge may consider all the circumstances in fixing the punishment up to a prescribed maximum there is little need for formal doctrines of extenuating circumstances. There are, however, instances of special circumstances which reduce an offense to one of lesser degree, such as the provocation of the accused by the victim which will reduce murder to manslaughter or, in some U.S. jurisdictions, to second-degree murder. In England the jury, on a verdict of guilty of capital murder, may recommend the accused to mercy, but the judge is nevertheless bound to pass the sentence of death and the recommendation must be considered by the home secretary. In some states of the United States and in the federal jurisdiction the jury itself has the power to elect, in a conviction of first-degree murder, between a sentence of death and one of life imprisonment.

Countries with Roman law background, such as continental European and Latin-American states, make much more use of prescribed minimum sentences for crime and consequently have had to develop more formal doctrines of extenuating circumstances. The Italian penal code gives a list of extenuating circumstances, among which are the fact that the accused acted from motives of honour, that he committed the offense in a state of intense emotion caused by grave misfortune, or that before the trial he repaired the injury by giving compensation. The French penal code provides for reduction of punishment according to a prescribed scale where the jury or court finds extenuating circumstances.
(G. B. J. H.)

EXTERRITORIALITY, a term used in international law to denote the immunities enjoyed by foreign states or international organizations and their official representatives from the jurisdiction of the country in which they are present. Exterritoriality, sometimes called extraterritoriality, extends to foreign states or international organizations as entities and to their heads, legations, troops in passage, war vessels, mission premises and other assets. It exempts them, while within the territory of a foreign sovereign, from local judicial process, police interference and other measures of constraint. The term stems from the fiction that such persons or things are deemed not to be within the territory of the sovereign where they are actually present. This doctrine was originated by the French jurist Ayraut (1536–1601) and gained wide currency because of its adoption by the classical writers on the law of nations such as Grotius and Pufendorf. The word exterritoriality or its foreign equivalent was not in use until the end of the 18th century. It gained a place in the legal vocabulary through its employment, if not creation, by Georg Friedrich von Martens (*q.v.*). His treatise on the law of nations, published in 1788, acquired an international reputation and was promptly translated into several languages, including English.

Today it is recognized that the fiction, if taken absolutely, overshoots the mark and that it is only a convenient justification for a congeries of immunities from local jurisdiction. Chief Justice John Marshall asserted no more than this in *The Exchange* (U.S. Sup. Ct., 1812) when he laid down the general principles of exterritoriality as governing American law; and it was so stated specifically by Lord Atkin when speaking for the judicial committee of the privy council in *Chung Chi Cheung* v. *The King* (1938). This case concerned a murder committed on a Chinese armed public vessel that was passing through British waters. The court held the murder to have been committed in British territory and subject to prosecution in the local courts after the wounded culprit had been brought ashore to a hospital. The actual scope of the immunities comprised in the doctrine of exterritoriality depends, according to the circumstances, on principles of customary international law as applied in a particular country, on specific statutory or executive regulation, or on international agreements, as in the case of international organizations such as the United Nations.

Heads of State.—One of the classical cases leading to the emergence of the exterritoriality doctrine was that of a foreign sovereign visiting a friendly country. It became recognized that no local jurisdiction, whether criminal or civil, could be exercised over him. The rule was later extended to republican heads of state. There is disagreement as to whether such immunity covers

his family and other members of his suite and whether it can be asserted when he enters the country incognito. In the United Kingdom the latter question has been decided in the affirmative. The privilege is waived if the foreign head of state resorts to a local court by bringing suit; it is lost if he is deposed in his homeland. In the latter case he is not personally liable for acts performed in his former official capacity, but he can no longer claim jurisdictional immunity. Thus in *Ex-King Farouk* v. *Christian Dior, Inc.*, involving an action for the price of clothing for Queen Narriman delivered before Farouk (Faruk) lost his throne, the court of appeal of Paris rejected the plea of exterritoriality and held the defendant personally liable. During World War II the United Kingdom conferred exterritoriality on the provisional governments and governments in exile by special legislation.

Diplomatic Agents.—The exterritoriality of ambassadors and other diplomatic representatives is likewise of long standing. When, for example, during the reign of Queen Anne, the Russian ambassador was arrested for debt, an international incident ensued and the famous Act Preserving the Privileges of Ambassadors, etc. (1708), was passed. The United States enacted a substantially identical statute in 1790 (22 U.S.C. §252). The Regulation Concerning the Rank of Diplomatic Agents by the congress of Vienna (1815) did not cover their exterritorial status. More than a century later, a convention on the immunities of diplomatic agents was concluded at the sixth International Conference of American States (1928) and is in force among 15 South and Central American nations. At the request of the UN general assembly in 1953, the International Law commission undertook the codification of diplomatic intercourse and immunities. The commission prepared a draft convention on that subject and, after adoption of its final form during the tenth session in 1958, transmitted it to the general assembly for further action. A United Nations Conference on Diplomatic Intercourse and Immunities was held in Vienna in the spring of 1961. It was attended by 81 nations and resulted in the signing of a Convention on Diplomatic Relations.

There appears to be general agreement that a diplomatic agent, during the term of his office, is totally exempt from both criminal and civil jurisdiction in the state where he is accredited. In *Bergman* v. *De Sieyes* (1948) a U.S. court accorded equally broad immunity to a foreign diplomat accredited to a third country while en route, although the controlling statute by its terms was not applicable in such case. According to British and U.S. practice, the immunity extends both to the family of the diplomatic agent, *i.e.*, at the least to his spouse and children living with him, and to his staff, including counselors, attachés, secretaries and clerks, as well as to his domestic servants, whether nationals or not. Other nations follow more restrictive rules. In view of this disparity the United Kingdom in 1955 passed legislation which restricted the broad immunities of the members of the family and the staff of diplomatic representatives to countries which accord corresponding privileges. The privileges of persons other than the diplomatic agent himself are merely derivative and subject to waiver by the latter. The mission and residential premises of diplomatic agents are immune not only from process by creditors but also from being entered by the police and other law enforcement officers. Whether and under what conditions they may be used to grant asylum to outsiders is controversial. An inter-American convention concluded in 1954 and ratified by seven states sanctions diplomatic asylum for political offenders and refugees.

Consular Officers.—Foreign consular officers do not enjoy exemptions from the local administration of justice to the same extent as the staffs of foreign diplomatic missions. In contrast to the situation existing in the case of diplomatic intercourse, the law governing consular immunities is less a matter of settled customary international rules than of bilateral or multilateral treaties. Among the latter type of agreements the convention relating to the duties, rights, prerogatives and immunities of consular agents, adopted in 1928 at Havana during the sixth International Conference of American States, is of special significance. It is in force in the United States, Brazil, Colombia, Ecuador, Mexico, Peru, Uruguay and six of the Central American republics. The International Law commission decided during its seventh session in 1955 to take up the topic of consular communication and immunities, and appointed a special reporter. The commission completed its work on the subject and adopted a final text of draft articles on consular relations at its 13th session in 1961. It was transmitted to the UN general assembly with the recommendation to convene an international conference.

It is generally accepted that consular officers do not incur civil or criminal liability in the state of their residence for acts connected with the performance of their functions and that they are immune from civil suits or criminal prosecution in the local courts to that extent. Beyond that they are subject to local laws and local jurisdiction. In the United States, however, civil suits against consuls and vice-consuls, except in matrimonial causes, are reserved to the federal courts. Analogous principles govern the testimonial privileges. Consular officers are entitled to decline to give evidence relating to matters within the scope of their official duties and to produce documents from their archives or official papers, especially if they deem such disclosure to be prejudicial to the interests of the sending state. Consular archives are inviolable and may not be searched by local authorities. There is mounting authority to the effect that the inviolability extends to the offices of the consulate, but the U.S. department of state has taken the position that a writ of habeas corpus may be issued to compel the release of a foreign citizen who is detained against his will in the consular premises.

Delegates to, and Officials of, the United Nations.—The United Nations, as a legal entity, its officials and the members of the delegations of the member states to the United Nations enjoy extensive procedural, fiscal and other immunities from the jurisdiction of the countries where they are present. In the vast majority of the member nations the matter is regulated by the Convention on the Privileges and Immunities of the United Nations, adopted by the general assembly in 1946 and acceded to by more than 60 countries. Separate and special arrangements, however, govern in the United States and Switzerland, in view of the fact that the former country includes the UN headquarters, while the latter country, without being a member of the organization, harbours the important UN European office at Geneva.

In the United States the privileges and immunities of the UN, its officials and employees, and of the representatives to the UN of the member states are governed by the charter, the Headquarters Agreement of 1947 (approved by public law 357, 80th congress) and the International Organizations Immunities act of 1945. By virtue of the applicable provisions, the ranking resident representatives of the member states, as well as such resident members of their staffs as are agreed upon, are entitled, within the United States, to the customary diplomatic immunities. Accordingly, for instance, they or their spouses may not be charged in United States courts with local traffic violations or sued for damages caused by negligent operation of motor vehicles. The officers and employees of the United Nations, if reported to and accepted as such by the state department, are likewise entitled to certain privileges and exemptions. Their jurisdictional immunities are considerably more restricted, however, as they extend only to acts performed by them in their official capacity and within their functions. Consequently, in the Judith Coplon case prosecution for espionage was not barred against a Soviet co-defendant because of his employment in the UN secretariat.

The headquarters district is inviolable and placed under the police power of the UN. Otherwise, federal, state and local laws remain applicable thereto and the U.S. courts retain their regular jurisdiction over acts committed on the UN premises. Thus, a visitor who commits a crime while in the UN headquarters district must be surrendered to the local authorities for punishment.

Foreign Troops and Men-of-War.—The extent of local jurisdiction over foreign troops has been a much-mooted question. Isolated dicta in decisions and statements by text writers during the 19th century supported the view that foreign armed forces, whether in transit through or stationed in a friendly country, were exempt from its jurisdiction. More recent authority, however, has restricted the immunity to one of a limited and functional character. Judicial opinions rendered after the outbreak of World

War II in Australia, Canada and the United States, as well as elsewhere, followed this trend. As a result, the matter is regulated by specific legislation, such as the Visiting Forces act, 1952, in the U.K., and by a network of international treaties, especially the NATO Status-of-Forces agreement (1951), which went into operation among all NATO members except Iceland and Germany. The latter country, however, contemplated accession to that agreement in 1959 and signed supplementary agreements thereto with the powers maintaining troops in its territory. Conversely, the immunity of war vessels, whether in transit through or at anchor in foreign territorial waters, from being seized or boarded by local authorities has remained unshaken. This status no longer implies that acts committed on such vessels while in territorial waters are deemed to have been committed abroad.

Foreign States and Their Assets.—International practice has accorded jurisdictional immunity to foreign states in their capacity as legal entities and extended the exemption of their property from seizure to assets other than men-of-war or mission premises. However, the exact scope of the controlling principles is subject to great variations and beset with technical intricacies. In the United Kingdom and in other members of the commonwealth the so-called absolute theory is followed; it prohibits impleading a foreign government, either directly or indirectly, for any reason. Moreover, in 1957, when reviewing the scope of the rule, the house of lords went very far in considering a foreign state as indirectly impleaded (*Rahimtoola* v. *Nizam of Hyderabad*). Conversely, in other countries there is an increasing trend toward differentiating between public activities and private activities, such as commercial operations, and restricting the doctrine of sovereign immunity to the former. The United States traditionally has followed the absolute theory with perhaps a somewhat more limited rule against indirect impleading. In 1952, however, the acting legal adviser of the U.S. department of state announced that his department would no longer support requests by foreign governments for immunity, except in accordance with the restrictive theory. Disagreement exists on the status of members of a federal union which have no foreign relations power of their own. French tribunals have denied sovereign immunity in such cases, while a United States court has reached the opposite conclusion (*Sullivan* v. *State of São Paulo*, 1941).

Extraterritorial Rights.—In the past many European powers, including Great Britain, and the United States concluded treaties with a number of Asian and African nations which permitted them to exercise jurisdiction over their own nationals in the territory of those countries. The conventions, establishing exterritoriality in this sense, were sometimes called capitulations (*q.v.*). They provided that the nationals of the states enjoying such extraterritorial rights, while living in the Asian or African countries signing agreements of this type, should not be tried in the local tribunals but should be brought before their own consular officers or special courts established by their governments for that purpose. The arrangements were defended on the ground that western civilization was too different from the customs and laws of the African and oriental peoples to have European or U.S. nationals subjected to the latter. These treaties were a source of international friction for years and were resented by the countries whose territorial sovereignty was limited thereby.

In the course of the 20th century all of these extraterritorial rights were terminated. Japan took the lead in obtaining the elimination of the treaty clauses of this type, the abolition taking effect in 1899. Turkey was able to achieve abrogation of the capitulatory rights in the wake of World War I. Great Britain and the United States signed treaties relinquishing their extraterritorial rights in China in 1943. Canada did the same in 1944. By the 1960s extraterritorial jurisdictional rights were accorded only to a limited extent and, generally, only with respect to foreign troops stationed in friendly countries. *See* also CONSUL; DIPLOMACY; INTERNATIONAL LAW, PUBLIC; NATIONALITY.

BIBLIOGRAPHY.—C. C. Hyde, *International Law, Chiefly as Interpreted and Applied by the United States*, 2nd ed., sec. 244 ff., 250 ff., 259 ff., 433 ff. (1945); Lord Halsbury, *Laws of England*, 3rd ed., vol. 7, pp. 265 ff., 269 ff. (1954); L. Oppenheim and H. Lauterpacht, *International Law*, 8th ed., vol. 1, sec. 115a, 143 ff., 318, 346 ff., 443 ff., 447 ff. (1955); M. Sibert, *Traité de Droit International Public*, 2 vol., sec. 184, 703 ff., 719 ff., 764 ff. (1951); E. R. Adair, *The Exterritoriality of Ambassadors in the Sixteenth and Seventeenth Centuries* (1929); S. A. Riesenfeld, "Sovereign Immunity of Foreign Vessels in Anglo-American Law," *Minnesota Law Review*, vol. 25, p. 1 (1940); E. Lémonon, "L'Immunité de juridiction et d'exécution forcée des États étrangers," *Annuaire de l'Institute de Droit International*, vol. 44 (1), p. 5 (1952); J. Dehaussy, "The Inviolability of Diplomatic Residences," *Journal du Droit International*, vol. 83, p. 597 (1956); J. B. Sialelli, "Note on Ex-King Farouk *v.* Christian Dior, Inc.," *Journal du Droit International*, vol. 84, p. 717 (1957); "Report on Status of Forces Agreements," *Record of the Association of the Bar of the City of New York*, vol. 13, p. 246 (1958); G. I. Draper, "The Exercise of Criminal Jurisdiction Under the Agreement Regarding the Status of Forces of Parties to the North Atlantic Treaty," American Bar Association, Section of International and Comparative Law, *Proceedings*, p. 55 (1957); United Nations Conference on Diplomatic Intercourse and Immunities, Final Act, *American Journal of International Law*, vol. 55, p. 1062 (1962); "Changed Policy Concerning the Granting of Sovereign Immunity to Foreign Governments," U.S. Department of State *Bulletin 984*, vol. 24 (1952); L. Cavare, "L'Immunité de juridiction des États étrangers," *Revue générale de droit international public*, vol. 58, p. 177 (1954); G. Perrin, "Les Privileges et immunités des representants des États aupres des organisations internationales," *Revue générale de droit international public*, vol. 60, p. 193 (1956); W. E. Beckett, "Consular Immunities," *British Yearbook of International Law*, vol. 21, p. 34 (1944); International Law Commission, *Report* covering work of its 13th session, p. 2 (1961). (S. A. R.)

EXTORTION, a criminal offense consisting of the unlawful exaction of money or property by threat or intimidation. In its English origin, the offense was restricted to illegal exactions by public officials, but statutes, especially in the United States, gradually extended the offense to many kinds of private coercion to compel a victim to part with his belongings. Thus, threats of harm to the person or property of another, or threats to accuse him of crime or to reveal a business or other secret may suffice. Some of these threats are occasionally singled out for special statutory treatment under the designation "blackmail" (*q.v.*).

The scope given to the offense of extortion in a particular legal system is determined partly by the content of the related offense, robbery. Robbery, formerly a capital offense in England and still so in a minority of U.S. states, is typically confined to taking property from the person or presence of the victim by violence or by threat to do an immediate physical harm. A few early decisions went so far as to include within robbery the threat to charge one with a heinous offense such as sodomy. More remote and less terrifying threats then fall within the province of the extortion and blackmail statutes. It is sometimes said that in extortion the victim consents, although under duress, while in robbery his will is overwhelmed so that there is no consent; but it is extremely doubtful whether such a doctrinal distinction corresponds to any real psychological distinction. At the opposite end from robbery on the scale of coercion, the crime of extortion is defined with an eye to bargaining processes which are lawful and even favoured in our society, and which must therefore not be subject to possible criminal prosecution. For example, a union official may threaten to call a strike for higher wages, and a purchasing agent may threaten to discontinue buying unless the seller lowers his price. Such threats are criminal only if used to obtain money or property for the personal gain of the actor. (L. B. S.)

EXTRACT, in pharmacy, the name given to concentrated preparations of a vegetable or animal drug formed by evaporating all or nearly all of the solvent from solutions of active principles; tinctures are solutions that have not been subjected to evaporation. Liquid extracts are those of a sirupy consistency and are generally prepared by treating the drug with the solvent and causing the solution to be concentrated until it attains the desired consistency. Although ordinary extracts are thick and sticky preparations, some are dry, crumbly masses or powders.

END OF VOLUME EIGHT

PRINTED IN THE U. S. A. BY R. R. DONNELLEY & SONS CO.